4.01 Comprehensive Overview

G. C. Beroza, Stanford University, Stanford, CA, USA

H. Kanamori, California Institute of Technology, Pasadena, CA, USA

4.01.1 Introduction

In general usage, the term 'earthquake' describes a sudden shaking of the ground. Earth scientists, however, typically use the word 'earthquake' somewhat differently – to describe the 'source' of seismic waves, which is nearly always sudden shear slip on a fault within the Earth (see **Figure 1**). In this article, we follow the scientific usage of the term, and focus our review on how earthquakes are studied using the motion of the ground remote from the earthquake source itself, that is, by interpreting the same shaking that most people consider to be 'the earthquake'. The field defined by the use of seismic waves to understand earthquakes is known as earthquake seismology.

The nature of the earthquakes makes them intrinsically difficult to study. Different aspects of the earthquake process span a tremendous range in length scales – all the way from the size of individual mineral grains to the size of the largest plates. They span a tremendous range in timescales as well. The smallest micro-earthquakes rupture faults for only a small fraction of a second and the duration of even the very largest earthquakes can be measured in hundreds of seconds. Compare this with the length of strain accumulation in the earthquake cycle, which can be measured in decades, centuries, and even millenniums in regions of slow strain rate. The evolution of fault systems spans longer times still, since that can require the action of thousands of earthquakes. At different physical dimensions or temporal scales, different physical mechanisms may become important, or perhaps negligible. Earthquakes occur in geologically, and hence physically, complicated environments. The behavior of earthquakes has been held up as a type example of a complex natural system. The sudden transformation of faults from being locked, or perhaps slipping quasistatically, to slipping unstably at large slip speeds, as is nearly universally observed for earthquakes, also makes them a challenging physical system to understand.

Figure 1 Earthquakes are due to slip on faults within the Earth. In large earthquakes, fault slip can reach the Earth's surface. Photo shows that surface rupture of the 1906 San Francisco earthquake offset this fence horizontally by ~8.5 ft. Plate 1-B, US Geological Survey Bulletin 324 – from US Geological Survey Photographic Library, San Francisco Earthquake, plate 41.

Despite these challenges, seismologists have made tremendous progress in understanding many aspects of earthquakes – elucidating their mechanisms based on the radiated seismic wavefield, determining where they occur and the deep structure of faults with great precision, documenting the frequency and the regularity (or irregularity) with which they occur (and recur) over the long-term, gaining insight into the ways in which they interact with one another, and so on. Yet, the obvious goal of short-term prediction of earthquakes, that is specifying the time, location, and size of future significant earthquakes on a timescale shorter than decades, remains elusive. Earthquakes are different in this sense from nearly all other deadly natural hazards such as hurricanes, floods, and tornadoes, and even volcanic eruptions, which to varying degrees are predictable over a timescale of hours to days. The worst earthquakes rank at the very top of known disasters. The deadliest known earthquake killed over half a million people in a matter of minutes. A level of sudden

destruction that no other catastrophe in recorded history – either natural or human made – has attained.

Our inability to predict earthquakes is one reason they cause such apprehension. This lack of a precursory warning is compounded by the fact that they strike so abruptly. No one can see an earthquake coming and it is only a matter of seconds from the initial perception of the first arriving waves of a large earthquake before dangerous strong ground motion begins. Moreover, large, damaging earthquakes occur infrequently (fortunately) at any given point on the Earth relative to a human lifespan. This means that most people who experience a major earthquake are doing so for the first time. Finally, there is something fundamentally unsettling about the movement of the 'solid' earth.

The unpredictability, sudden onset, and their unfamiliarity make earthquakes a uniquely terrifying phenomenon. As testament to this, other extreme and catastrophic events in the affairs of humankind – if they are devastating enough – are described with the simile, 'like an earthquake'. The point of origin of an extreme event of any kind is often described as 'the epicenter', a term borrowed from seismology.

The unpredictability of earthquakes also renders them difficult to study. Since we do not know where, and particularly when, large earthquakes will strike, collecting data on earthquakes has to be approached passively. Seismologists deploy instruments to measure seismic waves where we expect earthquakes to occur and then wait for nature to carry out the experiment. The wait can last decades, or more, for a large earthquake. Inevitably, with finite budgets and finite patience, this leads to seismic monitoring instruments being widely, and hence too thinly, dispersed in an attempt to gather data from at least some earthquakes, wherever and whenever they might occur.

Finally, the combination of unpredictability, sudden onset, long intervals between events, and unfamiliarity means that the risk created by earthquake hazards is extremely difficult for both policymakers and the general public to contend with. Because earthquakes are not predicted and occur infrequently relative to other hazards, it is understandably tempting for governments and individuals to focus on the many immediate and predictable problems that impact society more frequently. The unpredictability and sudden onset of earthquakes, however, mean that once an earthquake begins, it is too late to do much more than duck and cover.

4.01.2 Seismicity

4.01.2.1 Earthquake Size

Perhaps the best known earthquake source parameter is magnitude. Earthquake magnitude, and the characteristics of earthquake behavior that it is used to define, such as Gutenberg–Richter statistics (discussed further in Section 4.01.2.5), are purely empirical observations in the sense that they rely only on measurements of seismic waves as recorded on seismographs and do not require much in the way of assumptions about earthquake source physics. It is impossible to completely represent a complex physical process like an earthquake with a single number that measures its size. Nevertheless, several definitions of earthquake size have proven extremely useful for reaching a better understanding of earthquakes.

Earthquake size is traditionally measured by one of various magnitude scales. M_L, the local magnitude scale, was devised by Richter in the early 1930s (Richter, 1935). He was cataloging data from the southern California seismic network, and although locations, depths, and origin times for many earthquakes were available, there was no measurement of earthquake size. So, he invented one using the relation,

$$M_L = \log(A) - \log(A_o)$$

where A is the measured amplitude of the seismic trace on a standardized Wood–Anderson seismograph (**Figure 2**) at a distance of 100 km, and A_o is the amplitude of a reference earthquake with $M_L = 0.0$. The particular definition of the zero level is that an earthquake of magnitude 0 has an amplitude of 0.000 001 m at a distance of 100 km. Thus, an earthquake of magnitude 3.0 has an amplitude of 1 mm at the same distance and a magnitude 7.0 earthquake would have an amplitude of 10 m. Wood–Anderson instruments record on photographic paper and saturate at an amplitude of about 20 cm.

Until recently, quoted magnitudes for local earthquakes were often correctly described as the Richter magnitude. The old Wood–Anderson instruments, though state-of-the-art when the magnitude scale was developed, are not used now. Instead, if one wants to calculate the local magnitude, data from modern instruments are degraded to mimic recordings from the old instruments (**Figure 3**). Although the magnitude scale has proven extremely useful for

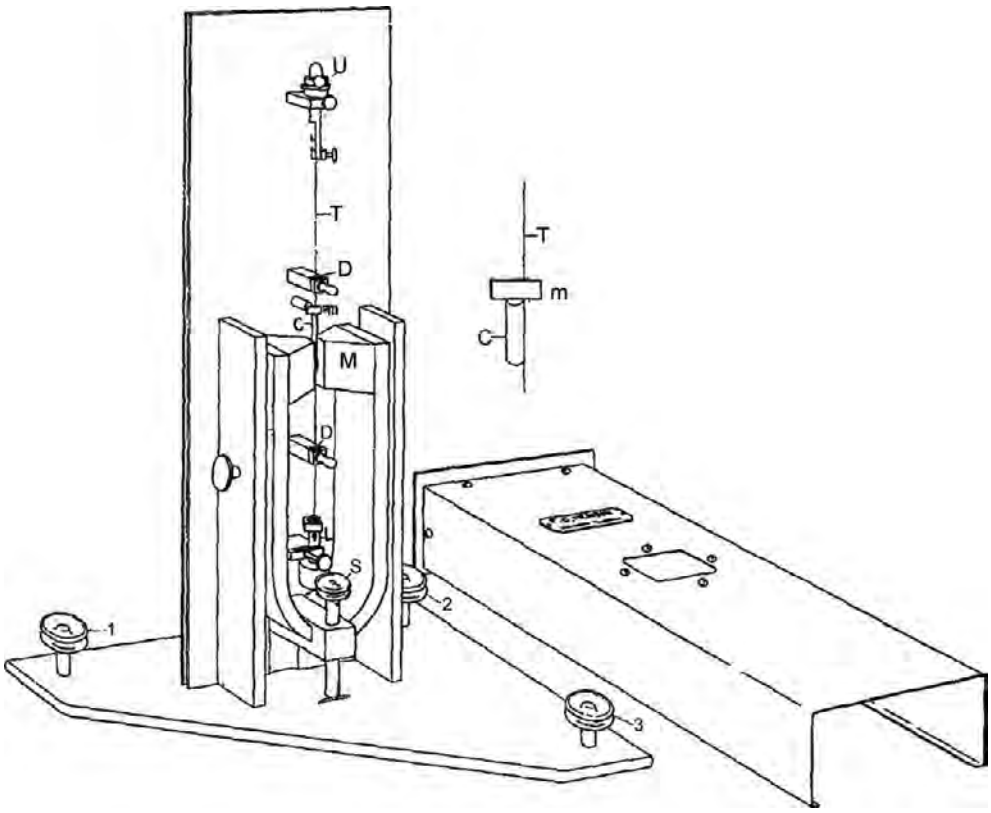


Figure 2 Diagram of a Wood–Anderson torsion seismometer, which is used to define M_L. Ground motion causes rotation of an inertial mass, C, which is attached to a thin wire under high tension, T. Seismograms are recorded on photographic paper using a beam of light reflected off a mirror, M, that is mounted on the inertial mass. From Anderson JA and Wood HO (1925) Description and theory of the torsion seismometer. *Bulletin of the Seismological Society of America* 15: 1–72.

routine cataloging of earthquake behavior, there are a several shortcomings of M_L, among them:

1. If the station is not 100 km from the earthquake, one must correct for propagation effects – this is difficult because at different distances, different wave types have the largest amplitude.
2. Excitation of waves can vary strongly with depth and this must be corrected for to compare the size of earthquakes at different depths.
3. Earthquakes have complex radiation patterns so that the azimuthal direction of the seismograph from the earthquake strongly affects the amplitude of the recorded waves.
4. The Wood–Anderson instruments respond primarily to ground motion at periods less than about 1 s. Once the size of an earthquake approaches magnitude 7.0, 1 s waves start to saturate in amplitude due to interference effects arising from the finite size and duration of the fault. This occurs even if the seismograph is operated at low gain and the waves are all on scale. Beyond $M_L \sim 7$, the local magnitude is no longer a reliable measure of earthquake size.

To some extent, difficulties 1–3 can be corrected for empirically and by averaging over many stations. Nevertheless, M_L is frequently uncertain by as much as 0.25 magnitude units and it does not extend to earthquakes recorded at distances in excess of 300 km. Hence it is not suitable for global earthquake monitoring. To remedy this, and to address difficulty 4, seismologists have invented a variety of different magnitude scales. Usually, it is one of these other magnitude scales that are used to characterize the size of an earthquake.

m_b, the body-wave magnitude scale, is defined using the initial P waves from earthquakes that occur very far away so that it can be used to compare earthquakes all over the world. It has been adjusted to agree approximately with the Richter magnitude at

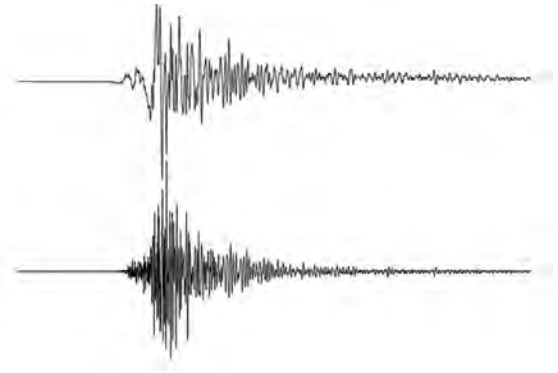


Figure 3 Upper seismogram shows data for the 28 Sep. 2004 Parkfield, CA, earthquake from a modern broadband instrument at station JRSC (Jasper Ridge, Stanford, California) and lower seismogram shows the same ground motion as it would have been recorded on a Wood–Anderson instrument. Time increases from left to right, and a total of 600 s is shown. Vertical axis shows north–south component of displacement. Maximum peak-to-peak displacement of the broadband record is approximately 0.4 cm. At frequencies less than 1 Hz, the Wood–Anderson instrument is flat to acceleration, which means that the displacement response rapidly decays. The lower seismogram is multiplied by a factor of 20 for comparison. Standard magnification for these instruments was 2800, which would have slightly exceeded the dynamic range of the photographic recording system for this earthquake.

short distances. Since it uses short-period waves, it also saturates above magnitude 7.0. The application of the body-wave magnitude scale has varied with time – particularly in whether or not the maximum amplitude of the entire P-wave train, rather than just the first few cycles, is measured. Prior to 1964, it was determined using the formula,

$$m_b = \log(A/T) + Q$$

where A is the amplitude, T is the period, and Q is a factor that corrects for distance and depth. Unlike the local magnitude scale, the amplitude measurement for the body-wave magnitude represents the ground motion amplitude after correcting for instrument response, rather than the amplitude as recorded on the instrument. Since the advent of the Worldwide Standardized Seismographic Network (WWSSN), m_b has been determined from the high-frequency channel, for which the response is strongly peaked at a period of 1s.

M_s, the surface-wave magnitude scale, is a measure of the amplitude of surface waves (Rayleigh waves) at a period of 20 s, as measured very far from the event. This magnitude is often quoted by the press as the Richter magnitude. The internationally adopted definition (Vanek *et al.*, 1962) of the surface-wave magnitude scale is

$$M_s = \log(A/T)_{max} + 1.66\log\Delta + 3.3$$

where A is the peak amplitude of vertical ground motion (as for the body-wave magnitude it is corrected for instrument response) and Δ is the angular distance from the earthquake to the seismograph. This definition of the surface-wave magnitude closely resembles the original definition of Gutenberg and Richter (1936), which used a fixed period of 20 s. M_S is usually a better estimate of earthquake size, especially for larger earthquakes, than either m_b or M_L. It only saturates for 'great' earthquakes when M_s is much over 8. For intermediate and deep earthquakes, however, it does a poor job of representing earthquake size because such events do not strongly excite surface waves.

The magnitude scales described above do not tie magnitude to a physical parameter. An obvious physical parameter of interest that is strongly linked to the strength of an earthquake is the total amount of energy radiated in the form of seismic waves. This is often referred to as the radiated seismic energy, or the seismic energy for short. Gutenberg and Richter (1942, 1956) attempted to relate the seismic energy to the surface-wave magnitude. For the seismic energy, E_s, expressed in joules, they found:

$$\log E_s = 1.5M_s + 4.8$$

This relationship assumes a proportionality between the energy radiated as Rayleigh waves at approximately 20 s period and the entire radiated field. As with the surface-wave magnitude scale, this is a reasonable assumption until M_s approaches 8. For larger earthquakes, the effects of fault finiteness start to become important for 20 s Rayleigh waves, and the surface-wave magnitude is no longer representative of the total radiated energy. Nevertheless, this relationship shows that the radiated energy increases very rapidly with magnitude. A unit increase in magnitude corresponds to a factor of 32 ($10^{3/2}$) increase in the seismic energy.

M_0, the seismic moment, is another physical parameter related to the size of an earthquake and has become the parameter that seismologists most often use to describe the size of an earthquake. The seismic moment is a measure of the size of the equivalent force system needed to generate the waves that the earthquake faulting generated. M_0 was first

calculated by Aki (1966) in his study of the Niigata, Japan, earthquake. It is a more robust estimate of the size because it is done, in theory, at infinite period, and radiation pattern effects due to the orientation of faulting are explicitly accounted for when it is calculated. Another advantage it has is that it can be applied to aseismic transients, which by definition do not radiate detectable seismic waves. Seismic moment does not saturate and is directly proportional to the amount of slip across a fault times the fault area. If the faulted area is A, the shear modulus is μ, and the average slip over the faulted area is s, then the seismic moment, M_o, is

$$M_o = \mu A s$$

The MKS unit for seismic moment is newton-meter. Earlier work in seismology typically used dyne-centimeter, the CGS unit, to express seismic moment.

M_w, the moment magnitude scale, is a magnitude scale based directly on the measurement of the seismic moment so it does not saturate. M_w agrees with the M_L and M_s scales before they saturate, but retains the same relationship to seismic moment thereafter (Hanks and Kanamori, 1979). Assuming that the seismic moment is expressed in N m, the moment magnitude relation can be written:

$$\log M_o = 1.5 M_w + 9.05$$

and hence the moment magnitude scale is determined from the seismic moment by the relation,

$$M_w = 2/3 \log M_o - 6.03$$

A unit increase in M_w does not mean a factor-of-10 increase in earthquake size. Rather, the increase in size of the earthquake, as measured by M_0, is about factor of 32 ($10^{3/2}$) for each magnitude unit.

As is apparent from all of the above definitions, magnitude scales do not vary from 1 to 10, but are open ended. In practice, even M_w is limited on the high end both by the strength of the rocks that earthquakes occur in and by the size of the fault. On the low end, M_w is limited by the size of the smallest faults and/or fault patch that can slip unstably. Earthquakes that are recorded by sensitive borehole instruments or *in situ* in deep mines routinely have negative magnitudes. There are many other magnitude scales, most of which are tailored to suit either the practicalities or the diverse goals of earthquake, volcano, and nuclear explosion monitoring. These include magnitude scales derived from measurements as diverse as the duration of the coda of scattered waves following an earthquake, the amplitude of guided waves in the continental crust, and the amplitude of the tsunami that an earthquake generates.

Instrumental seismology began in the last decade of the nineteenth century; however, many earthquakes of interest happened before then. For such earthquakes, seismologists must resort to more subjective and indirect estimates of earthquake magnitude. The most common approach to determining the magnitude of preinstrumental earthquakes is through the interpretation of seismic intensity observations.

Intensity is different from magnitude in that it is not a measure of the intrinsic size of an earthquake; rather, it is a local measure of the strength of shaking as perceived by people or as it affects buildings or natural features. There are many different intensity scales, but the most widely used is probably the modified Mercalli intensity (MMI) scale (Richter, 1958). Traditionally intensities are denoted with Roman numerals, and the MMI scale ranges from I to XII. **Table 1** contains the MMI scale as abridged by Bolt (1993). Earthquake magnitude can be determined from the intensity distribution in several ways. Some methods use the area encompassed by intensities of a certain level (Toppozada, 1975), whereas others use individual intensity observations directly (Bakun and Wentworth, 1997). Toppozada *et al.* (1981) used intensity distributions to estimate the magnitudes of California earthquakes from the preinstrumental period. Because the stronger intensities are direct measures of damage in earthquakes, there have been attempts to relate intensities to measures of ground motion, like peak ground acceleration (PGA), that are thought to correlate strongly with damage (Gutenberg and Richter, 1956). More recently, instrumental recordings of ground motion have been used to estimate 'instrumental intensities'. Wald *et al.* (1999) find the following relation between PGA and the MMI level, I_{mm}:

$$I_{mm} = 3.66 \log(\mathrm{PGA}) - 1.66 \; (\sigma = 1.08)$$

Intensity data has also been used as a proxy to study strong ground motion from significant earthquakes where such data are lacking, such as for the 1906 San Francisco, CA, earthquake (Boatwright and Bundock, 2005). The intensity map of the 1886 Charleston, South Carolina, earthquake is shown in **Figure 4**.

Table 1 The modified Mercalli intensity scale

I.	Not felt except by a very few under especially favorable circumstances.
II.	Felt only by a few persons at rest, especially on upper floors of buildings. Delicately suspended objects may swing.
III.	Felt quite noticeably indoors, especially on upper floors of buildings, but many people do not recognize it as an earthquake. Standing automobiles may rock slightly. Vibration like passing of truck.
IV.	During the day felt indoors by many, outdoors by few. At night some awakened. Dishes, windows, doors disturbed; walls make creaking sound. Sensation like heavy truck striking building. Standing automobiles rocked noticeably.
V.	Felt by nearly everyone, many awakened. Some dishes, windows, and so on broken; cracked plaster in a few places; unstable objects overturned. Disturbances of trees, poles, and other tall objects sometimes noticed. Pendulum clocks may stop.
VI.	Felt by all, many frightened and run outdoors. Some heavy furniture moved; a few instances of fallen plaster and damaged chimneys. Damage slight.
VII.	Everybody runs outdoors. Damage negligible in buildings of good design and construction; slight to moderate in well-built ordinary structures; considerable in poorly built or badly designed structures; some chimneys broken. Noticed by persons driving cars.
VIII.	Damage slight in specially designed structures; considerable in ordinary substantial buildings with partial collapse; great in poorly built structures. Panel walls thrown out of frame structures. Fall of chimneys, factory stack, columns, monuments, walls. Heavy furniture overturned. Sand and mud ejected in small amounts. Changes in well water. Persons driving cars disturbed.
IX.	Damage considerable in specially designed structures; well-designed frame structures thrown out of plumb; great in substantial buildings, with partial collapse. Buildings shifted off foundations. Ground cracked conspicuously. Underground pipes broken.
X.	Some well-built wooden structures destroyed; most masonry and frame structures destroyed with foundations; ground badly cracked. Rails bent. Landslides considerable from river banks and steep slopes. Shifted sand and mud. Water splashed, slopped over banks.
XI.	Few, if any, (masonry) structures remain standing. Bridges destroyed. Broad fissures in ground. Underground pipelines completely out of service. Earth slumps and land slips in soft ground. Rails bent greatly.
XII.	Damage total. Waves seen on ground surface. Lines of sight and level distorted. Objects thrown into the air.

4.01.2.2 Earthquakes in the Context of Plate Tectonics

Scientists have long recognized that some portions of the Earth are characterized by frequent earthquake activity, while others have essentially no earthquake activity whatsoever. **Figure 5** shows the first global map of earthquake activity (Mallet, 1858), which shows a remarkably accurate view of global seismicity, particularly since it predated the invention of the seismograph by decades.

In 1954, Beno Gutenberg and Charles Richter published the second edition of *Seismicity of the Earth*. To this day, their book provides a remarkably complete description of the locations, depth, and size of earthquakes, as well as the distribution and types of volcanoes, worldwide (**Figure 6**). What the book lacks is a unified model to tie all those observations together, explaining the where and the why of the Earth's earthquake and volcanic activity. The theory of plate tectonics, which was established within a decade following publication of that second edition, provides that context. *Seismicity of the Earth* underscores both just how much was known about the systematics of earthquake and volcanic activity before the theory of plate tectonics was formulated, and what a wide range of phenomena plate tectonics explains.

Earthquakes delineate plate boundaries; the vast majority of earthquakes occur at plate boundaries (**Figures 7** and **8**), and the type of boundary – transcurrent, divergent, and convergent – exerts a profound effect on the nature of earthquake activity on it. Thus, it is natural to organize a discussion of earthquakes and volcanoes by plate boundary type. In this section, we are only able to offer a cursory review of the systematics of earthquakes and volcanoes.

4.01.2.2.1 Transcurrent plate boundaries

Most transcurrent plate boundary earthquakes, in which plates slide past one another horizontally on strike-slip faults, occur on oceanic transform faults deep under the ocean surface. Such oceanic transform fault earthquakes do not pose much hazard to humankind. That and their remoteness from seismic monitoring instruments renders them difficult to study; however, some transform plate boundaries traverse continental crust where they do pose a significant hazard and are more easily monitored.

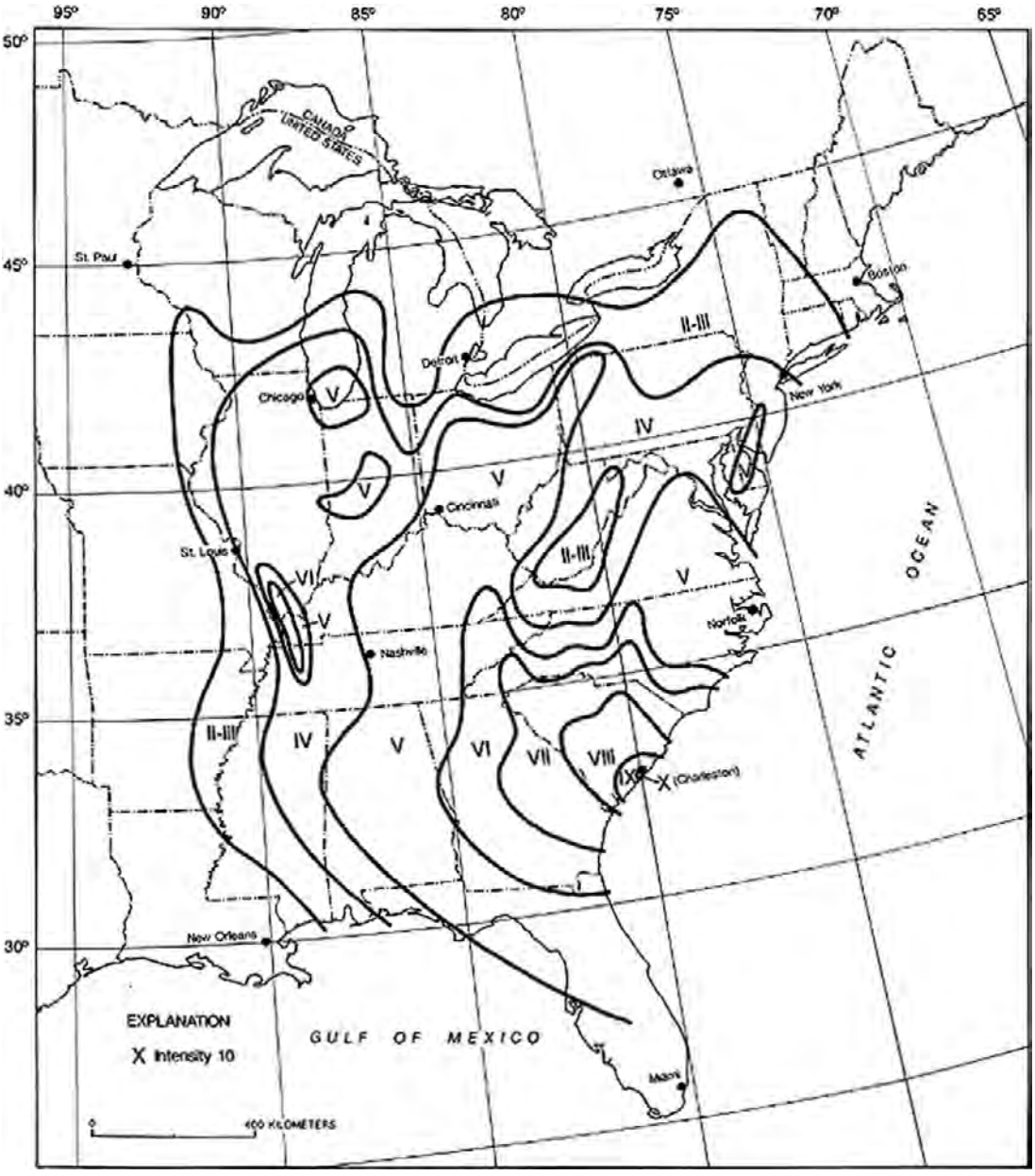


Figure 4 Isoseismal map for the 1886 Charleston, South Carolina, earthquake. The lines enclosing areas of equal intensity are referred to as 'isoseismal' lines. From Stover CW and Coffman JL (1993) Seismicity of the United States, 1568–1989 (Revised). US Geological Survey Professional Paper 1527. United States Goverment Printing Office, Washington.

Transform fault earthquakes, whether in the oceans or on continents, are limited in their size by the depth extent of the seismogenic crust, which is at most a few tens of kilometers, and the fact that transform faults tend to cut through the Earth's crust at angles very close to vertical.

Despite this limitation on size, continental transform fault earthquakes that occur in places such as Turkey, New Zealand, and California can be devastating, and include some of the most infamous earthquakes in history. When they occur in continental crust, the shallow extent of the seismogenic zone means that waves are generated everywhere very close to the Earth's surface and hence have little opportunity to spread out or attenuate, and thus reduce in amplitude, before impacting the built environment. Although strike-slip deformation predominates at transform plate boundaries, such plate

Figure 5 Global distribution of earthquakes as represented by Mallet and Mallet (1858). The principal zones of seismicity are well represented, with the exception of the mid-ocean ridge system, for which the earthquakes were too small and too remote from land to detect. This map preceded instrumental seismology and was based entirely on felt reports of earthquakes. Reproduced from Mallet R and Mallet JW (1858) *The earthquake catalogue of the British Association, with the discussion, curves, and maps, etc.* Transaction of the British Association for the advancement of Science, London.

boundaries are not perfectly parallel to the direction of relative plate motion. For this reason, there are often significant reverse and normal faulting earthquakes at such boundaries too, particularly for continental transform boundaries, which typically involve much more broadly distributed fault systems than oceanic transforms.

4.01.2.2.2 Divergent plate boundaries

Earthquakes at plate boundaries where plates diverge from one another on normal faults have the least societal impact of any type of plate-boundary earthquakes. Most such plate boundaries occur as mid-ocean ridges situated several kilometers beneath the ocean surface. Exceptions to this include Iceland and the Galapagos, where the effects of mid-ocean ridges and hot spots are superimposed, and East Africa, where the divergent plate boundary is not yet mature enough to have separated sufficiently for an ocean to form. Divergent plate boundaries are characterized by effusive basaltic volcanism through which new crust is formed.

A great deal of the deformation at these plate boundaries appears to take place aseismically. The largest normal faulting earthquakes that occur on extensional plate boundaries are small relative to the largest earthquakes at other plate boundaries owing to the fact that the high temperatures in newly formed crust limit the depth extent of the seismogenic crust. Normal faulting predominates at extensional plate boundaries, but extensional rifts are often irregular, and it is not uncommon for strike-slip earthquakes to occur in such settings across local transform faults within a rift. In some cases, these earthquakes are significant.

Mid-ocean ridge volcanism for the most part occurs under several kilometers of ocean. Hence, until the 1950s, the extent and systematics of the mid-ocean ridge system was unknown. Nevertheless, as the source of new oceanic crust, it is a key element of plate tectonics, and the creation of oceanic crust at mid-ocean ridges is the key behavior that distinguishes the theory of plate tectonics from the theory of continental drift. The mid-ocean ridge volcanoes form the longest mountain chain on Earth – some

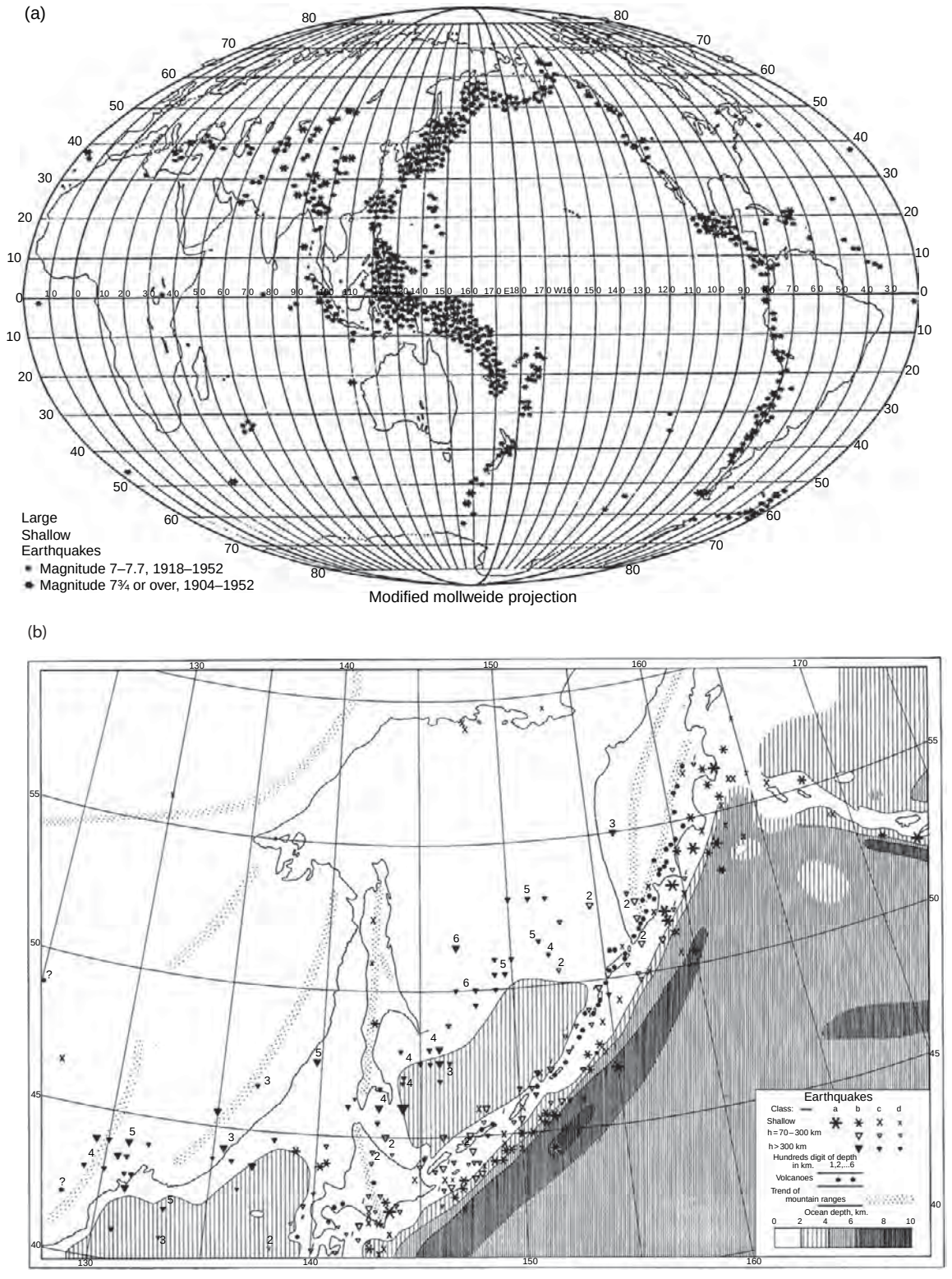


Figure 6 (a) Seismicity map showing worldwide distribution of shallow earthquakes. (b) Close-up of subduction zone seismicity including Honshu, the Kurile Islands, and Kamchatka. The principal features, including trench offshore, increasing depth of earthquakes to the northwest, and volcanoes, were subsequently explained by the theory of plate tectonics. From Gutenberg B and Richter CF (1954) *Seismicity of the Earth and Associated Phenomenon*, 2nd edn. Princeton: Princeton University Press.

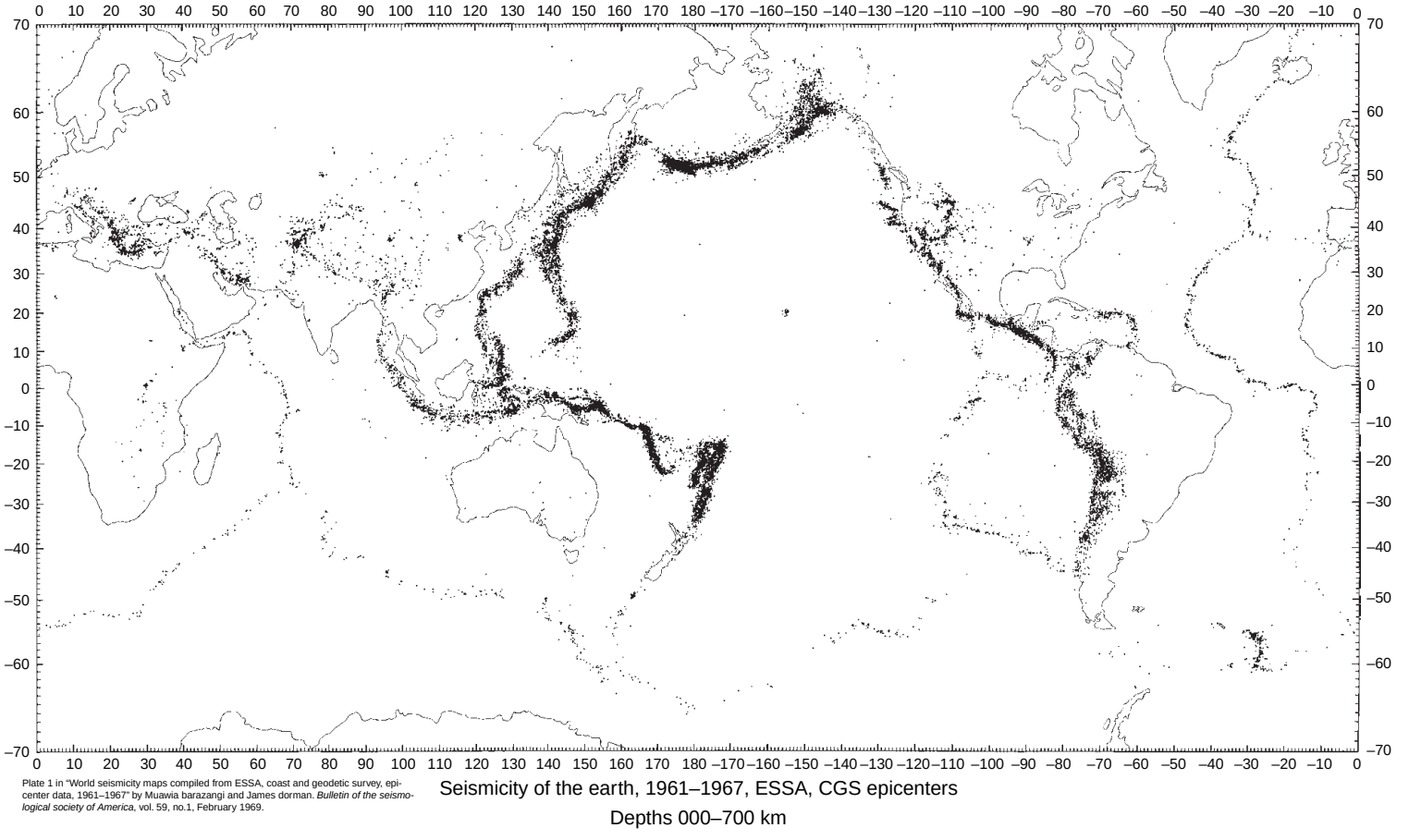


Figure 7 Global seismicity as shown by Barazangi and Dorman (1969) from 1961 through 1967, which comprises 29 553 earthquakes of magnitude greater than 4.0. The narrow bands of seismicity running through the Earth's oceans are readily apparent in this figure.

(a)

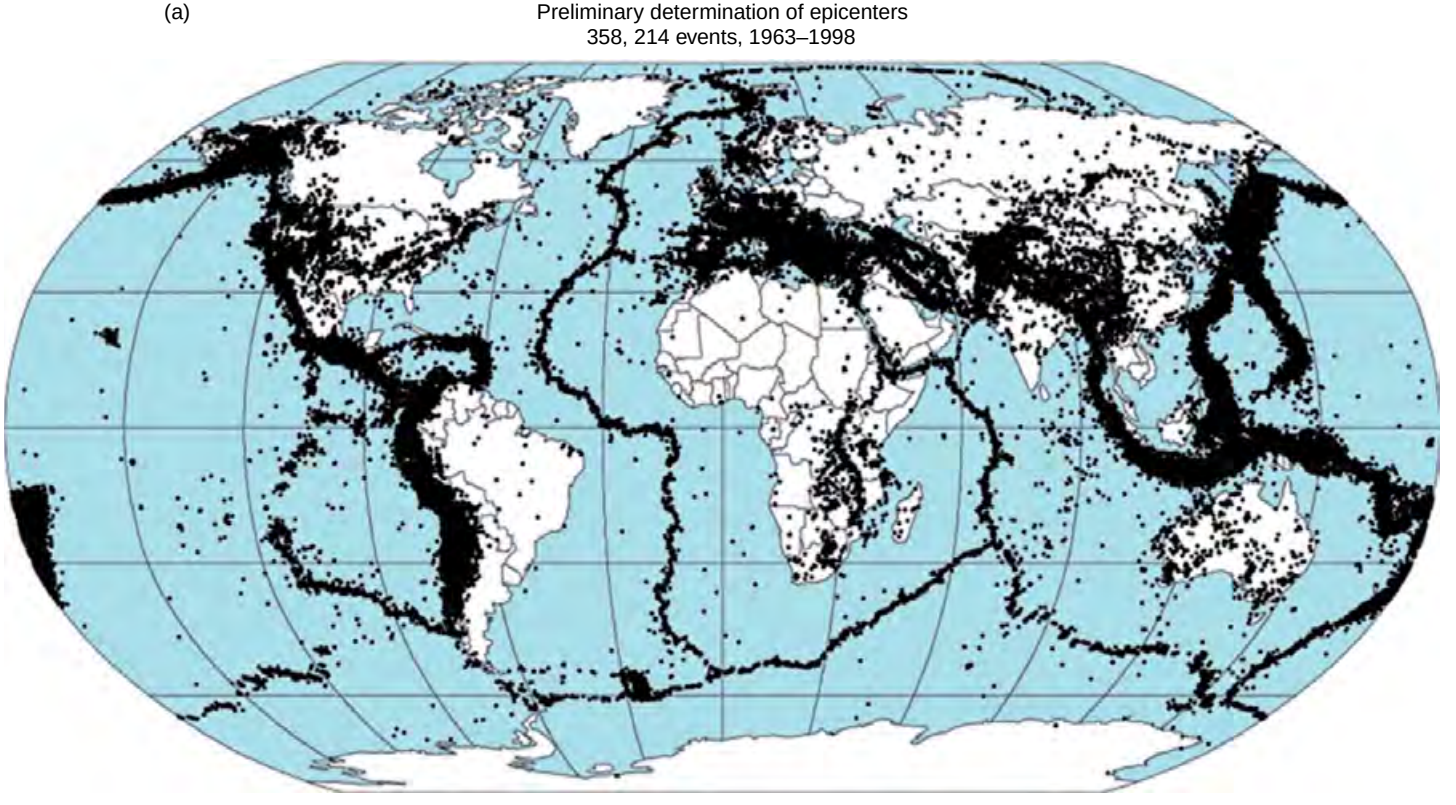


Figure 8 (*Continued*)

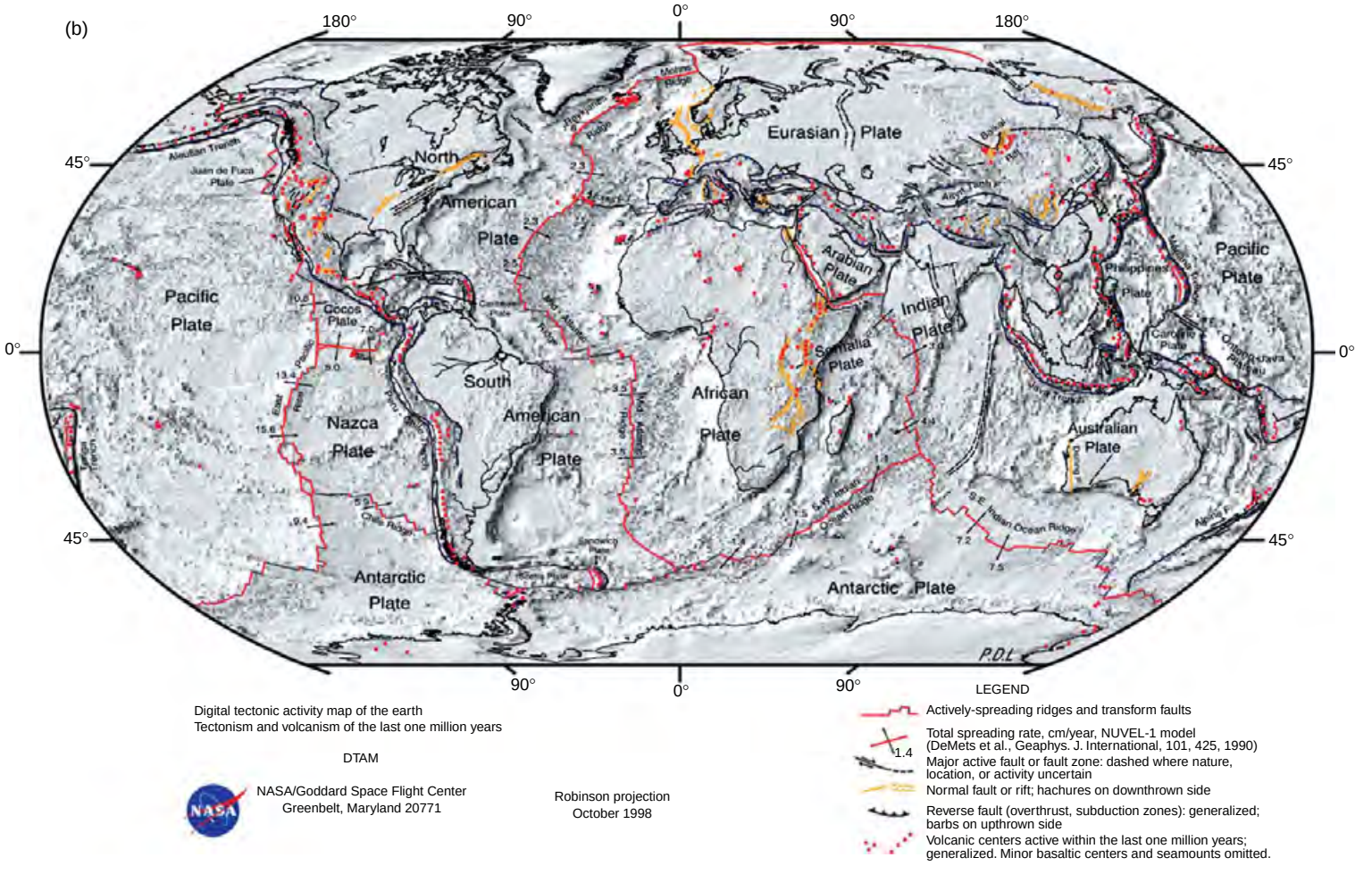


Figure 8 (a) Seismicity of the Earth for a 25 year time interval. Each black dot represents an earthquake and shows how earthquake activity limits global tectonic plate boundaries. From http://denali.gsfc.nasa.gov/dtam/seismic/. (b)Topography/bathymetry and the major tectonic plate boundaries of the Earth for a 25 year time interval. Note the close correspondence between plate boundaries and earthquake activity in (a). From NASA via Wikipedia Commons.

40 000 kilometers in total extent. The hydrothermal plumbing of these magmatic systems is complex, depends on the spreading rate, and gives rise to a range of unusual hydrothermal deep-sea vents. Although eruptions on mid-ocean ridge systems have been inferred from signals detected on hydrophone arrays, they have not yet been directly observed.

4.01.2.2.3 Convergent plate boundaries

By almost any measure, earthquakes at convergent plate boundaries dominate the seismicity of the Earth. At such plate boundaries, all factors work toward the generation of frequent large earthquakes. Relatively cool crust is subducted into the Earth, resulting in reduced temperatures at depth, which increases the depth extent of the seismogenic zone. Moreover, unlike other plate boundaries, the plate interface of some subduction zones traverses the seismogenic zone at a very shallow angle – in places less than 10° from horizontal – meaning that the seismogenic width of the plate boundary in places like Sumatra, Alaska and Chile, can be hundreds of kilometers across. As a result, the great majority of large earthquakes, and all earthquakes much larger than magnitude 8, occur at convergent plate boundaries with reverse faulting mechanisms. The largest recorded earthquakes are all of this type, including the M_w 9.5 1960 Chile earthquake, the M_w 9.3 Alaska earthquake of 1964, the M_w 9.2 Sumatra earthquake of 2004, and the M_w 9.0 Kamchatka earthquake of 1952. These four earthquakes dominate the total seismic activity of the Earth over the past 100 years. Of these largest recorded earthquakes, the Sumatra earthquake is the only one that caused widespread and large-scale devastation, and that was due to the tsunami it generated. There are several factors that mitigate the damaging effects of these earthquakes somewhat. One is the depth of faulting, which is several tens of kilometers. Another is that the strongest shaking in an earthquake is usually in the direction of rupture, which for typical subduction geometries means that the strongest shaking can be expected to occur up-dip and offshore.

A particular kind of earthquake occurs in a region known as the 'outer-rise'. The outer rise is an area of relatively high topography, seaward of the trench, caused by flexural stresses from the downward bending of the subducting plate. Outer-rise earthquakes are thought to occur in response to these localized stresses in the subducting plate (Chapple and Forsyth, 1979). Despite the fact that convergent plate boundaries accommodate compressional deformation, both reverse and extensional normal faulting earthquakes are known to occur in the outer rise. These outer-rise earthquakes can be quite large. The damage due to shaking from them is usually less severe than for earthquakes on the subduction interface because they are more distant from land; however, outer-rise earthquakes can give rise to large and devastating tsunamis, such as in the case of the 1933 M_w 8.4 Sanriku event.

In some subduction zones, such as Japan, there is considerable deformation of the overriding plate as well. In the case of Japan, the upper-plate deformation is largely strike slip and is widely distributed. In cases where the plate convergence direction is highly oblique to the subduction zone, the slip is usually partitioned into trench-normal reverse slip on the subduction megathrust, and trench-parallel strike slip in the overriding plate (Fitch, 1972; Yu *et al.*, 1993). Several explanations have been offered for this partitioning. Savage (1983) suggested that the down-dip edge of the coupled zone would lead to localized horizontal shear arising from the strike-slip component of relative plate motion at that locale. Beck (1983) suggested that shear strain in the overriding plate was controlled by a weak zone corresponding to the volcanic arc. Another explanation is that slip partitioning would occur on distributed fault systems if each fault will have a smaller total area than one large fault. Thus, the total resisting force, which is the product of the frictional strength and the fault area, will be lower and hence less force will be required to cause fault slip (McCaffrey, 1992). In subduction zones with oblique convergence, the strike-slip component of motion can be accommodated with less total frictional resistance if it occurs on a vertical strike-slip fault. Of course, this comes at the expense of having slip in more earthquakes. The tradeoff between the two factors has been explored by Michael (1990), who determined simple geometries for which partitioning was energetically favorable.

Intermediate- and deep-focus earthquakes, usually defined as earthquakes at depths of 70–300 km and 300–680 km, respectively, occur exclusively at convergent plate boundaries within subducting lithosphere. **Figure 9** shows the global depth distribution of intermediate- and deep-focus earthquakes. In some cases, the subduction zone that gave rise to the lithosphere at depth may no longer be active at the surface, such as for the Spanish deep-focus earthquake of 1954 (Chung and Kanamori, 1976). The mechanism of intermediate- and deep-focus earthquakes is a topic of continuing debate as Earth materials at such great pressures and temperatures are expected to deform

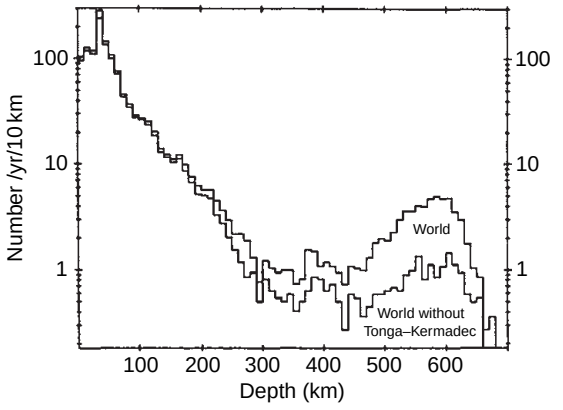


Figure 9 Depth distribution of earthquakes for the period 1/1964-2/1986 with m_b 5 and larger. There is a steady decrease of seismicity up to approximately 300 km depth followed by a marked increase in seismicity below 500 km, with a maximum at approximately 600 km depth. The distribution is similar, whether or not the prolific source of deep earthquakes in the Tonga–Kermadec region is included. From Frohlich C (1989) The nature of deep focus earthquakes. *Annual Review of Earth and Planetary Sciences* 17: 227–254.

plastically; however, several mechanisms have been put forward to explain their occurrence.

Deep earthquakes are most numerous in places such as Tonga, where cold material is rapidly subducted to great depth. The Tonga–Kermadec region accounts for nearly two-thirds of the seismicity of greater than 300 km depth. The largest known deep-focus earthquake is the M_w 8.3 Bolivia event of 1994 (Wu *et al.*, 1995). A thorough investigation of deep earthquakes can be found in Chapter 4.11. **Figure 10** shows the depth distribution of earthquakes in the Tonga subduction zone.

The spatial distribution of intermediate-depth earthquakes is less well understood than that of deeper earthquakes, though plate bending and subduction zone coupling appear to play important roles (Astiz *et al.*, 1988). The largest known intermediate-depth earthquake is the magnitude 8.1 Ryuku Islands event of 1911. There is a great deal of variation in the behavior of intermediate-depth earthquakes. For both intermediate- and deep-focus earthquakes, we lack the kinematic plate tectonic boundary conditions that enable us to constrain the long-term probabilities of interplate earthquakes; hence long-term forecasting of such earthquakes is by necessity based solely on statistical empiricism.

Some convergent plate boundaries involve continental crust on both plates. This occurs most notably along the Alpide belt, which includes the collision of Africa, Arabia, and India with southern Eurasia. Although in places, such as the eastern Mediterranean, this involves the subduction of small relict ocean basins, for the most part continental collision zones entail the collision of relatively buoyant and weak continental crust that is not easily subducted. Deformation in these regions can extend many hundreds of kilometers from the main collision zone. In East Asia, the buoyancy of continental crust leads to the rise of the Himalayas and the Tibetan Plateau, as well as large strike-slip fault systems that allow buoyant crust to effectively flow away from the convergence zone. These zones give rise to some of the largest known strike-slip earthquakes, such as the great M_w 8.3 1905 Mongolia earthquake (Okal, 1992), and include major strike-slip fault zones, such as the North Anatolian Fault in Turkey.

As with earthquakes, much of the systematics of subduction zone volcanism was recognized before the advent of plate tectonics. Island arc volcanoes occur over subducting or recently subducted slabs. The melt that drives island arc volcanism is created when phase changes in the subducting slab give off water and other volatiles. When these come in contact with the surrounding mantle, the melting point is lowered, which in turn creates a buoyant melt that ascends and results in explosive volcanism. For detailed explanation of volcanic processes, *see* Chapter 4.12.

4.01.2.2.4 Intraplate earthquakes

Intraplate earthquakes occur far from plate boundaries and thus are not explained by the theory of plate tectonics. We distinguish here between earthquakes that occur within broad zones of plate boundary deformation, such as those in the Basin and Range province of western North America, and truly intraplate earthquakes, such as those in the New Madrid seismic zone of central North America. Earthquakes that fall into a sort of gray area between plate boundary earthquakes and true intraplate earthquakes include earthquakes in central and southern India as well as earthquakes in the East African Rift system. Locales of true intraplate earthquakes include central Australia, Europe's Rhinegraben, central Brazil, and large parts of oceanic crust under the Indian and Pacific Oceans.

Why do intraplate earthquakes occur? About half of all intraplate earthquakes are observed to occur in failed continental rifts (Johnston and Kanter, 1990). These are likely to be weaker than the un-rifted continental crust around them, so it is reasonable to conclude that such intraplate earthquakes exploit this relative weakness. This does not appear to provide a

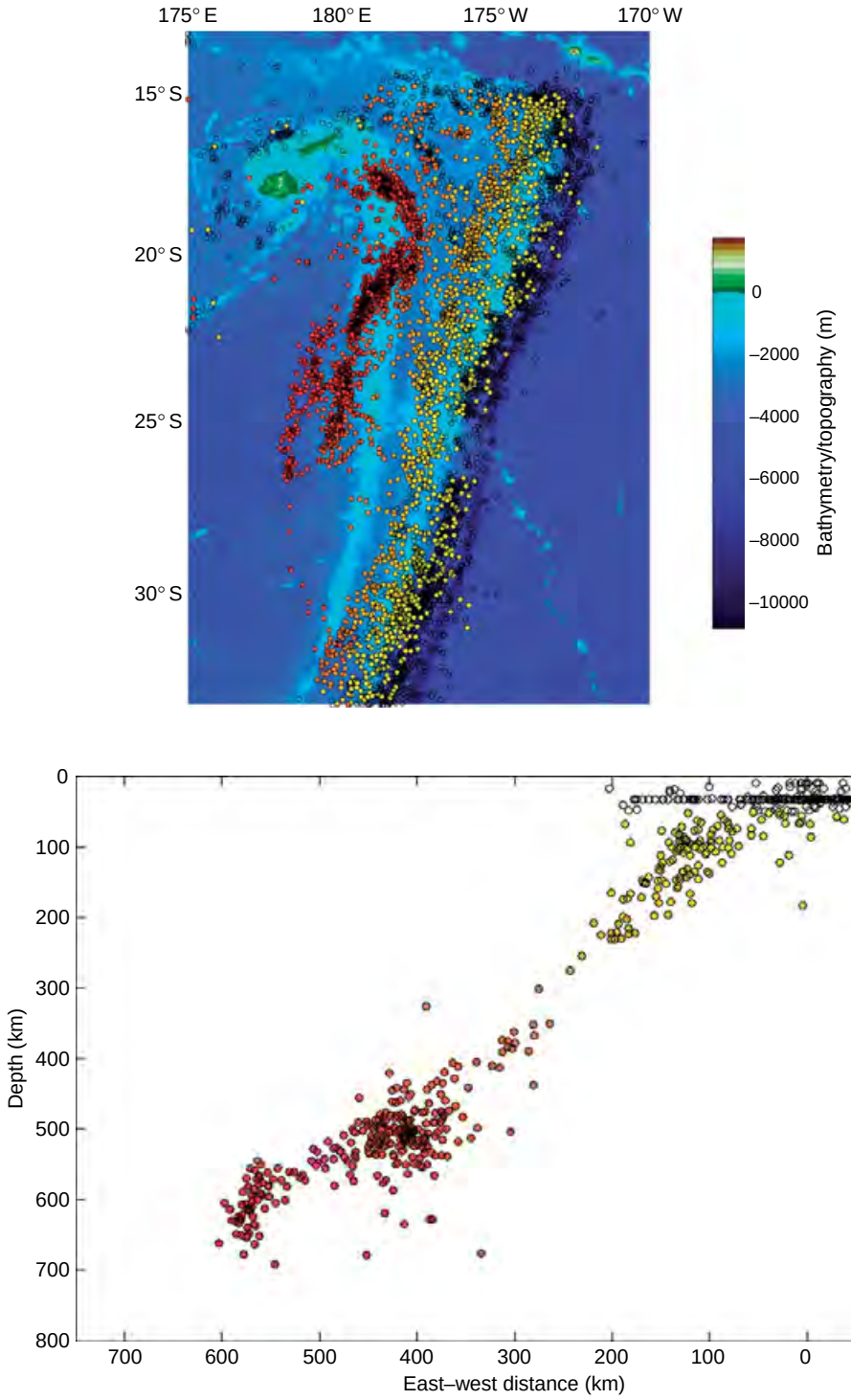


Figure 10 Top panel shows $M > 5$ seismicity (from CISN catalog for 1964–2006) in the Tonga region, color-coded by depth of hypocenter. Tonga trench is dark blue lineation in bathymetry. Open symbols represent earthquakes < 50 km depth, yellow 50–100 km depth, and increasing red tone denotes increasing depth up to nearly 700 km. Horizontal lineation at 15 km is default depth used when depth is indeterminate. Lower panel shows E–W cross section at 24–27°S latitude. Rapid subduction of old lithosphere results in the world's most prolific source of intermediate and deep-focus earthquakes.

universal explanation, however, as intraplate earthquakes in places such as Australia and oceanic plates do not occur on such structures. Intraplate earthquakes respond to the ambient stress field, which supports the possibility that they are occurring due to relative lack of strength, rather than excess of stress. Other explanations for the driving mechanism behind intraplate earthquakes include stresses induced by the emplacement or sinking of plutons (Kane, 1977), post-glacial rebound (Arvidsson, 1996), and stress concentration due to the presence of a weak zone in otherwise strong crust (Sbar and Sykes, 1973; Campbell, 1978).

Regardless of their mechanism, it is clear that intraplate earthquakes cannot account for a very large fraction of lithospheric deformation. If they did, then global plate motion models would fail closure tests. Strain rates due to intraplate deformation have an upper bound of 4×10^{-10} yr^{-1} (Gordon, 1998). Johnston and Kanter (1990) found that the cumulative seismic moment of intraplate earthquakes was about 0.5% of that of the Earth's total. Despite this small percentage, intraplate earthquakes do pose a significant seismic hazard and are occasionally quite large, such as in the 1886 Charleston earthquake, with an estimated moment magnitude of 7.6 or the New Madrid earthquake sequence of 1811–12, which by some estimates included three earthquakes of moment magnitude approximately equal to 8.

4.01.2.2.5 Hot spot volcanism

As with intraplate earthquakes, there are important types of volcanoes for which plate tectonics does not have a ready explanation. These are known as hot spot volcanoes – areas of profuse volcanism that persist for tens of millions of years. Examples of hot spot volcanoes include some of the most famous and well-studied volcanic systems on Earth, such as Hawaii, Kerguellen, and Iceland. Yellowstone is a classic example of a continental hot spot volcano. Hotspot volcanism in the Earth's oceans tends to be characterized by effusive basaltic eruptions, whereas the interaction of basaltic magma with continental crust can lead to large-scale explosive rhyolitic volcanism. The ultimate magmatic source of hot spot volcanism remains a matter of some debate (Montelli *et al.*, 2004), with some evidence pointing to a source in the lowermost mantle while other evidence favors a shallower origin.

4.01.2.3 The Largest and Deadliest Earthquakes

The 12 largest instrumentally recorded earthquakes (updated from Kanamori, 1977) are listed in **Table 2** All of them are plate-boundary earthquakes in subduction zones. The largest instrumentally recorded earthquake was the 1960 Chile earthquake, which had a moment magnitude of 9.5 (instrumentation adequate for calculating magnitudes has only been around since about 1900). The 1964 Good Friday earthquake in Alaska had a surface-wave magnitude of 8.4 and a moment magnitude of 9.2. Put another way, the Alaska earthquake was 20 times larger than would be estimated from the surface-wave magnitude.

The 1989 Loma Prieta earthquake had a moment magnitude of 6.9. For comparison, the 1960 Chile earthquake involved nearly 7000 times the seismic moment of the Loma Prieta earthquake. The 1964 Alaska earthquake had nearly 3000 times the seismic moment of Loma Prieta. The 1906 San Francisco earthquake had a moment magnitude of approximately 7.9 (this is more uncertain than the other estimates due to the sparse instrumentation at the time); hence the seismic moment for that earthquake was about 30 times higher than for the Loma Prieta earthquake. There is no evidence for events larger than the 1960 earthquake in the geologic record, and the extreme length, width, and slip of the 1960 Chile earthquake are impressive; however, given that the instrumental era in seismology spans only slightly more than 100 years, yet larger earthquakes have to be considered possible.

Magnitude scales quantify the physical size and geologic impact of earthquakes, but their impact on human affairs depends strongly on where, when, and how they occur. **Table 3** lists the 10 deadliest

Table 2 The 12 largest earthquakes

Year	Location	M_w
1960	Chile	9.5
1964	Alaska	9.2
2004	Sumatra	9.2
1952	Kamchatka	9.0
1906	Ecuador	8.8
1965	Aleutians	8.7
2005	Indonesia	8.6
1957	Aleutians	8.6
1950	Assam	8.6
1963	Kuriles	8.5
1923	Kamchatka	8.5
1938	Banda Sea	8.5

Table 3 The 10 deadliest earthquakes

Year	Location	Fatalities
1556	Shanshi, China	830 000
2004	Sumatra, Indonesia	283 106
1976	Tangshan, China	255 000
1138	Aleppo, Syria	230 000
856	Damghan, Iran	200 000
1927	Tsinghai, China	200 000
1920	Gansu, China	200 000
893	Ardabil, Iran	150 000
1923	Kanto, Japan	143 000
1948	Turkmenistan	110 000

Credit: US Geological Survey, Department of Interior/USGS.

Figure 11 Print depicting the firestorm that followed the 1923 Kanto, Japan, earthquake. The earthquake and firestorm that followed devastated the Tokyo–Yokohama metropolitan area and are estimated to have killed 143 000 people. From http://www.rekihaku.ac.jp/e-rekihaku/109/pic19.jpg. Courtesy of National Museum of Japanese History.

Figure 12 Photo showing nearly complete devastation in the Lunan district of the city of Tangshan from the 1976 Tangshan, China, earthquake (Huixian *et al.*, 2002). Tangshan was a major industrial center with a population of 1 million before the earthquake, which occurred directly under the city at 3.42 a.m. About 93% of residential buildings and 78% of commercial buildings in the city were destroyed (Yong *et al.*, 1988). The officially reported death toll is 255 000, but estimates are as high as 655 000. Reproduced with Housner GW, Lili X, and Duxin H (2002), with permission.

earthquakes. In the case of the 2004 Sumatra earthquake, nearly all of the fatalities were caused by the tsunami that the earthquake generated. In the case of the 1923 Kanto earthquake, the firestorm that followed played a major role (**Figure 11**). In the 1556 Shansi and 1920 Gansu earthquakes, the widespread collapse of caves used as dwellings and carved into poorly consolidated loess, magnified the losses tremendously.

It is difficult to assess how reliable these numbers are. Certainly for earthquakes that occurred hundreds of years ago, the reliability can be questioned. Even in the modern age, the devastation and confusion following such massive catastrophes, and the need to focus on recovery rather than counting fatalities in their immediate aftermath, means that the true death tolls for these events will never be known accurately. For example, given the widespread and extreme devastation caused by the 2004 Indian Ocean earthquake and tsunami, the number of people lost in that catastrophe will never be known. Moreover, in some cases, political sensitivities may have influenced the reporting of casualties, contributing to the uncertainty. For example, unofficial estimates of the death toll of the 1976 Tangshan earthquake (**Figure 12**) range as high as 655 000. Following the 1999 Izmit, Turkey, earthquake (not listed), authorities simply stopped counting when the death toll reached 18 000. Some estimates of the number of fatalities are twice that high. Whatever their precise toll, one thing is clear: earthquakes can be devastating. At their worst, they rank among the most extreme, sudden catastrophes humankind has to cope with.

Due to the rapid growth of Earth's population, the concentration of that growth on the earthquake prone areas of the Pacific rim, and the urbanization accompanying that growth, vulnerability to earthquakes is growing. This is true despite steady and considerable progress in earthquake-resistant engineering. Bilham (1995) suggests that an earthquake within the next 30

years with over a million fatalities is not out of the question.

Loss of life represents the most important vulnerability to earthquakes, but it is not the only one. Economic losses in earthquakes can be grave as well. For example, by some estimates (Risk Management Solutions Inc., 1995) a repeat of the 1923 Kanto, Japan, earthquake would likely result in direct economic losses of over 2 trillion dollars (44–70% of Japan's GDP) – a financial event that would be without historical precedent, and which would have far-reaching, global, consequences.

4.01.2.4 Historic and Prehistoric Earthquakes

For the purposes of discussion, we define 'historic' to denote an earthquake for which there are written accounts, but that occurred before the dawn of instrumental seismology in the last decade of the nineteenth century, so that there is no instrumental record of ground motion. We use the term 'prehistoric' to denote an earthquake that occurred so long ago that there is no written record of it.

Some earthquakes have had a significant impact on the course of history. Crushed skeletons in ruins of the ancient world provide compelling evidence that destructive earthquakes occurred in the cities of Troy, Meggido, and Jericho. They may be the inspiration for the Old Testament story about the fall of the walls of Jericho. Meggido (Armageddon) is thought to have been the inspiration for the biblical story of the apocalypse (Nur and Ron, 1997). The 1755 Lisbon earthquake made a deep impression on philosophers in eighteenth-century Europe. The earthquake occurred on a Catholic holiday, and destroyed many churches and much of the city through a combination of strong shaking, fire, and tsunami. The Lisbon earthquake (**Figure 13**) coincides in time with, and may have helped bring about, the decline of Portugal as a major world power. Some have argued that the destruction and chaos brought about by the devastating 1923 Kanto earthquake led to fervent Japanese nationalism that helped fuel World War II. The expropriation of international relief funds for recovery from the 1972 Managua, Nicaragua, earthquake is thought to have helped bring about the communist takeover of that country shortly afterward. The devastation wrought by the 1976 Tangshan earthquake was followed within months by the death of Mao Tse Tung and the end of the Cultural Revolution. For a more detailed discussion on ancient earthquakes and the role of earthquakes in history, *see* Chapter 4.20.

Figure 13 Copper engraving depicting the 1755 Lisbon, Portugal, earthquake of 1 Nov. 1755 with fire ravaging the city and a tsunami destroying ships and the waterfront. Original in: Museu da Cidade, Lisbon. Reproduced in: O Terramoto de 1755, Testamunhos Britanicos – The Lisbon earthquake of 1755, British Accounts. Lisbon: British Historical Society of Portugal, 1990.

Prehistoric earthquakes are in some cases studied using archeology. It is far more common, however, to study them using geology. This field of earthquake science has come to be known as paleoseismology. The field of paleoseismology allows geologists to address questions, such as whether or not a fault is active and what the nature of earthquake recurrence is, that are difficult or even impossible to address by other means. For a detailed discussion of paleoseismology, *see* Chapter 4.19.

4.01.2.5 Earthquake Size Distribution

Small earthquakes are much more common than large earthquakes. This statement is usually quantified in seismology using the frequency–magnitude relation – that is, an equation that describes the relative frequency of large versus small earthquakes. The frequency–magnitude relation is widely observed to follow Gutenberg–Richter statistics, named for the same seismologists who developed the magnitude scale. If $N(M)$ is the number of earthquakes of magnitude M or larger, then frequency–magnitude relation can be expressed as (Gutenberg and Richter, 1956)

$$\log N(M) = a - bM$$

Two examples of earthquake populations that follow such a frequency–magnitude relation are shown in **Figure 14**. This linearity in semi-log space describes a power law, with the a-value (intercept)

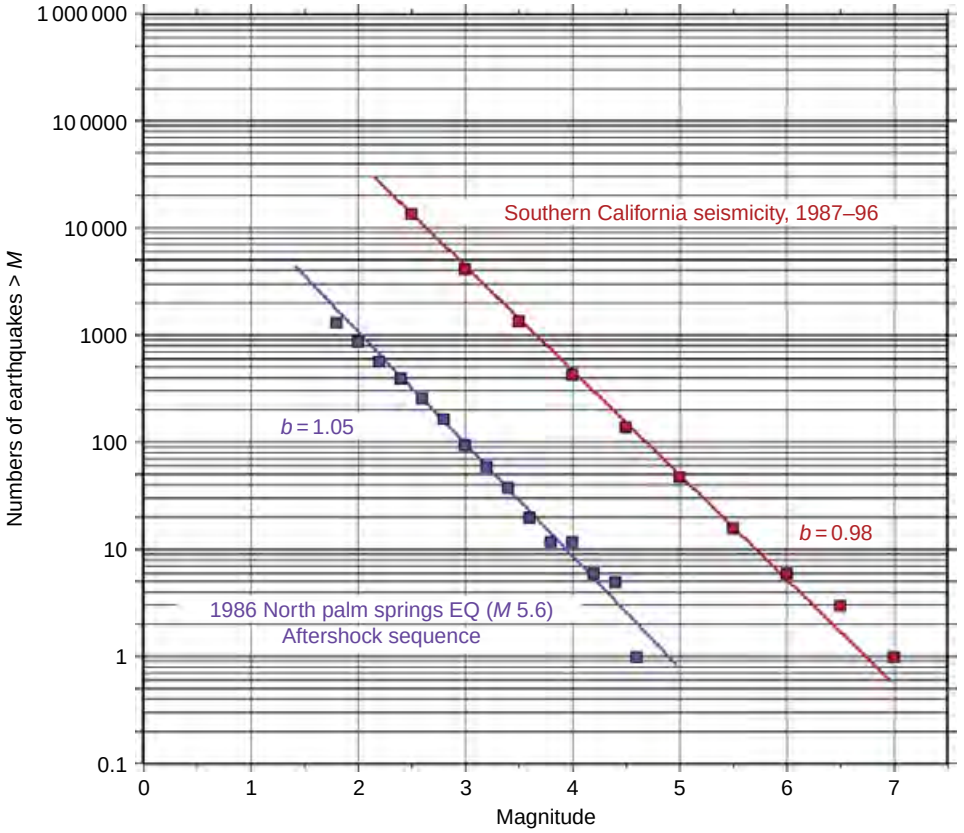


Figure 14 Fit of the cumulative form of the Gutenberg–Richter distribution for two earthquake populations: the 1986 North Palm Springs aftershock sequence (mainshock removed) and a decade of seismicity in all of southern California. In both cases, the *b*-value is approximately 1. From http://www.data.scec.org/Module/Pics/s2inset1.gif.

corresponding to the number of magnitude 0 or larger earthquakes in the population considered, and the *b*-value (slope) describing the relative number of large versus small earthquakes. While, the *a*-value is dependent on the population size, and hence has no obvious physical significance, the *b*-value describes the relative frequency of large and small events regardless of the population size. A *b*-value of 1 means that for each unit magnitude increase, there is a factor-of-10 reduction in the number of earthquakes.

For many earthquake populations, including the global seismicity data set, a *b*-value of ~1 describes the relative numbers of large versus small earthquakes. For individual faults, however, the frequency–magnitude relations may fail to describe earthquake populations adequately (Wesnousky, 1994), and to the extent that they do, there is evidence that the *b*-value for an individual fault may decrease systematically with the maturity of a fault system (Stirling *et al.*, 1996). For volcanic systems, *b*-values are typically much higher than for tectonic faults, sometimes reaching values as large as 3 (McNutt, 2005), indicating that the seismicity associated with active magmatic systems can be extraordinarily rich in very small earthquakes.

By combining the definition of the moment magnitude with the frequency–magnitude relation, it is straightforward to demonstrate that the largest earthquakes on a fault system dominate the slip budget. Combining the frequency magnitude relation with the definition of the moment magnitude yields the following proportionality between the cumulative seismic moment as a function of earthquake magnitude, M:

$$\sum M_o(M_w) \sim 10^{(1.5 - b)M_w}$$

From this, it is apparent that if the *b*-value is less than 1.5, the largest earthquakes will dominate the seismic moment budget. For the globally representative value of $b = 1$, the seismic moment contribution

increases by a factor of $10^{0.5}$, or approximately 3.2, with each increase in magnitude. The relationship between magnitude and seismic energy has the same dependence (Gutenberg and Richter, 1942). Thus, as was recognized by Richter in his original paper on earthquake magnitude, nearly all the energy is released by the largest earthquakes. A direct consequence of this is that little earthquakes do not relieve appreciable strain energy in large fault systems and cannot act as a 'safety valve' for large earthquakes.

4.01.2.6 Earthquake Location

An earthquake's hypocenter is defined as the temporal and spatial coordinates where seismic waves are first generated. Earthquake location has long been routine, but there continues to be important progress in developing more precise earthquake locations. The fundamental observations used to estimate earthquake locations are seismic-wave arrival times. Over the last decade, precise earthquake location methods (e.g., Waldhauser and Ellsworth, 2000) coupled with the use of waveform cross correlation on large waveform databases to reduce measurement error (e.g., Schaff *et al.*, 2004) have revolutionized our ability to resolve the fine structure of fault zones (**Figures 15** and **16**).

By using waveform cross correlation on large numbers of similar waveforms, seismologists have been able to reduce the uncertainty in arrival-time measurements. For similar earthquakes, it is straightforward to reduce the measurement error of arrival times to several milliseconds for regional seismic networks. Since the measurement error maps directly into location uncertainty, the result is greatly improved earthquake locations. Arrival-time measurement using cross correlation is not a new idea in seismology (Poupinet and Ellsworth, 1984), but only recently, with the widespread availability of massive waveform data sets and computers capable of analyzing them, has this approach reached its full potential.

The second advance that has led to improved locations is the expansion of the use of joint hypocentral location methods, particularly ones that are designed to minimize the effects of Earth structure on the solutions. These methods focus on precision (i.e., relative earthquake locations) rather than accuracy (absolute earthquake locations). These methods solve for the position of a large population of earthquakes simultaneously, and when used with differential arrival time measurements allow seismologists to reduce the effects of unmodeled velocity structure on arrival times. Through such techniques, which are analogous to adaptive optics in astronomy, it is possible to reduce earthquake location errors dramatically. For the example shown in **Figure 15**, Schaff *et al.* (2002) document a reduction in hypocentral uncertainty by an order of magnitude in their relocations. As with large-scale waveform cross correlation, enhanced computational capabilities have been an essential element in large joint hypocentral determinations.

An intriguing finding based on highly precise earthquake locations is the discovery of slip-parallel 'streaks' of seismicity (Rubin *et al.*, 1999; **Figure 16**). These streaks have been observed in diverse environments, including the Southeast Rift Zone of Kilauea Volcano, Hawaii, as well as on the San Andreas, Calaveras, and Hayward Faults in California. Waldhauser *et al.* (2004) examined several such streaks along the Parkfield segment of the San Andreas Fault and concluded that, in one case, they may mark the boundary between locked and slipping parts of the fault, while in another case they are more easily explained as a result of a discontinuity in geologic structure.

4.01.3 The Earthquake Source

Most of the measures of earthquake size and the determination of earthquake locations that we have discussed in the previous section, with the exception of the seismic moment, do not require a quantitative theory of the earthquake source. In this section, we outline some of the applications of earthquake source theory to seismology. For a much more detailed and complete treatment, *see* Chapter 4.02.

4.01.3.1 Point-Source Parameters

Quantitative studies of the earthquake source usually treat the Earth as an elastic continuum, with the exception of the fault itself, which is assumed to be a discontinuity having infinitesimal width. These are both approximations that are not uniformly valid and must be kept in mind when interpreting source characteristics. The source of nearly all earthquakes is shear slip on faults. To first order, faults represent planar discontinuities within the Earth and the shear slip occurring on them is parallel to that fault surface. Equation [51] of Chapter 4.02 expresses the displacement seismogram at a point, x, as a function of time, t,

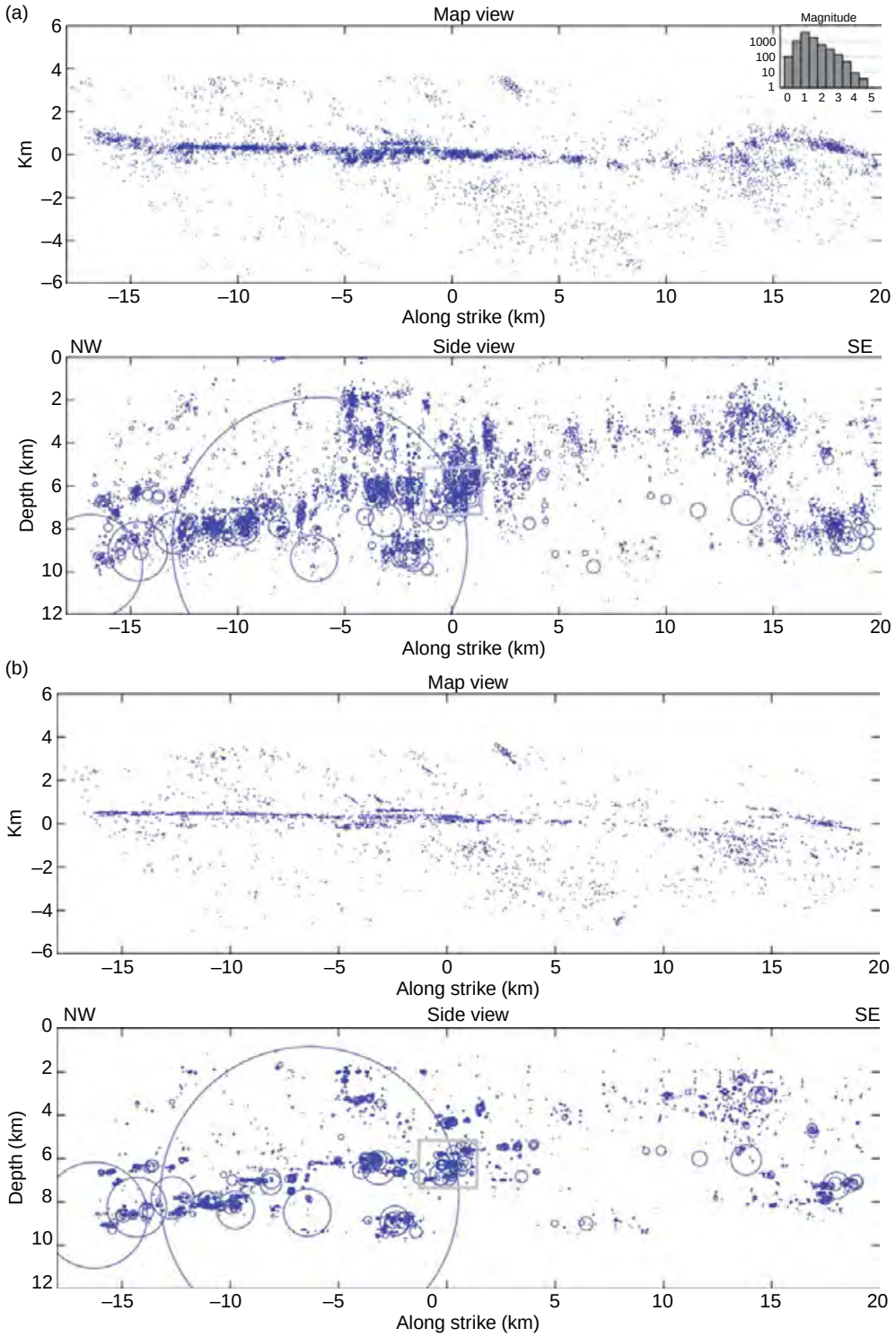


Figure 15 Seismicity on the Calaveras fault before (a) and after (b) double-difference earthquake relocation using cross correlation-based arrival-time measurements (Schaff *et al.*, 2002). Both map and side views are shown for the two cases, which show nearly 8000 earthquakes. Earthquakes are shown as circular sources assuming constant stress drop. The improved locations are evident in the narrow width of fault structures at depth and in the large number of repeating earthquakes found on this fault. Such improved locations allow testing of diverse hypotheses regarding earthquake fault mechanics. From Schaff DP, Bokelmann GHR, Beroza GC, Waldhauser F, and Ellsworth WL (2002) High resolution image of Calaveras Fault seismicity. *Journal of Geophysical Research* 107(B9): 2186 (doi:10.1029/2001JB000633).

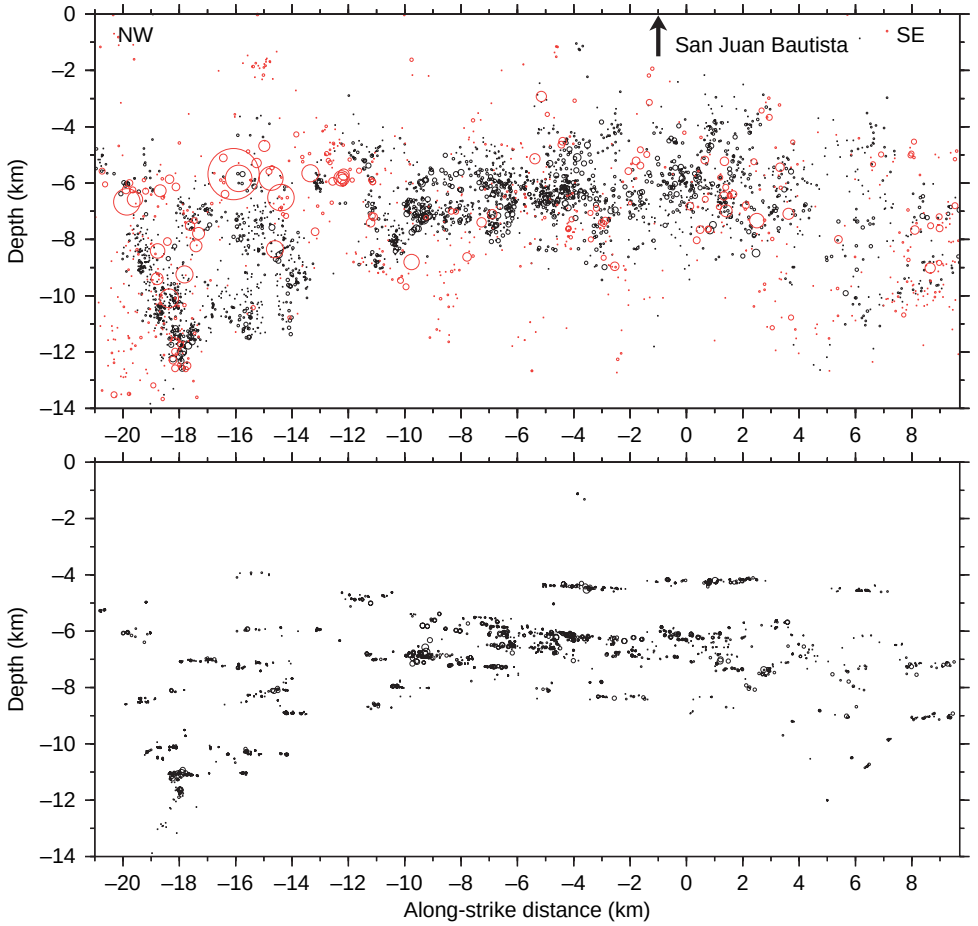


Figure 16 Network locations of microearthquakes on a vertical cross section along the San Andreas Fault near San Juan Bautista (Rubin *et al.*, 1999) are shown on the upper panel. Red circles in upper panel represent larger events that were not relocated because waveforms were clipped. Circles denote estimated slip area for larger events. Lower panel shows same events after relocation. The horizontal seismicity streaks, which were previously obscured by location errors, emerge in the relocated catalog.

as an integral over the fault surface and time of the fault slip history, $\Delta u_j(x_0,\ \tau)$:

$$u_i(x,\ t) = \int_0^t \int_{S_{x_0}} \mu(x_0)\Delta u_j(x_0,\ \tau)G_{ij,k}(x,\ t|x_0,\ \tau) \times n_k(x_0)\mathrm{d}^2x_0\mathrm{d}\tau$$

The other terms in the integral – the shear modulus, μ, the Green's function gradient, $G_{ij,k}$, and the fault plane normal vector, n_k – together represent the traction across the fault surface due to an impulsive point source acting at the receiver.

If the wavelengths considered are much larger than the rupture dimension, and if we record seismic waves sufficiently far away from the source, then we can make the point-source approximation. In this case, variations in the Green's function, G, with position on the fault, x_0, are negligible and the integration above becomes trivial. Other approximations include a high-frequency approximation, wherein only the far-field $(1/r)$ terms in the Green's function make important contributions to the seismogram. This approximation is valid once the receiver is several wavelengths from the source $(\omega r/\beta >> 1)$ and allows the integral above to be simplified since only terms that can be calculated using ray theory need be retained. An intermediate assumption, analogous to the Fraunhofer approximation in optics, is that the Green's function varies over the fault plane, but only in phase. This leads to a simplification of the integral above into one in which the G is outside the integral and results in an expression for the far-field pulse shape, which represents the dependence of the far-field wave amplitude with time. The

development of the theory and approximations outlined above are covered in greater detail in Chapter 4.02.

4.01.3.2 Sense of Faulting from First Motions

The polarity of the first arriving P waves can be used to determine a focal mechanism, which is a representation that plots the polarity of waves as a function of takeoff angle from the source. For a shear-slip source, the P waves will have four lobes, two of which have compressional first-arriving P waves and two of which have dilatational first-arriving P waves (**Figure 17**). At 45° from these lobes are two planes for which the P-wave amplitude is nodal. One of these planes is the true fault plane of the earthquake; the other is known as the 'auxiliary' plane. From the polarity of the P waves, it is impossible to distinguish these two planes from one another. Indeed, in the point-source approximation, the entire seismogram, not just the first-motion polarity, is identical for slip on either one of these planes. This leads to a fundamental fault-plane/nodal-plane ambiguity. Without independent information, such as geologic information on the fault orientation, alignment of earthquake hypocenters, or finite-source effects in the waveforms (i.e., a breakdown of the point-source assumption), it is impossible to tell the two planes apart. The sense of faulting as determined using the P-wave radiation pattern using earthquake focal mechanism solutions provided a key test of the theory of plate tectonics in which the sense of motion on oceanic transform faults was found to be consistent with the plate tectonic prediction (Sykes, 1967).

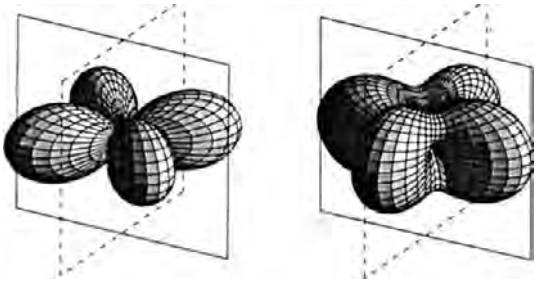


Figure 17 Radiation patterns for a strike-slip fault showing the amplitude of P waves on the left and the amplitude of S waves on the right. The four lobed P-wave radiation pattern has two compressional and two dilatational lobes. The fault plane and auxiliary plane, for which P-wave amplitude is nodal, separate the four lobes. The lobes of the radiation pattern are separated by quadrants in the usual depiction of the focal mechanism. From Julian BR, Miller AD, and Foulger GR (1998) Non-double-couple earthquakes 1. Theory. *Reviews of Geophysics* 36: 525–549.

4.01.3.3 Moment-Tensor Representation

A more general representation of a point source of seismic waves is provided by the seismic moment tensor. The moment tensor can be used to represent the seismic source as a pair of force couples – the so-called double-couple solution, which is the point-source equivalent force distribution for a dislocation source. The determination of seismic moment tensors for earthquakes is now routine such that there exist catalogs of moment-tensor solutions at both the regional and global scale. Perhaps the best known of these is the Harvard centroid moment-tensor (CMT) catalog (Dziewonski *et al.*, 1981) and its successor, the Global CMT Project, which contains moment-tensor solutions for over 20 000 earthquakes worldwide (**Figure 18**). Similar catalogs of earthquake mechanisms, whether based on first-motion focal mechanisms or waveform-based moment-tensor inversions, are available for thousands of locally recorded earthquakes where local seismic networks provide the data to make such work possible.

A strength of the moment-tensor representation is that it can be used to describe sources other than shear

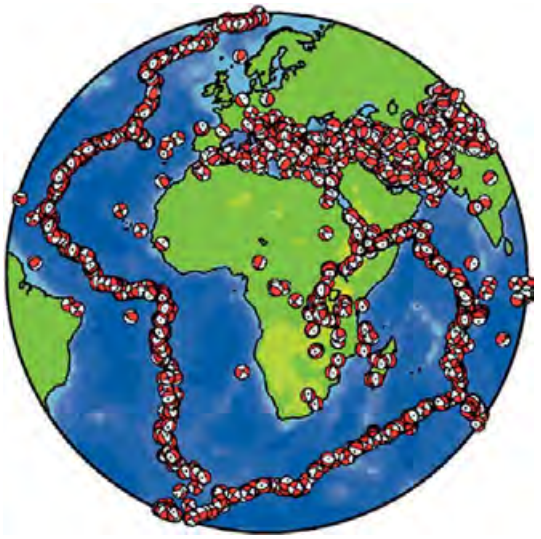


Figure 18 First-motion focal mechanism representation of CMT solutions from the Harvard catalog for earthquakes around the continent of Africa for the 25 year period (1976–2000). Taken from *Living on an Active Earth: Perspectives on Earthquake Science*, National Academy Press, 2003.

slip on faults. Sources with isotropic (volumetric), or deviatoric, but not planar, shear slip have been studied using seismic waves, most often in volcanic environments. Single-force solutions have also been applied to describe the sources of seismic waves for large landslides, such as the one that initiated the 1980 eruption of Mt. St. Helens (Kanamori and Given, 1982). For more detail on moment tensors and their application to the study of earthquakes, *see* Chapters 4.02 and 4.16.

4.01.3.4 Seismic Energy

A point-source parameter that is important, but for which estimates are far from routine, is the radiated seismic energy. The radiated energy is typically calculated by measuring the energy flux at a point, or points, where the wavefield of an earthquake is recorded, and from that inferring the total energy flux in all directions from the earthquake source. The radiated energy is broadband, with most of the energy being radiated at frequencies higher than a frequency known as the 'corner' frequency (Singh and Ordaz, 1994).

Seismic sources can be characterized by an overall length scale. Waves of sufficiently low frequency will have corresponding wavelengths that are much larger than that source dimension. Waves of sufficiently high frequency, on the other hand, will have corresponding wavelengths that are much smaller than that of source dimension. In this latter case, the amplitude of seismic waves will suffer from interference effects, and hence have smaller amplitudes in the far field. The crossover frequency, at which wavelengths are comparable to the seismic source dimension, is known as the corner frequency. The corner frequency can be used to infer the characteristic size of the seismic source.

Strictly speaking, because the corner frequency defines the frequency at which finite-source effects start to become important, the seismic energy is not truly a point-source parameter. As a scalar measure of earthquake strength, however, it is natural to consider it as such. Obtaining the necessary bandwidth with sufficient signal-to-noise ratio, and correcting for propagation effects over that wide bandwidth, is what makes the radiated energy so difficult to calculate reliably (Ide and Beroza, 2001). The study of seismic energy, its determination and its implications for the physics of earthquakes, is the subject of a recently published AGU monograph.

4.01.3.5 Extended-Source Models of Earthquakes

The success seismologists have had in cataloging basic point-source earthquake parameters, particularly location, origin time, moment tensor, and seismic moment, contrasts with attempts to estimate higher-order source parameters. This is natural in that these extended source measurements involve either spatial or temporal dimensions, that is, slip duration, directivity, stress drop, moment-rate function, that can be difficult to resolve. Methods to recover these parameters remain an area of active research.

Studies of the earthquake source are sometimes divided into 'kinematic' and 'dynamic' studies. The adjective 'kinematic' is used in seismology to describe models of an earthquake that estimate source parameters without explicit consideration of the failure mechanism on the causative fault. This is not necessarily a negative property. With our incomplete knowledge of fault mechanics, it is likely that dynamic models, which seek to incorporate what we think we know about earthquake physics, may be incorrectly biased by those assumptions. The hope is that observations of earthquake kinematics will inform dynamic models of fault behavior and that dynamic models will motivate kinematic modeling that tests dynamic hypotheses. We start our discussion of kinematic source models with point-source models.

4.01.3.5.1 Kinematic source models

The moment-tensor representation that is used to describe earthquakes in the point-source approximation can be extended to finite source using an expansion in higher-order moments (Backus and Mulcahy, 1976). McGuire *et al.* (2002) applied this formalism to recover the degree-2 moments of large earthquakes recorded at teleseismic distances. They found that in 80% of the earthquakes studied, rupture was unilateral (rupture initiated at one end of the rupture zone) rather than bilateral (rupture initiated near the center of the rupture zone). For the most part, seismologists have not applied the formalism of higher-order moments to the study of the earthquake source. This may be due to a desire to image the source directly as the spatial and temporal evolution of slip on a fault.

A great deal of work has gone into imaging earthquake rupture directly as slip on a fault (or faults) using seismic waves. For some earthquakes, important additional constraints have been provided by geodetic measurements or field measurements of

surface rupture. The chapter by Ide details approaches taken for imaging earthquake rupture.

As shown by Ide *et al.* (2005), the nonuniqueness of the fault-slip inverse problem is extensive. When there are insufficient near-source data, such as for the 1999 Izmit, Turkey, earthquake, slip models developed by different investigators bear little resemblance to one another. In contrast, well-recorded earthquakes like the 1999 Chi Chi, Taiwan, earthquake will yield much more robust results. For all but extremely large events (Ishii *et al.*, 2005), extended-source images are difficult to recover without near-source data. As a result, most of the information available on source finiteness derives from well-instrumented areas like Japan and California.

A large earthquake may rupture multiple fault segments. Modeling of strong ground motion recordings of large earthquakes suggests that fault rupture typically propagates at about 80% of the shear-wave velocity, although there is evidence that the rupture velocity can locally exceed the shear-wave velocity (Bouchon *et al.*, 2001; Bouchon and Vallée, 2003; Dunham and Archuleta, 2004). The slip velocity across the fault is much more difficult to constrain, but is thought to be of the order of meters per second in a large event (Heaton, 1990). In addition to constraints on the evolution of slip in earthquakes, it is clear from all studies that have sufficient resolution that slip is strongly variable with position on the fault (Mai and Beroza, 2002). **Figure 19** shows

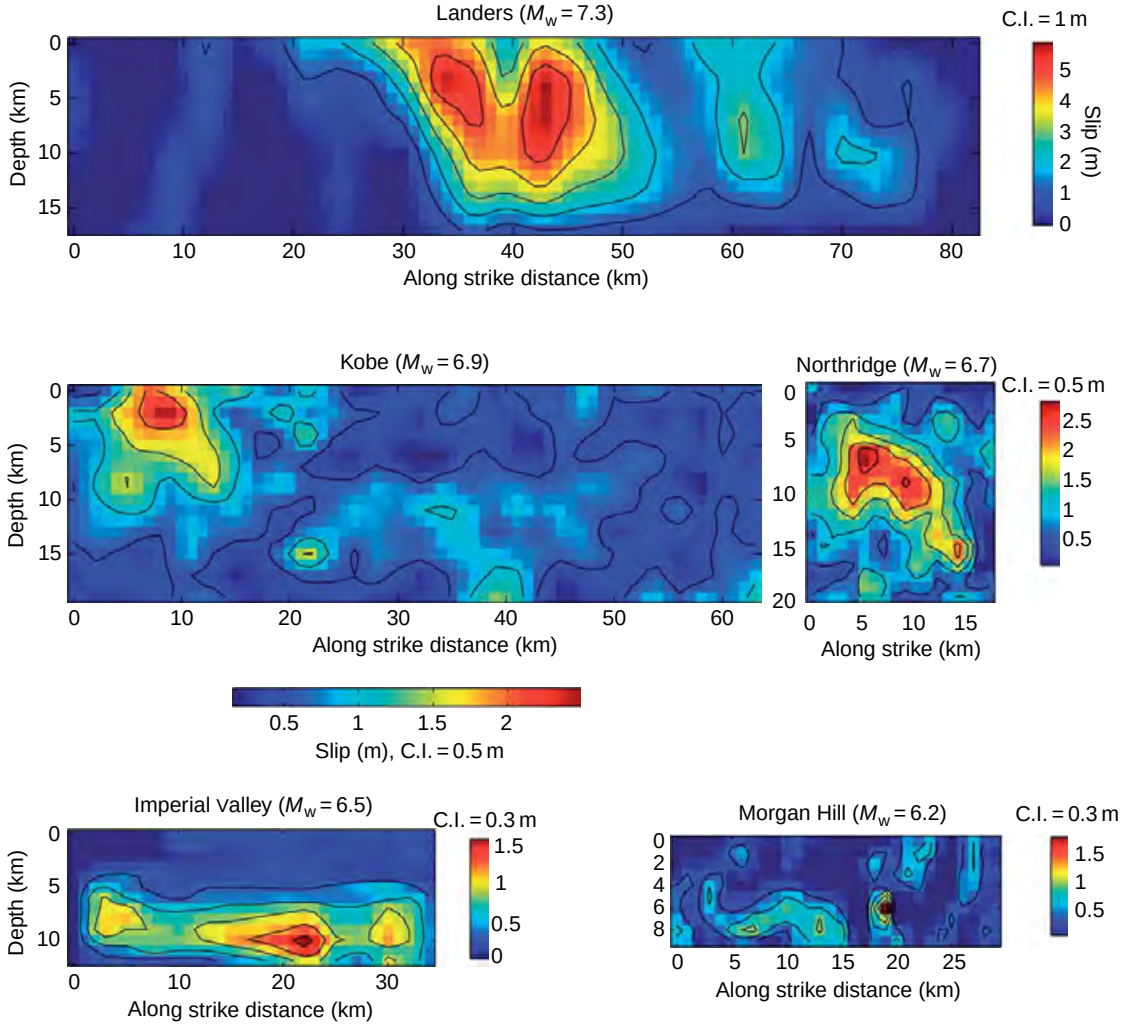


Figure 19 Side view of slip models for five different earthquakes, including the 1992 Landers, CA, the 1995 Kobe, Japan, the 1994 Northridge, CA, the 1979 Imperial Valley, CA and the 1984 Morgan Hill, CA earthquakes. In each case, slip is strongly variable over the fault surface. From Mai PM and Beroza GC (2002) A spatial random-field model to characterize complexity in earthquake slip. *Journal of Geophysical Research* 107(B11): 2308 (doi:10.1029/2001JB000588).

examples of variable slip models for five different earthquakes.

4.01.3.5.2 Dynamic source models

Dynamic source models differ from kinematic models in that they explicitly consider the stress associated with fault slip and attempt to model stress, slip, and the evolution of rupture using a failure criterion. In some cases, the evolution of the rupture is prescribed, and the relevant physical parameters are calculated. Quasi-dynamic rupture models (Quin, 1990) fall within this category. In others, the fault is allowed to evolve based on the initial conditions and a failure criterion for the fault surface. A dynamic model such as this is referred to as a 'spontaneous' model, because, once initiated, the rupture propagates spontaneously as long as the energy balance allows it to be self-sustaining (Kostrov and Das, 1988). A third class of dynamic rupture models does not assume the fault plane *a priori*, but instead allows it to evolve based on a failure criterion for the medium. This is potentially quite important because as rupture speed increases, out-of-plane stresses come to dominate stresses in the direction of rupture, and hence favor nonplanar faulting.

Dynamic faulting models are typically based on fracture mechanics – an approach that was originally developed for tensile opening-mode cracks. In ideal elastic–brittle crack models, the crack tip separates unslipped from slipping fault surface. Stress in the unslipped region immediately ahead of the crack tip is infinite, but it is an integrable singularity such that the amount of energy expended at the crack tip is finite. Such a singularity is unrealizable in the real Earth, so models have been developed that require a breakdown zone ahead of the crack tip in which shear traction remains finite as the crack tip approaches. For a detailed treatment of the application of fracture mechanics to modeling earthquake dynamics, see Rice (1980) and Chapter 4.03.

Dynamic rupture modeling is challenging for several reasons. The need to keep track of stress during rupture means that the entire past history of faulting within the cone of causality is required to calculate the stress on each point on the fault. The need to simultaneously track the large scale of the entire fault rupture and the small-scale heterogeneity near the crack tip presents computational challenges. If grid size is not sufficiently small, then dynamic rupture models can develop an unphysical grid-size dependence. In the presence of material contrast across the fault or of strong velocity weakening, there is the additional complication of numerical instabilities to deal with. Finally, to model multiple earthquake cycles, steady strain accumulation with time scales of centuries and sudden release with timescales of seconds, must both be resolved.

A complete discussion of dynamic source modeling is well beyond the scope of this chapter. Recent work in this area has addressed the possibility that material contrast across a fault zone might lead to a direction of preferred rupture (Andrews and Ben-Zion, 1997), whether strong velocity weakening occurs during seismic faulting (Perrin *et al.*, 1995), the origin of heterogeneity over many earthquake cycles (Lapusta *et al.*, 2000), and attempts to generate dynamic rupture models that emulate the salient properties of kinematic rupture models as in **Figure 20** (Peyrat *et al.*, 2001).

4.01.3.6 Volcano Seismology

As with earthquakes, seismology provides a key source of information on how volcanoes work. Unlike earthquakes, eruptions on well-instrumented volcanic systems are routinely predicted and seismological monitoring plays a key role in this capability. Volcanic systems are often highly seismic with literally millions of earthquakes. Volcanoes exhibit a rich range of behavior that leads to diverse seismic signals.

Figure 21 shows a suite of long-period earthquakes at several volcanoes.

One signal of particular note is volcanic tremor. Volcanic tremor has a range of behaviors, but it often occurs as ground motion primarily at a discrete set of frequencies. There are several theories for the generation of tremor. A notable example is that due to Julian (1994), who suggests that tremor is generated by a nonlinear interaction coupling between fluid flow and the surrounding elastic material. Among the predictions of this tremor model is that, with increasing flow, tremor that is initially harmonic will undergo a cascade of period doubling as the tremor becomes chaotic. Since that paper was published, period doubling has been observed on several volcanoes including Arenal in Costa Rica (**Figure 22**).

Chapter 4.12 covers volcanic processes and Chapter 4.13 covers volcano seismology in detail.

4.01.4 Slip Behavior

The nature of slip on earthquake faults varies widely. It can occur suddenly, as in earthquakes. It can also

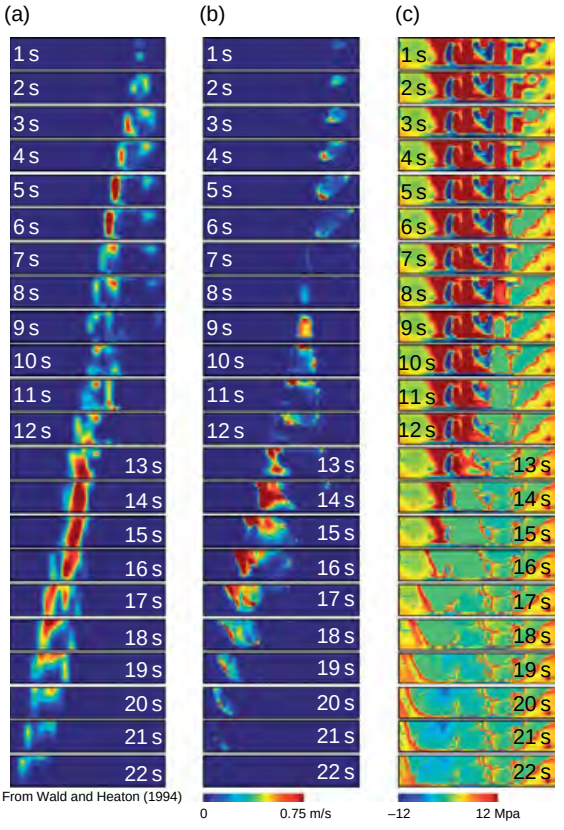


Figure 20 Comparison of kinematic and dynamic rupture models for the 1992 Landers, CA, earthquake. (a) Snapshots of kinematic model from Wald and Heaton (1994); (b) snapshots of dynamic rupture models of Peyrat *et al.* (2001), (c) stress change accompanying dynamic model. Both models fit the long-period strong ground motion data, but predictions of strong ground motion at higher frequencies would vary markedly between the two models. From Peyrat S, Olsen KB, Madariaga R (2001) Dynamic modeling of the 1992 Landers earthquake. *Journal of Geophysical Research* 106: 26 467–26 482.

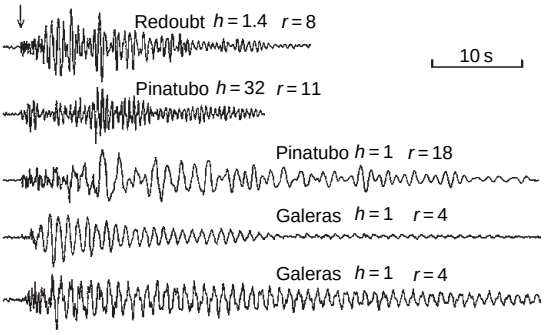


Figure 21 Long-period (LP) earthquakes recorded at Redoubt, Pinatubo, and Galeras volcanoes at source–receiver distance, *r*, and source depth, *h*. These events are thought to reflect pressure fluctuations at depth within a volcano's magmatic system and are useful as a possible precusor to eruption. The lowermost LP earthquake has such a long coda that it verges on harmonic tremor. From Chouet B (1996) Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature* 380: 309–316.

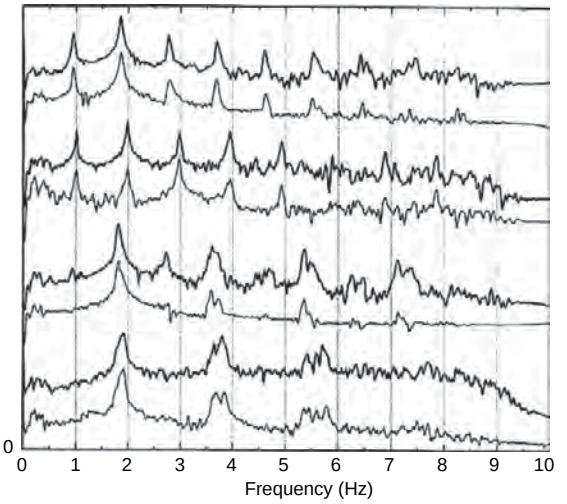


Figure 22 Spectra of velocity seismograms from two seismic stations during four different 30 s time windows during tremor episodes at Arenal Volcano, Costa Rica (Hagerty *et al.*, 2000). The peaks at ~1, 3, 5, and 7 Hz are sometimes present and sometimes absent, which is a signature of period doubling. From Hagerty MT, Schwartz SY, Garces MA, and Protti M (2000) Analysis of seismic and acoustic observations at Arenal volcano, Costa Rica, 1995–1997. *Journal of Volcanology and Geothermal Research* 101: 27–65.

occur quasi-statically, as in faults creep, in which faults slip so slowly that no detectable seismic waves are generated. Documenting the nature of slip behavior and understanding the reasons for its variation is a key goal of earthquake seismology.

4.01.4.1 Brittle Failure

The source of nearly all earthquakes is sudden shear slip across a fault. This sudden slip is sometimes referred to as 'brittle' failure. The domain over which brittle failure occurs is dependent on factors such as mineralogy, pressure, and temperature. Because each of these factors vary strongly with depth, synoptic models of earthquake fault behavior usually characterize the nature of slip as a function of depth.

At very shallow depths – such as the upper few kilometers of continental crust in California – geological materials are weak and unable to store much elastic strain energy. This renders them unable to support sudden slip in earthquakes and is reflected

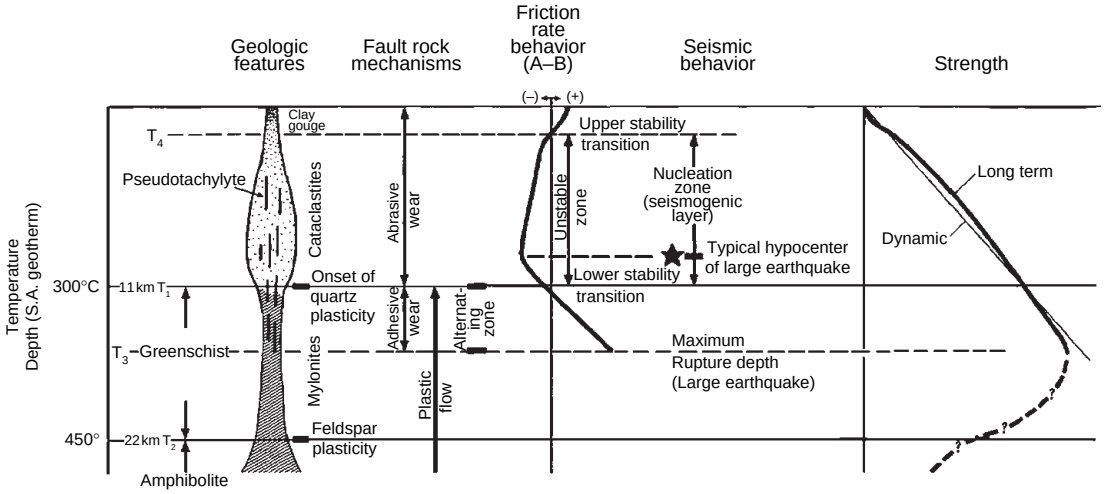


Figure 23 Aspects of shear zones and faulting in the Earth. Increasing temperature with depth is shown on the left axis. Geologic manifestation of deformation is shown next to that together with the underlying deformation mechanism. In the middle is frictional behavior, which closely follows the seismogenic zone. Schematic diagram from Scholz CH (1989) Mechanics of faulting. *Annual Review of Earth and Planetary Sciences* 17: 309–334.

in the depths of the shallowest micro-earthquakes. On the San Andreas Fault in California, for example, the depths of the shallowest micro-earthquakes are several kilometers. The San Andreas Observatory at Depth (SAFOD) project targets some of the shallowest seismicity on the creeping section of the fault by piercing the fault at a depth of approximately 3.2 km. Above this, it is likely that faults slip relatively stably, unless dynamically driven during large earthquake ruptures that originate at greater depths. At somewhat greater depths, rocks have the potential to store elastic strain energy, and the friction on faults is such that they can slip unstably. This depth range is called the seismogenic zone. In California, the seismogenic zone extends to a depth of approximately 15 km. In oceanic crust, where the mineralogy is characterized by higher melting points, the lower boundary of the seismogenic zone is deeper.

The depth of the lowermost earthquakes, where the mode of deformation changes from brittle to ductile, is sometimes termed the brittle–ductile transition, or, less commonly, the seismic–aseismic transition. The depth of this transition is controlled by mineralogy, temperature, and pressure. High strain rates, such as those occurring during an earthquake, can result in unstable faulting at greater depths than earthquakes nucleate. High strain rate during aftershock sequences can have a similar effect (Schaff *et al.*, 2002). **Figure 23** shows a schematic interpretation of the depth dependence of geological processes and their seismological consequences.

Intermediate and deep-focus earthquakes pose an interesting case. The variation with depth of deep earthquakes and the cutoff of the very deepest earthquakes appear to be related to phase changes in the mantle. Because phase changes result in more compact crystal structure, seismologists have long sought a volumetric (isotropic) component to the mechanism of deep-focus earthquakes. They have found, however, that earthquakes at all depths are dominated by deviatoric failure, with deep earthquakes having at most small volumetric components. The fact that deep earthquakes have far fewer aftershocks further distinguishes them from shallow seismicity (Frolich, 1989). In any case, it is difficult to understand how deep earthquakes can undergo brittle failure at temperatures and pressures where rocks are expected to deform plastically. Possible mechanisms behind deep earthquakes – among them thermal runaway, phase transitions, and transformational faulting – are discussed in Chapter 4.11. Regardless of the mechanism, the clear dependence of deep-earthquake behavior on the thermal parameter (**Figure 24**) illustrates the important role played by relatively cold subducted lithosphere at great depth.

4.01.4.2 Creep

If slip on a fault occurs slowly enough, then it is possible that no detectable seismic waves will be generated. Such slip is referred to as 'aseismic'. It has long been known that slip occurs aseismically on some

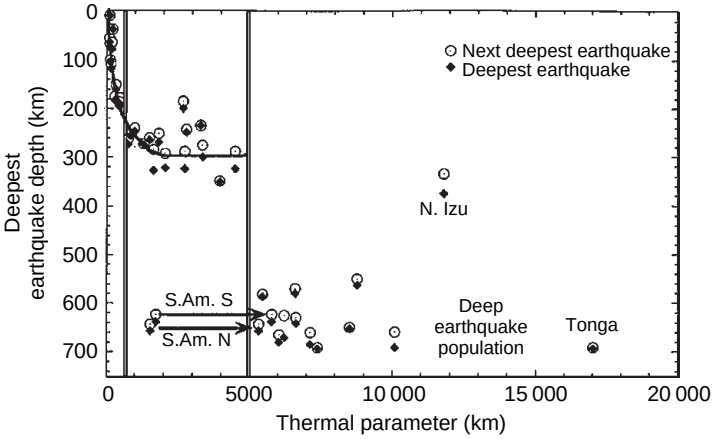


Figure 24 Depth of the deepest earthquake vs the thermal parameter, which is defined by Molnar *et al.* (1979) as the product of plate age and vertical descent rate, for different subduction zones (from Kirby *et al.*, 1996). A threshold thermal parameter of 5000 appears to separate two domains of behavior.

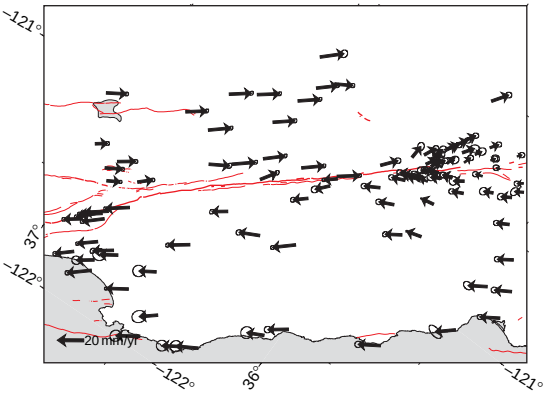


Figure 25 From R. Bürgmann (personal communication), showing faults in red. The San Andreas system of faults crosses the center of the figure. The section between the branching of the fault near San Juan Bautista, on the left of the figure, and Parkfield, the area of dense observations on the right, is known as the creeping section of the San Andreas Fault. GPS velocity vectors, shown with arrows, and 95% confidence ellipses show nearly block motion of the two sides of the fault (i.e., there is very little velocity gradient with distance away from the fault). Farther to the northwest and to the southeast, the velocity gradient indicates that the fault is locked between large earthquakes.

faults, such as the San Andreas Fault in central California (Nason and Tocher, 1970). Such slip has come to be called 'creep' and the section of the San Andreas Fault where most of the fault motion occurs aseismically has come to be known as 'the creeping section'. **Figure 25** shows Global Positioning System (GPS) velocity vectors adjacent to the creeping section of the San Andreas Fault in central California and **Figure 26** shows that creep on this section of the fault at the surface closely matches that inferred over the depth of the seismogenic zone. Once global plate motion models provided predictions about long-term slip rates across plate boundaries, a comparison of earthquake history with seismic slip rates indicated that many plate boundaries have far too little seismic slip (Brune, 1968). Similarly, geodetically measured deformation following significant earthquakes is always observed to be much greater than can be accounted for by slip in aftershocks. The natural inference in both instances is that much of the total fault slip must occur aseismically. An important question about this aseismic slip is whether it is continuous or episodic.

Observations of creep on the San Andreas Fault indicate that creep occurs episodically in a range of sizes that is not unlike the Gutenberg–Richter distribution that applies to ordinary earthquakes (Wesson, 1988). Aseismic transients at depth are sometimes referred to by the oxymoron 'silent earthquakes'. Because they do not radiate seismic waves as ordinary earthquakes do, these events are very difficult to detect; however, they can be detected geodetically using GPS, interferometric radar interferometry, creepmeters, tiltmeters, or strainmeters. As these measuring systems become more widespread, the detection of aseismic transients is becoming more common (**Figure 27**). GPS measurements of crustal deformation over subduction zones has led to the discovery of silent earthquakes as large as $M_W \sim 7.5$ (Kostoglodov *et al.*, 2003).

Intermediate between silent earthquakes and ordinary earthquakes is a class of events referred to as 'slow' earthquakes. Slow earthquakes are identified

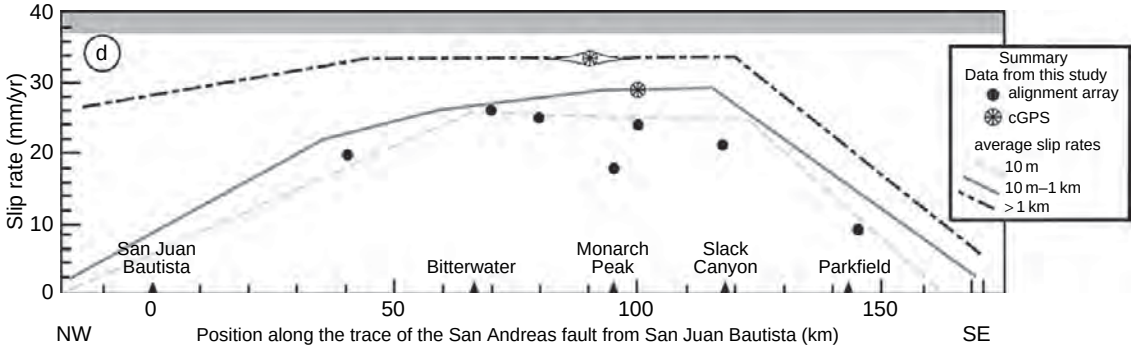


Figure 26 Depiction of the slip rate parallel to the strike of the San Andreas Fault as a function of distance along the fault from NW (left) to SE (right) as inferred from geodetic measurements. Pale dashed line shows rates determined from measurements within 10 m of the fault and can be considered the surface creep rate. Gray line shows measurements taken within 1 km of the fault and can be considered to represent the slip rate in the upper few kilometers. Dark dashed line shows larger-scale measurements, which include GPS measurements. These estimated slip rates reflect slip to greater depths. The similarity of the curves indicates that creep occurs throughout the depth of the seismogenic zone and the comparison with the model of Argus and Gordon (2001) for the motion of the Pacific Plate relative to the Sierra Nevada–Great Valley block indicates that much of that relative motion is taken up by creep at this latitude. The creep rate decays to zero at both the northwest and southwest end of the creeping section. From Titus SJ, DeMets C, and Tikoff B (2006) Thirty-five-year creep rates for the creeping section of the San Andreas Fault and the effects of the 2004 Parkfield earthquake; constraints from alignment arrays, continuous global positioning system, and creepmeters. *Bulletin of the Seismological Society of America* 96: S250–S268 (doi: 10.1785/0120050811).

by having low-frequency seismic waves that are much larger than would be predicted based on their strength as measured by seismic waves at high frequencies. They are thought to be slip events in which the rupture velocity and/or slip velocity are significantly lower than ordinary earthquakes. Oceanic transform faults are one environment in which slow earthquakes are common (Okal and Stewart, 1982). They also occur in subduction zones, with some giving rise to devastating tsunamis (Kanamori, 1972). A new class of slow events (**Figure 28**) has recently been discovered in the accretionary prism of the Nankai trough of Japan (Ito and Obara, 2006). Recently discovered slow events show interesting scaling properties (**Figure 29**).

An earthquake initiates at a point known as the hypocenter. Thereafter, as an earthquake grows, there is a boundary between the part of the fault that is slipping, or has already slipped, and the part of the fault that is as yet unruptured. This boundary is known as the rupture front and the speed at which the rupture front propagates is known as the rupture velocity. The effect of rupture velocity on seismograms was first recognized for the 1952 Kern County, CA, earthquake (Gutenberg, 1955) from the asymmetry of the surface-wave radiation. Seismological estimates of the average rupture velocity for crustal earthquakes are typically in the range of ~70–90% of the S-wave velocity, or on the order of 2.5–3.0 km s^{-1}. There is no evidence that rupture velocity varies with magnitude, with even microearthquakes having rupture velocities in the same range (Frankel *et al.*, 1986). Rupture velocities that are a large fraction of the shear-wave velocity lead to pronounced directivity in strong ground motion, particularly for shear waves, and lead to strong fault-normal pulses of ground motion in the near field of major faults (Somerville *et al.*, 1997).

From the point of view of earthquake dynamics, it is not clear why earthquakes rupture at these velocities. The simplest models based on ideal elastic–brittle fracture theory hold that shear failure ought to accelerate quickly to a limiting velocity that depends on the rupture mode: either the shear-wave velocity for anti-plane rupture or the Rayleigh-wave velocity (about 92% of the shear-wave velocity) for in-plane rupture (Freund, 1990). Fracture mechanics models also predict that out-of-plane stresses will grow as the limiting velocity is approached, which should promote rupture bifurcation and a lower rupture velocity.

Although sub-Rayleigh rupture velocity is found for most earthquakes, there is increasing evidence that supershear rupture, that is, rupture velocities that exceed the shear-wave velocity, may be an important property of large strike-slip earthquakes. Supershear rupture velocities are expected in

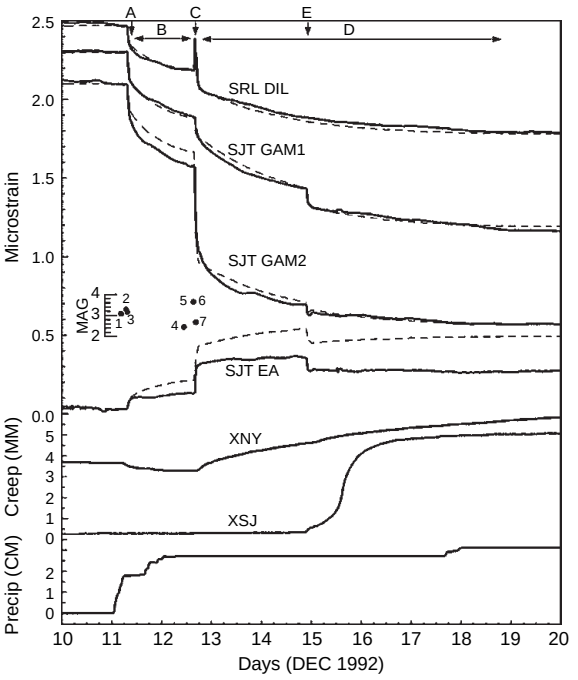


Figure 27 An aseismic transient recorded on the San Andreas Fault near San Juan Bautista. Strain data for stations SRL and SJT, with effects of tides and atmospheric pressure fluctuations removed, are shown as solid lines. Creep meters (XNY and XSJ) are also shown. Labels A–E denote aseismic transients. Model of strain due to aseismic slip on the San Andreas is shown with dashed lines. Small earthquakes occurred at the same time, but are orders of magnitude too small to explain the observed signals. From Linde AT, Gladwin MT, Johnston MJS, and Gwyther RL (1996) A slow earthquake srequence on the San Andreas fault. *Nature* 383: 65–69.

slip-weakening rupture models (Andrews, 1976), and have been observed in laboratory fracture experiments (Rosakis *et al.*, 1999). Rupture velocity has apparently exceeded the S-wave velocity for at least several recent large earthquakes with the 1999 Izmit, Turkey (Bouchon *et al.*, 2001), the 2001 Kunlunshan, Tibet (Bouchon and Vallée, 2003), and the 2002 Denali, AK (Dunham and Archuleta, 2004), earthquakes, all showing characteristics of supershear rupture. If supershear rupture propagation should prove common, it will have important implications for earthquake hazards because the strong ground motion from supershear rupture is qualitatively different from the subshear case (Aagaard and Heaton, 2004). For a full discussion of theoretical, laboratory, and observational aspects of shear rupture velocity, *see* Chapter 4.06.

4.01.4.3 Aseismic Transients and Tremor

Perhaps the most exciting recent discovery in the field of earthquake seismology is that of large silent earthquakes and accompanying deep, nonvolcanic tremor (**Figure 30**). Deep, nonvolcanic tremor was discovered in Japan in the parts of the subduction zone where the Philippine Sea Plate is subducting (Obara, 2002). It occurs at depths where dehydration reactions are thought to liberate water from the subducting plate (Seno and Yamasaki, 2003) and hence it is natural to attribute the genesis of tremor to the movement of fluids at depth. Fluid generation of tremor is also thought to be operative in volcanic tremor that is observed in active

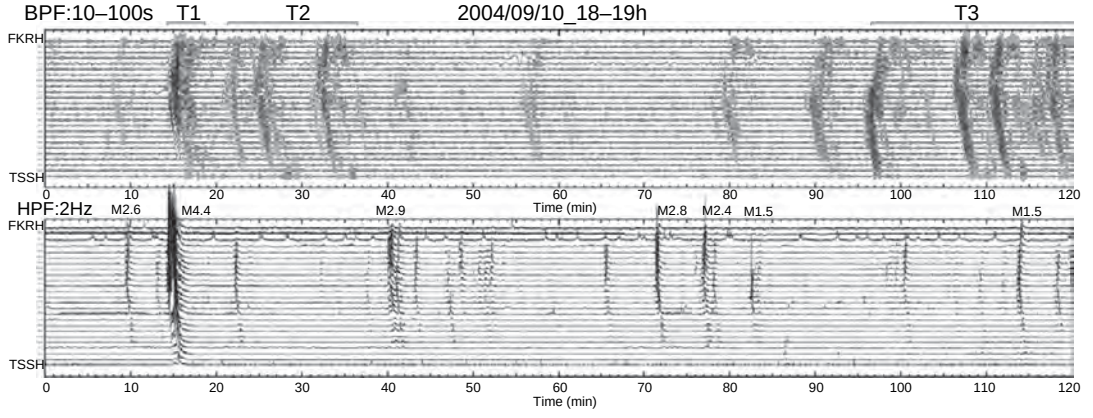


Figure 28 Times series showing very low frequency earthquakes discovered in the accretionary prism of the Nankai trough. Two hours of data are shown. Upper panel shows band-passed (10–100 s) ground motion. Lower panel shows rms ground motion for a high-pass filter with corner at 2 Hz. Labels at the top of the lower panel show magnitudes of earthquakes in the JMA catalog. Events in the upper panel at ~25, ~32, ~55, and ~90 min have no corresponding signature at high frequencies and are identified as very-low-frequency (VLF) earthquakes. From Ito Y and Obara K (2005) Very low frequency earthquakes excited by the 2004 off Kii Peninsula earthquakes: A dynamic deformation process in the large accretionary prism. *Earth Planets Space* 57: 321–326.

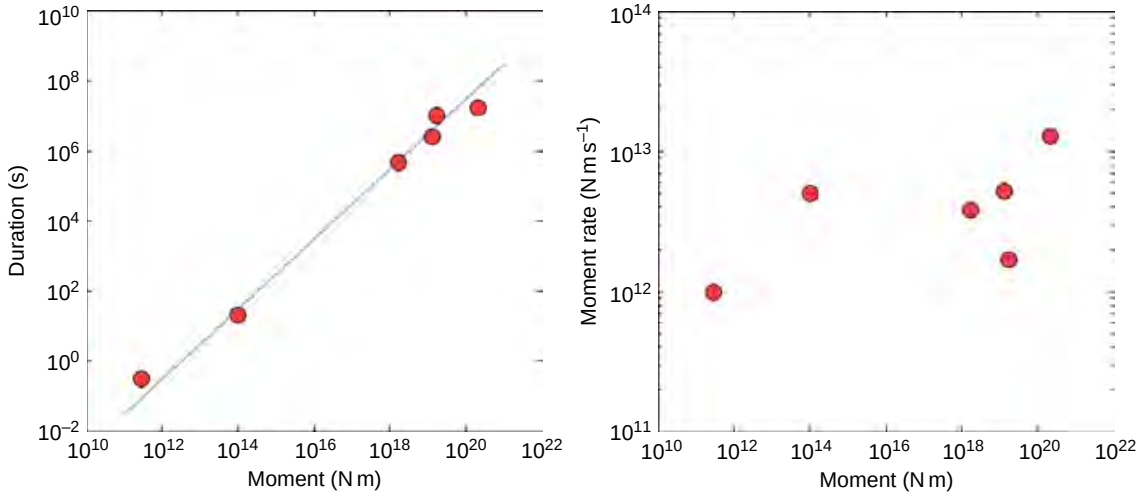


Figure 29 Left panel shows scaling of recently discovered anomalous earthquakes, including (from left to right) low-frequency earthquakes, very-low-frequency earthquakes, and slow slip events (*see* Chapter 4.15). The duration scales linearly with seismic moment over 8 orders of magnitude in slow earthquake size. Panel on the right shows the variation of moment rate with changing moment for the same population. The moment rate is nearly constant over 8 orders of magnitude in seismic moment.

magmatic systems (Chouet, 1988; Julian, 1994). Large silent earthquakes were discovered in Cascadia by Dragert *et al.* (2001) and subsequently found by Miller *et al.* (2002) to be occurring periodically with a recurrence interval of 14 ± 1 months. The remarkable regularity of recurrence of these silent earthquakes stands in apparent stark contrast to the more variable recurrence of more typical large earthquakes.

A near-perfect correlation between tremor and silent earthquakes in Cascadia was discovered by Rodgers and Dragert (2003) and is illustrated in **Figure 31**. A correlation between the two phenomena was subsequently documented in Japan as well (Obara *et al.*, 2004), where silent earthquake recurrence is less regular, but on the order of several months in Eastern Shikoku. The two phenomena are so closely intertwined that they are commonly referred to with an acronym ETS, for episodic tremor and slip. Longer-term aseismic transients, lasting years, have been documented in several areas of Japan including the Bungo Channel (Hirose and Obara, 2005) and the Tokai region (Ozawa *et al.*, 2005). In each of these cases, the aseismic transients appear to arise due to aseismic slip on the plate interface just below the locked seismogenic zone. Given this location, they are certain to accelerate the loading of the source region of large megathrust earthquakes and hence should be capable of triggering such events.

The source mechanism of tremor is not yet completely understood. The location of the tremor source at depths at which metamorphic dehydration reactions are expected and the fact that tremor is observed for Philippine Sea Plate subduction, but not Pacific Plate subduction, in Japan both point to a key role for fluids in the process (Julian, 2002). Fluids may either couple directly into ground motion as they are thought to do in the case of volcanic tremor, or they may act to trigger both phenomena. The discovery by Shelly *et al.* (2006), that low-frequency earthquakes that occur during tremor episodes apparently represent slip on the plate interface, indicates that at least part of the tremor signal arises from shear slip. For more details on aseismic transients, *see* Chapter 4.15.

4.01.5 Physics of the Earthquake Source

In this introductory chapter, we have focused on the phenomenology of the earthquake source as it is studied using seismology. For the most part, we have not focused on the physical mechanisms underlying earthquake behavior. There is much about these physical mechanisms that we have yet to understand. In this section, we attempt to tie together the physics behind the observations.

4.01.5.1 Friction

Most earthquakes occur on pre-existing faults. The friction acting across these faults provides the

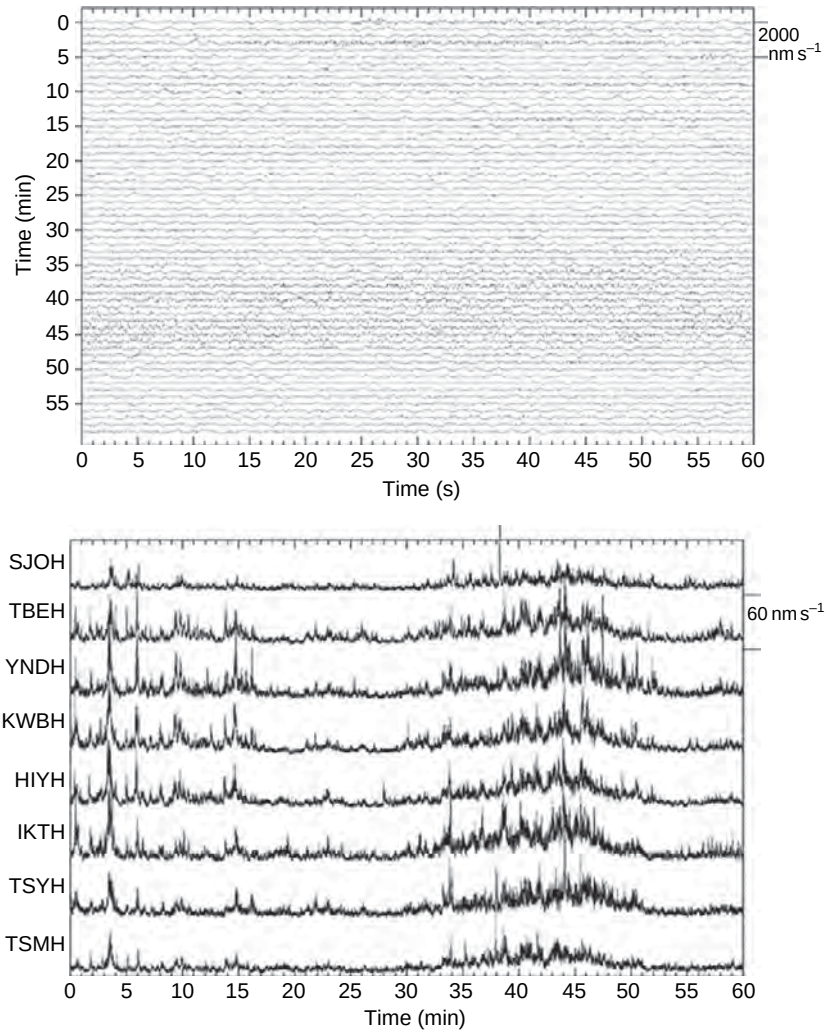


Figure 30 An example of signals from deep nonvolcanic tremor as observed by Obara (2002). The upper panel shows ground motion at a single station of the Hi-Net borehole array of high-sensitivity/low-noise seismic network. The tremor is strongest from 40–47 min and is characterized by a chaotic signature with no clear, impulsive arrivals. Such signals could be attributable to a range of sources, but their similarity at widely spaced stations and the high phase velocity between them indicates that the signals have a common source, originating deep within the Earth. The lower panel demonstrates the similarity of tremor signals at different stations using smoothed waveform envelopes. From Obara K (2002) Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* 296: 1679–1681.

resistance to keep them from slipping in between earthquakes. Laboratory measurements of friction indicate that for most rocks at seismogenic depth, the coefficient of friction should be ~0.6 (Byerlee, 1978). For earthquakes to propagate unstably, however, friction has to drop from a static level to some lower sliding level, as in the slip-weakening model (see next section). Even slip-weakening behavior, however, does not allow for faults to recover their frictional strength once slip has occurred, so realistic faults require a more complex frictional behavior.

More complex frictional behavior is, in fact, observed in laboratory experiments, and these second-order frictional effects may control much of fault behavior in the Earth. Static friction is not constant, but depends on the history of slip on the interface, with friction increasing as the logarithm of time of stationary contact in laboratory experiments (Dieterich, 1972). A related property is that dynamic sliding friction depends logarithmically on the slip rate. The sign of this dependence determines whether the dependence is weakening or

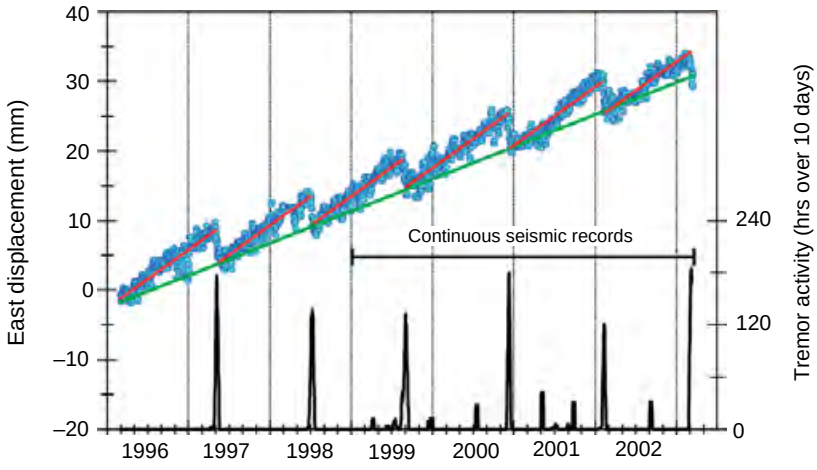


Figure 31 Comparison of slip and tremor activity for Victoria, BC, Canada (Rodgers and Dragert, 2003). Small blue circles show position of a GPS site near Victoria with respect to the interior of North America. Green line shows long-term eastward motion of the site. Red saw-tooth line shows motion and includes reversals in motion approximately every 14 months. These and displacements observed at other GPS stations have been modeled as aseismic slip on the plate interface at depth. The bottom black graph shows the number of hours of tremor activity observed on southern Vancouver Island. The correlation between episodic tremor and aseismic slip is remarkable and has led to the two phenomena being referred to as episodic tremor and slip (ETS). From Rodgers G and Dragert H (2003) Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* 300: 1942–1943.

strengthening with increasing slip rate (Scholz *et al.*, 1972). A third property is that when the slip rate on a sliding surface changes suddenly from one steady rate to another, the coefficient of friction evolves to a new steady-state value after a critical amount of slip, D_c, has occurred. D_c is thought to represent the slip required to renew the true contact area between two rough surfaces (Dieterich, 1978). These three properties – the time dependence of static friction, the rate dependence of dynamic friction, and the evolution of sliding friction with slip – have been unified in a theory of rate- and state-dependent friction laws that describe the evolution of friction (Dieterich, 1979; Ruina, 1983).

An important implication of these laws for earthquakes is that the effect of velocity arises from the combination of two effects – the direct velocity effect, and the evolution of friction with slip – that have opposite signs. Whether slip can occur unstably or must occur stably depends critically on their relative size. Rocks with steady-state velocity strengthening will always slip stably, while rocks with steady-state velocity weakening can slip unstably. The transition from steady-state velocity weakening to steady-state velocity strengthening behavior is thought to control the lower (Tse and Rice, 1986) and upper (Marone and Scholz, 1988) depth extent of the seismogenic zone. (*see* Chapters 4.04 and 4.05).

4.01.5.2 Energy Budget

In the time between large earthquakes, seismogenic faults remain frictionally locked. As plate motion continues, the crust around a fault will gradually distort and store energy in the form of gravitational and elastic strain energy. When the fault starts to slip, this gradually accumulated potential energy is suddenly converted to different forms. Some of it is used to overcome the initial resistance to shear failure. This component of the energy is referred to as the fracture energy. As slip continues during the earthquake at some lower frictional level, energy is converted to heat due to the work done against the resisting frictional force. Another term in the earthquake energy budget is the component that is radiated in the form of seismic waves. This is the seismic energy referred to in Sections 4.01.2.1 and 4.01.3.4. Seismic energy is gradually dissipated and converted to heat elsewhere due to anelastic wave propagation effects. The energy balance can be expressed using a simple equality (Kostrov and Das, 1988) involving the radiated seismic energy, E_s, the change in potential energy, ΔE_p, the energy expended extending the crack, ΔE_γ, and the energy expended overcoming friction, ΔE_f:

$$E_s + \Delta E_p + \Delta E_\gamma + \Delta E_f = 0$$

Of the terms in this equation, we have discussed the radiated seismic energy, a quantity that can be measured using seismic waves. The seismic energy can (with considerable uncertainty) be determined from seismological observations and it is natural to ask what constraints the seismic energy places on the other terms in the energy balance.

A further observational constraint comes from measurements of the slip. Together with the friction, the slip determines how much energy is dissipated as heat on the fault. This is not as simple as it sounds, however, as it is unrealistic to assume that sliding friction is constant. Rate- and state-variable friction laws indicate otherwise, and there are other mechanisms, discussed in the subsequent section, that are expected to introduce much larger dynamic variations in fault strength.

The fracture energy has been estimated for a number of earthquakes. The total fracture energy is proportional to the faulted area, so if these measurements represent fracture energy on a planar fault, they would only be of consequence for very small earthquakes. Seismological estimates of the shear fracture energy are quite large, on the order of 10^6 J m^{-2}. This is orders of magnitude larger than what is measured in the laboratory (Wong, 1982). One interpretation of this discrepancy is that the dissipation of energy being measured includes energy lost due to various forms of yielding near the rupture front (Andrews, 2005). Wilson *et al.* (2005) suggested that finely pulverized rock observed in the near field of the San Andreas Fault, and the tremendous surface area created in that process, might be large enough to account for seismologically inferred fracture energies.

In the slip-weakening model (Ida, 1972), the resistance of the fault to motion is slip controlled and decreases from some maximum when slip first begins, to a lower sliding frictional level once the slip-weakening distance is attained. Thereafter, slip can continue at the sliding frictional level of resistance. Seismologists have attempted to place constraints on the slip-weakening displacement. They have found that it is quite large in major earthquakes, reaching a value of approximately 0.5–1 m for the Kobe earthquake (Ide and Takeo, 1997). Here again, the inferred values are orders of magnitude larger than would be inferred by extrapolating measurements from the laboratory. At least some of this discrepancy may result from a lack of adequate resolution of the slip velocity function (Guatteri and Spudich, 2000).

Other information on the energy balance may come from magnitude dependence of the radiated seismic energy. If the true fracture energy is only an important component of the energy budget for very small earthquakes, then one would expect to see more seismic energy radiated per unit fault slip for larger earthquakes. Seismic moment is more easily calculated than fault slip, so seismologists typically consider changes in the ratio of seismic energy to seismic moment, i.e., E_s/M_0. A possible source of a change in E_s/M_0 comes from the possibility that the resistance to fault motion may drop dramatically with increasing fault slip (Kanamori and Heaton, 2000), as discussed in the next section. Another interpretation is offered by Abercrombie and Rice (2005), who replace the slip-weakening distance with a slip-weakening function for which the the resistance to fault slip continues to decrease with increasing fault slip. As with more dramatic weakening mechanisms, E_s/M_0 is predicted to increase as an earthquake grows larger.

So what do the data indicate? There is a lot of scatter in the measurements, but representative values for large earthquakes are on the order of 5×10^{-5}. Some observations of the magnitude dependence of seismic energy suggest that the ratio of seismic energy to seismic moment increases as a function of earthquake size. That is, earthquakes become more efficient at radiating seismic waves as they grow larger. Measurements from a number of studies of earthquakes over a wide range of earthquake size indicate that E_s/M_0 is approximately a factor of 10 larger for M_W 7 earthquakes than it is for M_W 2 earthquakes (Kanamori and Heaton, 2000). Ide and Beroza (2001) found that many published E_s/M_0 measurements could be biased to low values for small earthquakes because of biases introduced by inadequate corrections for path effects or the limited instrumental bandwidth (**Figure 32**). Thus, whether or not the observed scaling of E_s/M_0 with earthquake size is a true property of earthquakes remains an open question.

4.01.5.3 Microscopic Processes

There are many processes that occur at the microscopic scale, which may have no direct seismological signature, yet still have a profound effect on and may ultimately control macroscopic earthquake behavior. In this section, we briefly review several such phenomena.

4.01.5.3.1 Hydrological processes

An important unknown in earthquake science is the importance of pore fluids on the faulting process. Fluids support normal stress, but do not support shear stress. Thus, one way in which pore fluids can effect faulting is by changing the effective normal stress. Briefly, the criterion for slip under the Coulomb failure criterion

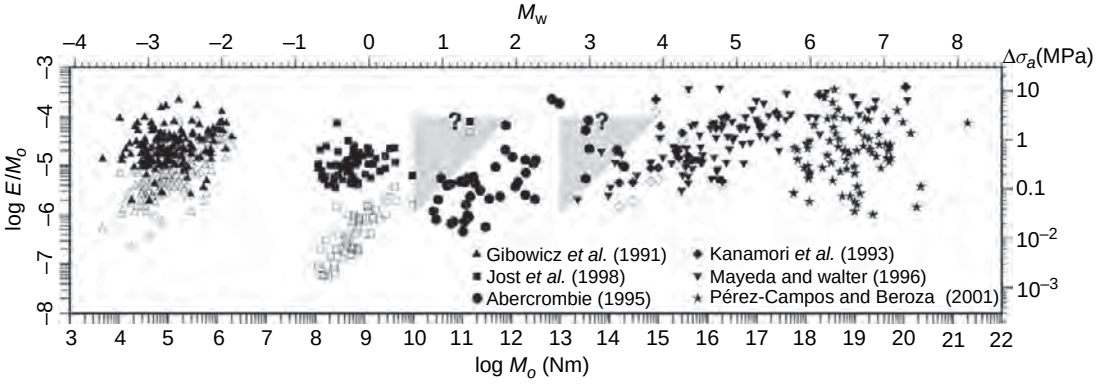


Figure 32 Plot of E_s/M_0 vs M_0 from several studies. Open symbols show original measurements, while filled symbols show values once seismic energy possibly missing due to limited bandwidth is restored. Gray triangles show areas where earthquakes are excluded from two studies due to limited bandwidth. Ide S and Beroza GC (2001) Does apparent stress vary with earthquake size? *Geophysical Research Letters* 28: 3349–3352.

(e.g., Harris, 1998) states that fault slip will occur when the shear stress, τ, acting across the fault exceeds the sum of the cohesion, C, and the effective normal stress multiplied by the coefficient of friction, μ:

$$\tau > \mu(\sigma - P) + C$$

where the term in the parentheses is referred to as the effective normal stress and is the difference between the fault normal stress, σ, and the pore pressure, P. The cohesion is usually assumed to be negligible. By this criterion, one way in which earthquakes can be triggered is by an increase in the pore pressure, P.

This effect has been cited by Hubbert and Rubey (1959) as a possible explanation for the movement of large, nearly horizontal thrust faults due to the large frictional resistance expected from the considerable overburden. Fluids have been shown capable of artificially triggering earthquakes in controlled experiments (Healy *et al.*, 1968; Raleigh *et al.*, 1976); thus, it is reasonable to expect that they might play a similar role for tectonic earthquakes. Both the lack of a heat flow anomaly (Brune *et al.*, 1969) and the nearly fault-normal stress orientation of some major fault zones (Zoback *et al.*, 1987) could be understood if the effective normal stress is low, due to high pore fluid pressure, such that the fault slips at low shear stress. There is independent evidence that fluids and faults interact (Sibson *et al.*, 1981), and that high pore pressures occur in fault zones (**Figure 33**), at least at shallow depths (Unsworth *et al.*, 1997). Dehydration embrittlement, in which phase changes that expel water decrease the effective normal stress in subducting lithosphere, is cited as a possible enabling mechanism for intermediate-depth earthquakes (Kirby *et al.*, 1996).

Another way in which fluids can influence faulting is through poroelastic effects. In a poroelastic medium, pore pressure will decrease where the mean normal stress decreases and increase where the mean normal stress increases. The immediate change in pore pressure, ΔP, for what is known as undrained conditions can be related to the change in the mean normal stress, $\Delta\sigma_{kk}$, by the equation,

$$\Delta P = B\Delta\sigma_{kk}$$

where Skempton's coefficient, B, a function of elastic moduli for drained and undrained conditions, is equal to 1 for fluid saturated soils and ranges between 0.5 and 0.9 for a range of rock types (Rice and Cleary, 1976). With time, this undrained state evolves to the drained state as fluids flow and re-equilibrate in response to pore pressure gradients.

Pore fluids can trigger earthquakes by inducing shear stress through a process referred to by Booker (1974) as consolidation. Consolidation in this usage refers to the process in which, in places where the mean normal stress and pore pressure increase, normal stress and pore pressure then gradually decrease as pore fluids migrate away from areas of high pore pressure. Where the mean normal stress and the pore pressure decrease, the mean normal stress will decrease further as pore fluids migrate into areas of lower pore pressure. The net effect is to couple normal and shear stress through pore–fluid migration. Booker proposed this as a candidate mechanism for the gradual decay described in Omori's law of aftershock decay.

Yet another way in which fluids can influence earthquakes and faulting is through chemical effects

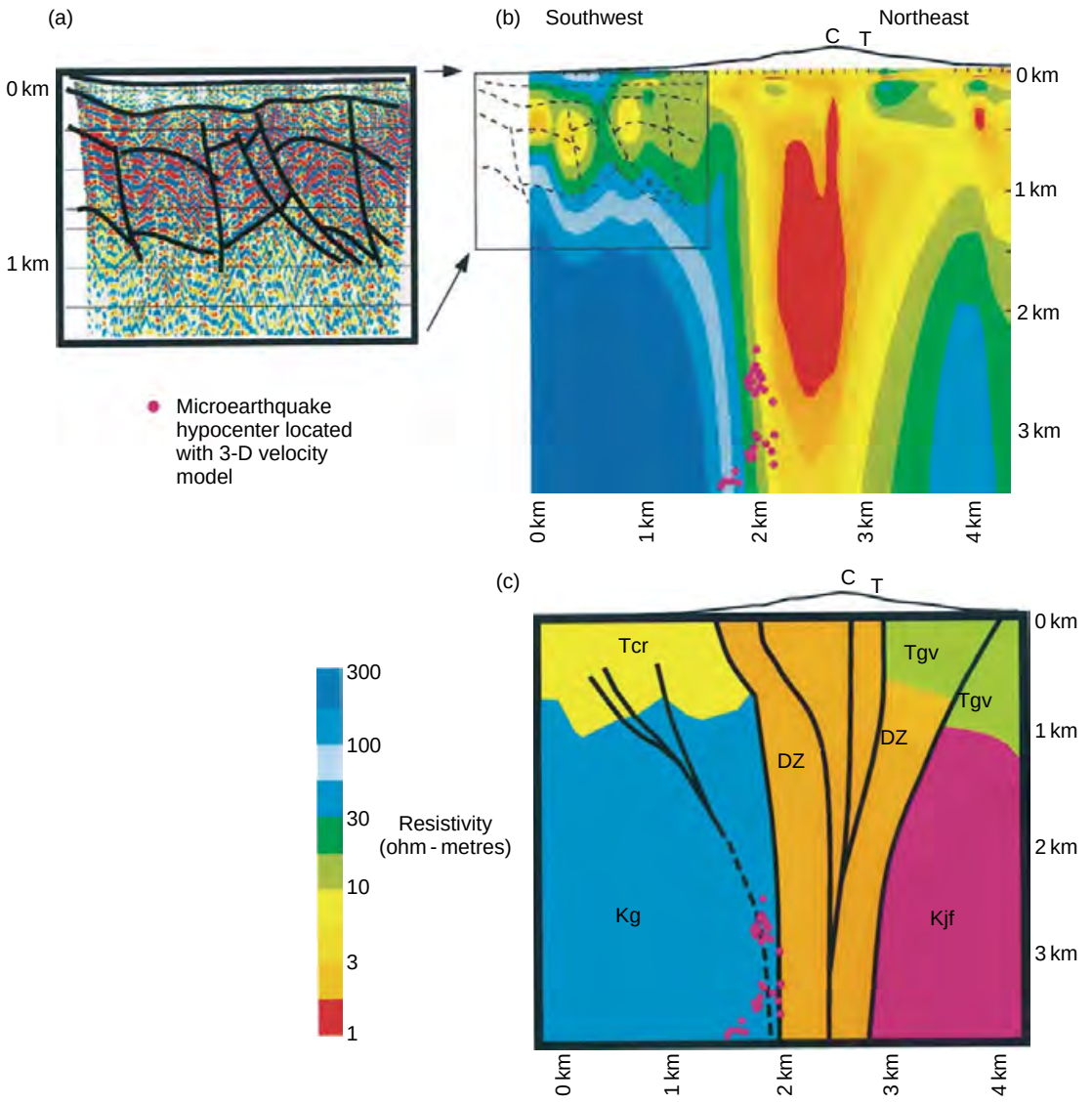


Figure 33 (a) Migrated seismic reflection data on profile normal to the San Andreas Fault. Depth of 1 km corresponds to two-way travel time of 1 s. (b) Resistivity model based on magnetotelluric data. Fault trace is creeping at the surface is marked with a 'T' and is shown along with surface topography. (c) Geologic interpretation: Kg is Salinian granite, Tcr denotes Coast Range sedimentary units, Tgv – Tertiary Great Valley sequence, Kjf – Franciscan formation, and DZ – damaged zone. Reproduced from Unsworth MJ, Malin PE, Egbert GD, and Booker JR (1997) Internal Structure of the San Andreas fault at Parkfield, California. Geology 25: 359–362, with permission from GSA.

and stress corrosion (Scholz, 1968). It is also worth noting that other fluids, notably magma and carbon dioxide (Miller *et al.*, 2004), can have similar effects. For more details on the involvement of fluids in faulting, *see* Chapter 4.10.

4.01.5.3.2 Melting

Rapid shear slip during earthquake faulting will lead to shear heating. If this heating is large enough, it can lead to very large changes in temperature, and, possibly, to melting on the fault (Jeffreys, 1942; McKenzie and Brune, 1972). There are several variables that control whether or not melting occurs, and if it does, how extensive it is. The heat generated per unit area of fault will be the product of the slip and the frictional resistance, so the slip and sliding friction will in part determine the conditions for possible melt generation. In addition, any real fault zone will be a shear zone of finite thickness. The thicker the shear zone, the more distributed heat production will

be, and all other things being equal, the more slip required to initiate melting. The thermal conductivity will control how quickly heat diffuses away from the fault. Thermal conductivity of rocks is low enough that this will not likely be an important effect, unless either the slip rate is very low or pore fluids are present. The effect of melt on friction is not straightforward. Laboratory experiments suggest that the initial effect might be to increase fault strength due to viscous effects, but this effect might be diminished at high stress. In any case, further slip will lead to more heating and a reduction in viscosity, and hence strength (Fialko and Khazan, 2005). A detailed treatment of temperature effects during faulting can be found in Chapter 4.03.

4.01.5.3.3 Thermal pressurization

The same frictional heating discussed in the previous section, in the presence of pore fluids, can lead to thermal pressurization of those fluids well before the onset of melting. Thermal pressurization can have a strong effect on faulting by reducing the effective normal stress. The absence of geological evidence for melting near fault zones led Sibson (1973) to propose thermal pressurization as a mechanism for reducing dynamic fault strength. For undrained conditions, a temperature change of ΔT leads to a pressure change, ΔP of, $\Delta P = \Lambda \Delta T$, where, Λ is the thermal expansion coefficient of water. Representative values of Λ for upper crustal conditions are 0.78–0.93 $\mathrm{MPa^{\circ}C^{-1}}$ (Rempel and Rice, 2006). Thus, thermal pressurization is potentially a very large and efficient agent of fault weakening. Thermal pressurization was explored as a potential explanation for the lack of a heat flow anomaly on the San Andreas Fault by Lachenbruch and Sass (1980). For it to operate, fluids must be present and fault-zone permeability has to be low enough that the excess pressure does not diffuse into the host rock. Measurements of the permeability of fault zones and their surroundings are very low (Wibberley and Shimamoto, 2003), however, any dilatancy effects during rapid slip could rapidly increase that permeability and act to short circuit this weakening mechanism (*see* Chapter 4.03).

4.01.5.4 Fault-Zone Structure

Detailed geological observations place important constraints on the physical processes that operate during faulting in the Earth. The geometrical complexity of faults has a first-order effect on the faulting process. Geologic observations also provide constraints on the nature of the deforming or damaged zone in the direction normal to the fault. These observations are crucial for understanding the mechanics of faulting and bear directly on some of the most important problems in earthquake physics.

Seismological constraints on fault-zone structure are of much lower resolution, but they have the advantage of measuring properties of faults that are actively deforming and not exhumed. Earlier in this chapter, we outlined some of the recent progress on the fine structure of faults as revealed by precise earthquake locations (**Figures 11** and **12**). Precise earthquake locations bear directly on questions such as how wide the fault zone is at seismogenic depths and whether or not fault trace complexities observed at the surface reflect complexities in fault geometry at depth.

Another important constraint on the structure of fault zones comes from fault-zone guided waves (FZGWs). Because faults move geological units past one another, they tend to be planes that separate materials with different properties. Moreover, the mechanical, chemical, and/or hydrological alteration of materials within deforming fault zones leads to them being a locus of low seismic velocities. Variations in material properties either within or across fault zones lead to strong effects on seismic wave propagation. Waves that are trapped within the low-velocity material within several hundred meters of active fault zones are known as FZGWs. These are analogous to Love waves in vertically layered media in that they usually consist of critically reflected S waves within the low-velocity material. FZGWs have been observed in the subduction zone in Japan (Fukao *et al.*, 1983), continental normal faults (Leary *et al.*, 1987), and continental strike-slip environments (e.g., Li *et al.*, 1990). Waves that refract horizontally due to a large contrast in velocity across a fault zone are known as fault-zone head waves. These have also been observed in diverse environments, including both subduction zones (Fukao *et al.*, 1983) and continental strike-slip faults (e.g., McNally and McEvilly, 1977).

Fault-zone waves are very sensitive to the details of fault-zone structure and have the potential to place important constraints on the structure of active fault zones. If the geometric regularity of the waveguide is disrupted, then such waves may not propagate or may be strongly altered. For this reason, they can be used to explore fault segmentation. An outstanding question regarding FZGWs is just how deep the waveguide extends. Some investigators find that it spans the entire depth of the seismogenic zone

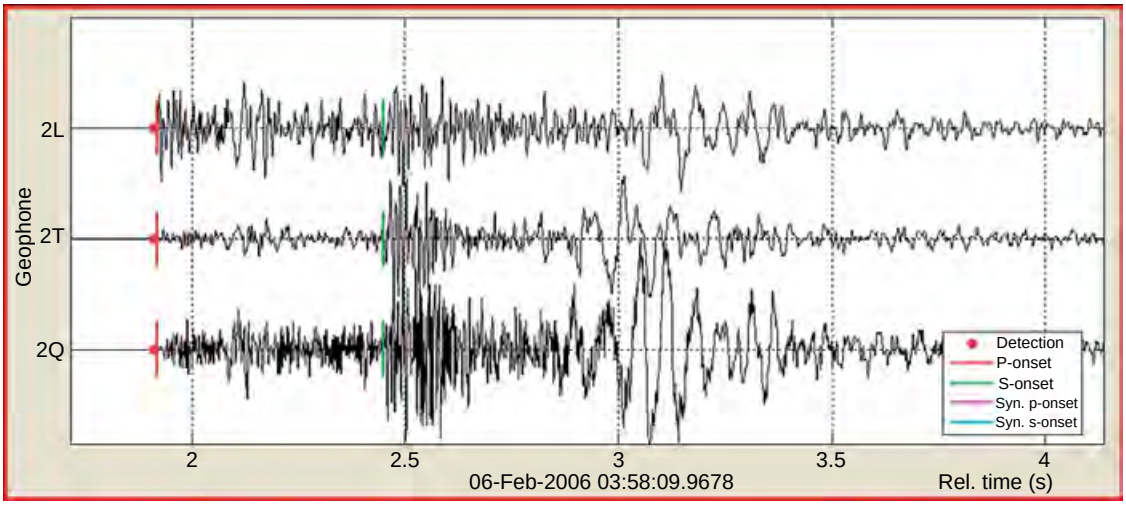


Figure 34 FZGWs observed at 2.7 km depth in the SAFOD borehole from a magnitude 1.3 earthquake. The presence of these waves at this depth is diagnostic of a 200-m-wide low-velocity zone associated with the San Andreas Fault. From Ellsworth *et al.* (2007) Seismology inside the fault zone: Applications to fault-zone properties and rupture dynamics. *Scientific Drilling Journal* (in press).

(Li *et al.*, 2000), while others argue that it may only extend through the upper few kilometers (Ben-Zion *et al.*, 2003). The recent observation of FZGWs in the SAFOD borehole through the San Andreas Fault suggests that the fault-zone waveguide extends to seismogenic depths (**Figure 34**).

Another important question regarding the fault zone waveguide is to what extent it is a permanent feature, and how much it is a transient that follows a large earthquake. Following the Landers earthquake, there is a clear and strong postseismic healing, or recovery of low velocities, indicating that at least some of the low-velocity fault zone is transient (Li *et al.*, 1998).

The relationship between low-velocity fault zones and the mechanical behavior of faults in earthquakes is important for earthquake faulting. Observations of surface deformation obtained by interferometric satellite radar interferometry (Fialko *et al.*, 2002) following the 2001 Hector Mine, CA, earthquake indicate that nearby fault zones are characterized by relatively compliant zones that are of the order of a kilometer wide. The effect of relatively low-strength materials is an important question that is just now being addressed.

4.01.5.5 Borehole Observations, Fault-Zone Drilling, and Seismicity in Mines

Another approach to the study of earthquakes and faulting is to make observations below the Earth's surface. This can take the form of recording earthquakes with seismometers deployed in boreholes, direct drilling into active faults, or studying seismicity in deep mines. This approach has a number of distinct advantages.

Borehole seismic observations allow earthquake scientists to make observations below the highly attenuating and strongly scattering layers of the Earth's surface. Borehole measurements have had an important impact on seismology as they allow a relatively unobscured view of seismic waves as they are radiated from the earthquake source (see the section on stress within the Earth below). The Hi-Net seismic network in Japan capitalizes on this idea as it uses sensitive borehole instruments to monitor the entire Japanese archipelago. It led directly to the discovery of deep, nonvolcanic tremor (Obara, 2002). The idea of making measurements in the very near field of a fault at depth is a key element of the ongoing SAFOD project.

Deep drilling allows direct sampling of active faults *in situ*, which is a tremendous advantage. It also allows scientists to make stress measurements at depth near active faults. However, fault-zone drilling is a logistically challenging and expensive undertaking. These disadvantages of fault-zone drilling have prevented its wider application. The expense of drilling rises rapidly with increasing depth and hence only relatively shallow depths are typically targeted. Fault-zone drilling also allows only discrete sampling of a few select points on a fault surface. Despite these drawbacks, fault-zone drilling has played, and will continue to play, an important role in earthquake science.

Notable deep drilling projects directly related to earthquakes include the ongoing SAFOD project, which has drilled through the actively deforming San Andreas Fault Zone near Parkfield, CA. Among the goals of the project are to drill through the source of a repeating earthquake sequence and to compare it with nearby fault-zone materials that exhibit creep rather than earthquakes. Another important fault-zone drilling experiment is the Taiwan Chelungpu Fault Drilling Project (TCDP). This project has drilled into the shallow part of the rupture of the M_w 7.6 1999 Chi Chi, Taiwan, earthquake in order to study fault-zone properties in the immediate aftermath of a large earthquake. Other fault-zone drilling projects of note include the Gulf of Corinth Rift Laboratory, which seeks to understand a continental back-arc rifting environment and the Nankai Trough Seismogenic Zone Experiment, which has the ambitious goal of drilling the accretionary prism of the Nankai Trough in Japan. These and other fault-zone drilling projects will provide important constraints on the fluids, fluid pressure, chemistry, permeability, deformation mechanisms, and stress as they relate to earthquake behavior. A detailed discussion of fault-zone drilling projects can be found in Chapter 4.22.

Another environment that is used to study earthquakes *in situ* is in deep mines. The act of mining removes load-bearing material and induces earthquakes. These earthquakes are extremely dangerous to miners and hence there is a great deal of interest in understanding and predicting them. They also have the potential to tell us much about how ordinary tectonic earthquakes work. Seismicity in mines offers some advantages over studying seismicity remotely. The location of mining operations is known and the likely areas of seismicity can be inferred from them. Another advantage is that seismic instruments can be deployed in three-dimensional arrays, which is not possible with strictly surface observations. Seismicity in mines is also at a scale that bridges the laboratory and usual seismological scale of observation. These are the principal motivations behind the multidisciplinary Natural Earthquake Laboratory in South African Mines (NELSAM) project, which is studying earthquakes in the western deep-level South African gold mine.

4.01.5.6 Earthquakes as a Complex System

Earthquakes do not happen in isolation, but interact strongly with each other. The same goes for faults, which might more properly be described as fault systems. A full numerical treatment of fault systems would require realistic geology, fault geometries, rheology, a full range of earthquake sizes, and multiple earthquakes. Although progress in being made on modeling earthquakes in ways that include each of these characteristics, combining them into a single comprehensive model is far beyond current capabilities. There is, however, an approach to studying earthquakes that seeks simplifications to the effect of these properties or characteristics, while at the same time attempting to extract the important system-level behavior that affects earthquakes. Studies of this sort treat the behavior of large earthquake populations statistically, with the goal of understanding the macroscopic behavior based on an understanding of the system at a microscopic level. An example of this sort of treatment, which seeks to explain the *b*-value of seismicity distributions as a manifestation of self-organized criticality is the study of Bak and Tang (1989). A review of this approach to understanding earthquake phenomena can be found in Chapter 4.23.

4.01.6 State of Stress, Nucleation, and Triggering

4.01.6.1 State of Stress

The orientation of stress within the Earth can be determined by diverse methods such as earthquake focal mechanisms, borehole breakouts, and the orientation of volcanic dikes. As shown in **Figure 35**, systematic compilations of stress orientations indicate that it is coherent over large scale lengths and generally consistent with what is expected from ongoing plate tectonic motion (Zoback *et al.*, 1989). The magnitude of these stresses at seismogenic depth, as well as the orientation of stress in the vicinity of major fault zones, is much less clear and the subject of ongoing research.

The lack of a detectable heat flow anomaly centered on the San Andreas Fault (Brune *et al.*, 1969), which is expected under the assumption that fault friction at depth is similar to that observed in the laboratory, indicates that friction is low, at least during slip in large earthquakes. The observation that stress along the San Andreas tends to be oriented such that the maximum compressive stress is close to fault normal (Zoback *et al.*, 1987) is consistent with the notion that fault slip occurs at low levels of shear stress. This conclusion has been disputed by Scholz (2000), who argued that stress rotates near the San Andreas Fault to angles that are consistent with

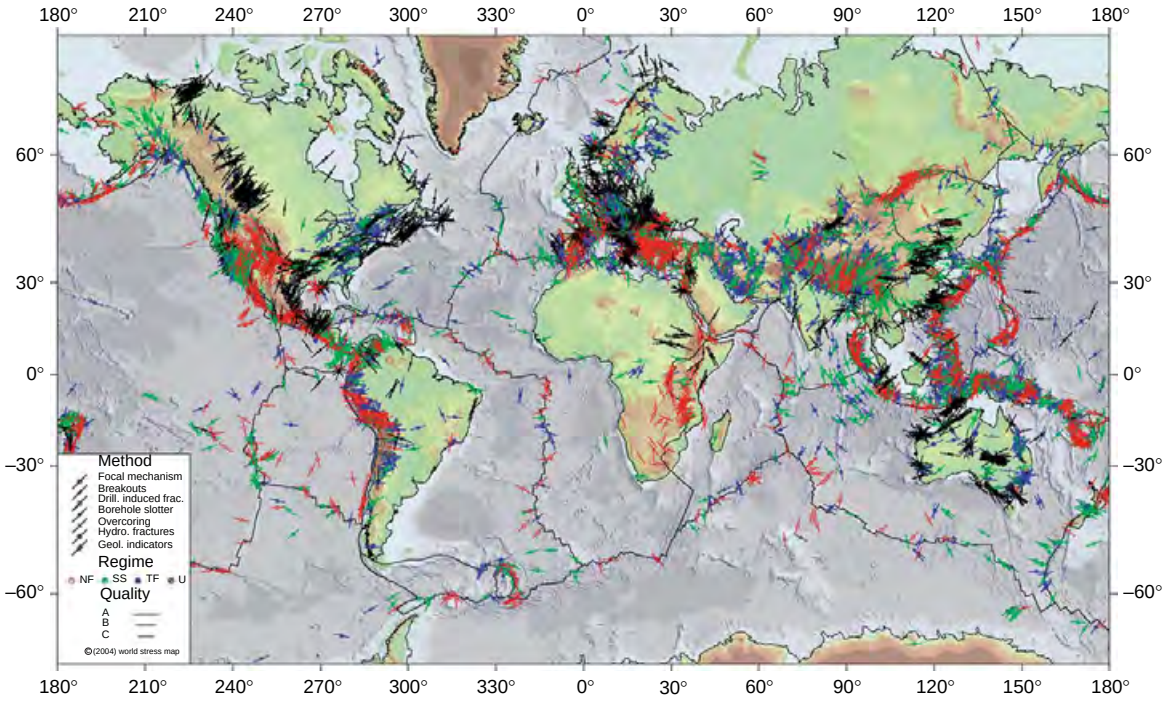


Figure 35 2005 version of the world stress map. Stress directions are determined by a range of techniques listed in the legend (Reineker *et al.* (2005) The release 2005 of the world stress map).

expectations from ordinary levels of friction. Further, borehole stress measurements suggest that stress increases with depth in accord with expectations from the laboratory. Scholz (2000) cites a much broader heat flow anomaly – approximately 80 km wide – centered on the San Andreas Fault (Lachenbruch and Sass, 1980) as the signature of heat generated during repeated large earthquakes. Larger permeabilities than have been measured in heat flow experiments would be required for convective heat transport to operate using crustal fluids, but if permeability is scale dependent, then it could reach such an extent. Recent work on geologic signatures of transient heating (d'Alessio *et al.*, 2003) and on transient heating in the immediate aftermath of large earthquakes (Kano *et al.*, 2006) should help to resolve this important and longstanding controversy.

Earthquake focal mechanisms provide constraints on the local orientation of the stress field. Earthquakes can also provide information on changes in stress. During an earthquake, shear stress on the fault changes – for the most part dropping on parts of the fault that slipped and increasing on parts of the fault that did not. If the area of slip can be determined, then a characteristic dimension of faulting can be inferred from it. The ratio of the average slip to the characteristic dimension can be thought of as a strain drop, and interpreted through various simple models of source geometry as a drop in stress. This 'stress drop' represents how much the stress decreased from its initial level before the earthquake to the final level after the earthquake – at least on the parts of the fault that slipped. Thus, stress drop provides indirect, model-dependent, but nonetheless useful constraints on the absolute level of stress. It is usually assumed that friction at depth is sufficiently large that shear stress does not drop below zero and act in the opposite direction following an earthquake, a condition that is termed 'dynamic overshoot'. Under this assumption, stress drops provide lower bounds on the level of shear stress acting on faults at seismogenic depths.

Stress drops (see Section 4.01.3.4) are determined seismically from a combination of the seismic moment and some measure of characteristic dimension. For shallow large earthquakes, the characteristic dimension can be estimated from various sources such as geodetic data, geological data (e.g., surface break), tsunami source, slip inversion, aftershock area, macroseismic data (e.g., intensity distribution). For small earthquakes, the source dimension cannot be directly determined, and is usually taken as the radius of a circular crack. The radius is determined

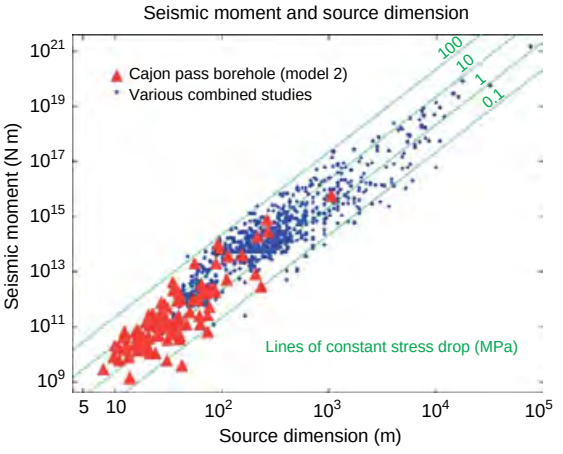


Figure 36 Abercrombie (1995) found from corner frequency measurements taken from a borehole seismometer that the relationship between source dimension and seismic moment for small earthquakes was the same as that for large earthquakes. Diagonal lines showing constant stress drop ($M_0 \sim r^3$) explain the trend in the data. From Abercrombie RE (1995) Earthquake source scaling relationships from −1 to 5 ML, using seismograms recorded at 2.5 km depth. *Journal of Geophysical Research* 100: 24015–24036.

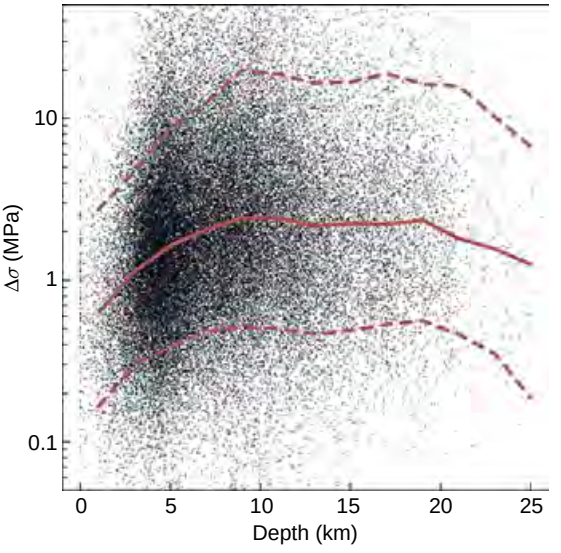


Figure 37 Shearer *et al.* (2006) were able to discern a depth dependence to stress drop for earthquakes in southern California. Black dots show individual event stress drops. Solid line shows median values with depth and 10th and 90th percentiles shown as dashed lines. From Shearer PM, Prieto GA, and Hauksson E (2006) Comprehensive analysis of earthquake source spectra in southern California. *Journal of Geophysical Research* 111: B06303 (doi:10.1029/2005JB003979).

by measuring the corner frequency – the frequency of radiated waves at which the finiteness of the source of the earthquake first becomes apparent – with the assumption that the rupture speed is approximately equal to the shear-wave speed. Data from earthquakes, large and small, show that stress drops follow a scaling relation of the form $M_0 \sim r^3$. Since the faulted area, a, follows $a \sim r^2$, this leads to the average slip Δu, scaling as $\Delta u \sim r$, or, in other words, the ratio of slip over length is constant (**Figure 36**).

Assuming that stress drop reflects the ambient stress, it is reasonable to expect that as depth increases, and the stress due to overburden changes with it, earthquake stress drops would increase as well. Such a correlation has proven difficult to find. In part, this may be due to the strong dependence of stress drop estimates on the corner frequency, which means that a small error in corner frequency will lead to a large error in the stress drop. By stacking over thousands of earthquakes, Shearer *et al.* (2006) have discerned just such a correlation (**Figure 37**). The depth dependence is much weaker, however, than would be expected if stress drops scaled with the overburden, and most of the dependence is in the upper third of the ~15 km depth extent of the seismogenic zone.

Another approach to determining stress drops is to use repeating earthquakes – earthquakes that repeatedly rupture a particular part of a fault – together with the assumption that these areas slip only seismically and keep up with long-term steady slip. For a given seismic moment, one can determine the average slip such that the slip matches the long-term rate. Stress drops determined by this approach (Nadeau and Johnson, 1998) are strongly scale dependent. Stress drops for earthquakes of magnitude ~6 are comparable to those based on corner frequency measurements, but they find that stress drops measured using slip-rate matching of repeating earthquakes increase rapidly with decreasing earthquake size such that for the smallest earthquakes they are 1 to 2 orders of magnitude higher.

4.01.6.2 Earthquake Nucleation and Short-Term Earthquake Prediction

Earthquake prediction can fairly be called the holy grail of seismology. While it may be that aspects of earthquakes are inherently unpredictable, the ability to predict behavior is the *sine quanon* of science. Thus, over the long term at least, we can make predictions about earthquakes based on what we really do understand. For example, it is safe to 'predict' that over the next century, most earthquakes will occur on plate boundaries. This is not what most people have in mind when the phrase 'earthquake prediction' is used.

Rather, it is understood to mean prediction of the location, size, and time of a specific earthquake to within a matter of days to months. Needless to say, this sort of prediction has not been realized, and there are many who maintain that the nature of the earthquake instability makes prediction on such short timescales intrinsically impossible.

The challenge of short-term earthquake prediction can be illustrated by comparing earthquakes with lightning – a phenomenon that shares some important similarities. Both earthquakes and lightning involve the catastrophic conversion of slowly accumulated potential energy to suddenly released kinetic energy. In the case of lightning, it is slowly accumulated electrical charge that suddenly flows as electrical current in a lightning bolt and radiates sound waves as thunder. For earthquakes, it is gradually accumulated elastic strain energy stored in the Earth's crust that suddenly accelerates the material on both sides of the fault and radiates energy in the form of seismic waves. The timescale of charge accumulation in a thunderhead can be measured in tens of seconds (order 10^1 s), whereas the timescale of strain accumulation in the Earth's crust can be measured in centuries (order 10^{10} s). Based on the timescale of energy accumulation, predicting the size, location, and time of an earthquake to within a day (order 10^5 s), is equivalent to predicting the amperage, location, and time of a lightning bolt to within a tenth of a millisecond (order 10^{-4} s). At face value, this sounds wildly unrealistic, but with lightning such a prediction is straightforward because lightning is preceded by a 'nucleation' process.

The Earth's lower atmosphere is normally a very effective insulator. In order for a lightning bolt to propagate between a cloud and the ground, it must first create a conductive path for the lightning strike to follow. It does this through the formation of a stepped leader – the piecewise ionization of air that creates a path from cloud to ground. As the air is ionized, it transforms from an insulator to a conductor and forms a channel for the flow of current. Once a channel connects the cloud to the ground, the circuit is completed, charge can flow, and the discharge of the main lightning strike occurs, typically traveling from ground to cloud. For a lightning strike, the formation of the stepped leader is the nucleation phase. The stepped leader immediately precedes the lightning bolt, and occurs over the entire length of the subsequent main event. The prediction of the time and location of a lightning bolt would be trivial once the stepped leader is observed. The key question for earthquake prediction is whether or not earthquakes have a similar nucleation process.

Laboratory (Dieterich, 1979) and theoretical (Andrews, 1976) models of earthquake initiation both indicate that aseismic nucleation must occur before slip can propagate unstably across a fault, so it seems highly likely that nucleation of some sort must occur before earthquakes. Assuming it does, the question becomes, how extensive is the earthquake nucleation process? If it occurs over a very limited part of the fault, and thereafter earthquake rupture propagates as a self-sustaining process, then earthquake prediction will be very difficult – perhaps impossible. In this scenario, termed the cascade model by Ellsworth and Beroza (1995), nucleation of small and large earthquakes might be identical. So for earthquake prediction to become a reality, not only would we need to observe the nucleation process, we would also have to know that the fault is ready to host a self-sustaining, sometimes termed 'spontaneous', rupture. This is a great deal more knowledge than we have about faults at present, and such information is likely to be hard to come by. If, on the other hand, the nucleation process scales with the size of the eventual earthquake (for instance if the size of the nucleation zone were proportional to the size of the eventual earthquake), then earthquake prediction would be a good deal more likely, though still extremely challenging. Another source of significant uncertainty is multiple-fault interaction. The M_w 7.9 Denali, AK, earthquake of 2002, for example, began as a reverse faulting event on the Susitna Glacier Fault, which led to a large strike-slip faulting earthquake on the Denali fault.

4.01.6.3 Earthquake Forecasting

The prospects for estimating earthquake likelihoods over timescales of decades and longer, sometimes known as earthquake forecasting, is considerably brighter. We know that the great majority of earthquakes occur either on plate boundaries or on previously identified fault systems, so for the most part, we know where earthquakes will occur. We also know quite accurately how rapidly plates are moving, so we have fault-slip boundary conditions that must be met over a fault, or a system of faults that comprises a plate boundary, over time. This knowledge, together with information such as the time of the last major earthquake, or assumptions of the size distribution of large earthquakes, is enough to start estimating earthquake likelihoods. While earthquake forecasts are made routinely now as a part of seismic hazard assessment, the long timescales over which

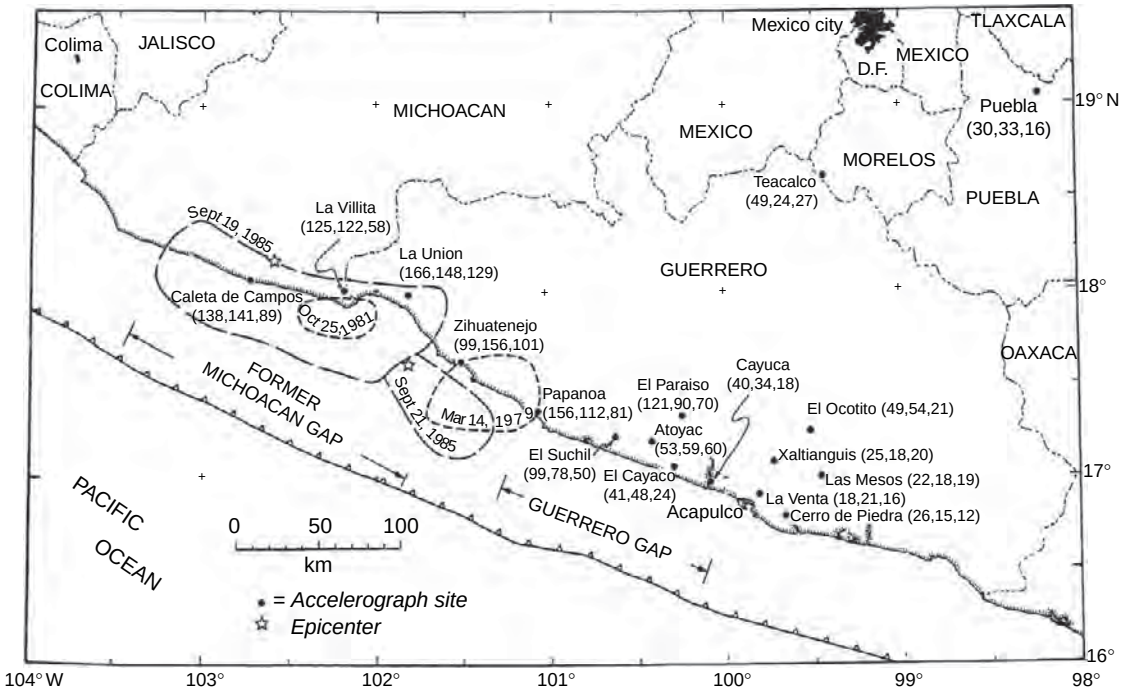


Figure 38 Map showing aftershock zones of the 19 and 21 Sep. Michoacan, Mexico, earthquakes. Note that the rupture of the two earthquakes abut, but that the 1985 rupture is much larger than, and entirely envelopes, the rupture zone of the 25 Oct. 1981 Playa Azul earthquake. If such behavior is common, then it will greatly complicate attempts to apply the seismic gap hypothesis. From Singh SK, Mena E, and Castro R (1988). Some aspects of source characteristics of the 19 September 1985 Michoacan earthquake and ground motion amplification in and near Mexico City from strong motion data. *Bulletin of the Seismological Society of America* 78: 451–477.

they apply make it difficult to test them rigorously. Moreover, to the extent that they have been tested, some studies have found them wanting.

The seismic gap hypothesis as formulated by Fedotov (1965), McCann *et al.* (1979), and Kelleher *et al.* (1973) holds that earthquakes are more likely to occur in areas prone to large earthquakes, but that have not recently had large earthquakes. Kagan and Jackson (1991) found that, as formulated, the seismic gap model has failed to predict where major earthquakes are more likely to occur. They interpret this to mean that earthquakes are clustered in time, adding that because the theory of plate tectonics only provides long-term constraints on slip rates, it does not imply the regularity of earthquake recurrence.

Another possible interpretation of the failure of the seismic gap hypothesis is that it needs to be refined to take into account complicating factors, such as the possibility that earthquakes are not characteristic, or that slip in earthquakes is strongly heterogeneous. As an example, consider the case of the 1979 Playa Azul and subsequent 1985 Michoacan, Mexico, earthquakes (Singh *et al.*, 1988; **Figure 38**). The two earthquakes occurred just 6 years apart and their spatial centroids are essentially the same, yet they clearly had very different, and most probably complementary, slip distributions. It should be possible to test a weaker form of the seismic gap hypothesis, such as whether or not earthquakes occur in areas that have not slipped recently, once geologic fault slip rates are taken into account. The Playa Azul and Michoacan earthquakes would pass this test, while failing the original seismic gap hypothesis. It is interesting to note that the other seismic gap shown in **Figure 38**, in the state of Guerrero, has not been the source of a large earthquake in the interval since Singh *et al.* (1988) published their paper. The Guerrero gap, however, has been the site of a $M_W \sim 7.5$ large silent earthquake (Kostoglodov *et al.*, 2003), an observation that complicates the interpretation of this seismic gap.

4.01.6.4 Static Stress Triggering

Over the past few decades, it has become increasingly obvious that our ability to forecast earthquakes is

strongly dependent on our understanding of how earthquakes interact with one another. This is a critical question for earthquake hazard assessment that will inevitably depend on essential aspects of earthquake physics. How does the occurrence of one earthquake influence the occurrence of another? The seismic gap hypothesis is a limiting case of this in which the effect of an earthquake on a subsequent earthquake at the same place on the plate boundary is considered. Beyond this, the most obvious such interaction is an aftershock sequence in which many smaller earthquakes occur in the aftermath of a larger 'mainshock'. Large earthquakes are sometimes implicated in the triggering of other large earthquakes too, with the most famous example being the twentieth century sequence of large earthquakes that eventually ruptured most of the North Anatolian Fault in Turkey (Toksöz *et al.*, 1979).

Earthquake interaction is thought to occur through stress triggering, that is, failure in a future earthquake is either encouraged or discouraged by the static stress change induced by an earthquake that precedes it. Invoking the same criterion as in Section 4.01.5.3.1, fault slip will occur when the shear stress, τ, acting across the fault exceeds the sum of the cohesion, C, and the effective normal stress multiplied by the coefficient of friction, μ:

$$\tau > \mu(\sigma - P) + C$$

where the term in the parentheses is referred to as the effective normal stress and is the difference between the fault normal stress, σ, and the pore pressure, P. The cohesion is usually assumed to be negligible. By this criterion, earthquakes can be triggered by either an increase in shear stress, τ, a decrease in normal stress, σ, an increase in the pore pressure, P, or some combination of the three. There is now a substantial body of literature that explores static stress triggering (Stein, 1999) and its implications for earthquake physics, for earthquake interaction, and for time-dependent seismic hazard assessment. **Figure 39** shows perhaps the clearest example of static stress triggering: the progression of earthquakes across the North Anatolian Fault in Turkey during the twentieth century (Stein *et al.*, 1997). As outlined in Section 4.01.5.1, friction on faults is thought to be dependent on slip rate and the past slip history of a fault. Changes in long-term earthquake probabilities due to stress steps induced by nearby earthquakes under rate- and state-variable friction laws have been explored by Parsons *et al.* (2000).

Perhaps less obvious than the triggering of earthquakes by static stress changes is the notion of a stress shadow, in which earthquakes are inhibited because the stress necessary to drive them is relieved by a nearby large earthquake. The most famous and convincing example of this may be the almost complete lack of significant earthquakes in northern California following the 1906 San Francisco earthquake (Ellsworth *et al.*, 1981). Over the 100 years since the 1906 earthquake, there have been several earthquakes near the ends of the 1906 earthquake rupture – most notably the 1989 Loma Prieta earthquake, but nowhere near the level of seismicity that occurred in the 50 years leading up to the 1906 earthquake (*WGCEP*, 2002). The topic of static stress change triggering of earthquakes is covered in considerably more detail in Chapter 4.08.

4.01.6.5 Dynamic Triggering

The examples cited above are of earthquakes either encouraging or inhibiting other earthquakes as a consequence of the static stress change that they induce. Another mode of earthquake interaction is through dynamic stress changes, that is, the transient stresses that are transmitted in the form of seismic waves. These were first clearly recognized in the aftermath of the 1992 Landers, CA, earthquake in which 'aftershocks' were triggered at extreme distances of more than 1000 km from the main shock (Hill *et al.*, 1993). At these distances, static stress changes induced by the earthquake are much smaller than the periodic stress changes induced by solid-Earth tides and are thus unlikely to act effectively as a trigger. Dynamic stress changes, on the other hand, are still quite substantial at great distances. This is attributable to the fact that static stress changes decay approximately as distance cubed once a distance comparable to the source extent is exceeded. In contrast, dynamic stresses associated with surface waves decay much more slowly.

Since the Landers earthquake, dynamic triggering at great distance has been documented in the 1999 Izmit earthquake, the 2002 Denali earthquake, the 2004 Sumatra earthquake, as well as in other events. In the case of Sumatra, micro-earthquakes were triggered at Mt. Wrangell, AK, at a distance of nearly 11 000 kilometers from the rupture zone. In this instance, the triggered earthquakes were observed to occur during the maximum vertical displacement of the large and extremely long-period Rayleigh waves (**Figure 40**) generated by that event (West *et al.*, 2005).

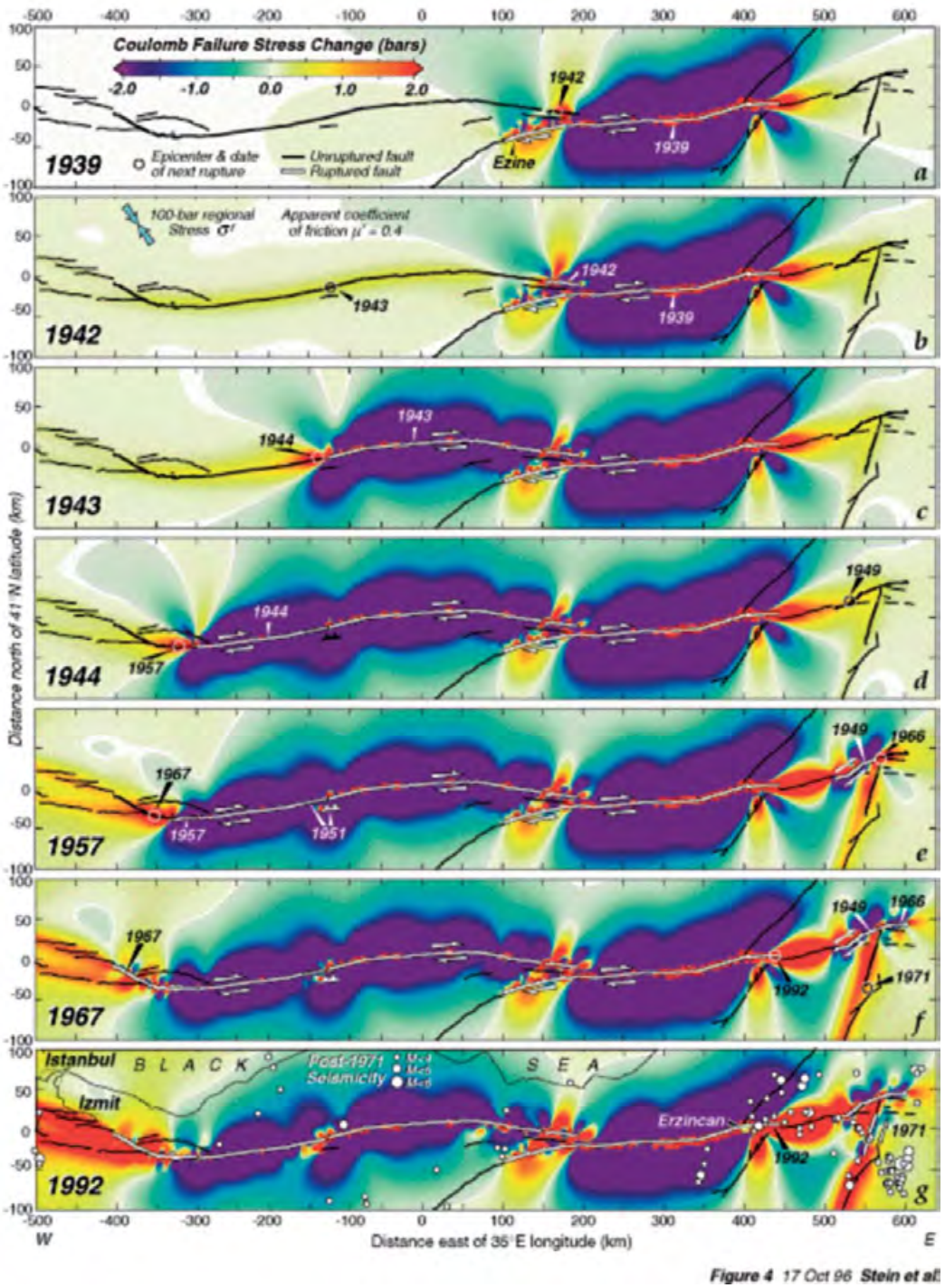


Figure 39 Cumulative Coulomb stress changes caused by large earthquakes on the North Anatolian fault in Turkey since 1939. In each panel, the epicenter of the next earthquake to rupture is circled. All but the 1943 epicenter lie in regions where the stress rose significantly, typically by 2–5 b, owing to a combination of previous earthquakes and deep fault slip. Relocated earthquakes (Engdahl *et al.*, 1998) are shown in the last panel, with the 1992 aftershock sequence removed. Calculations are made assuming a Poisson, elastic half-space. Slip is assumed uniform from 0 to 12.5 km depth. Fault is projected such that regional stress maintains a nearly fixed angle relative to the fault. From Stein *et al.* (1997) Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International* 128: 594–604.

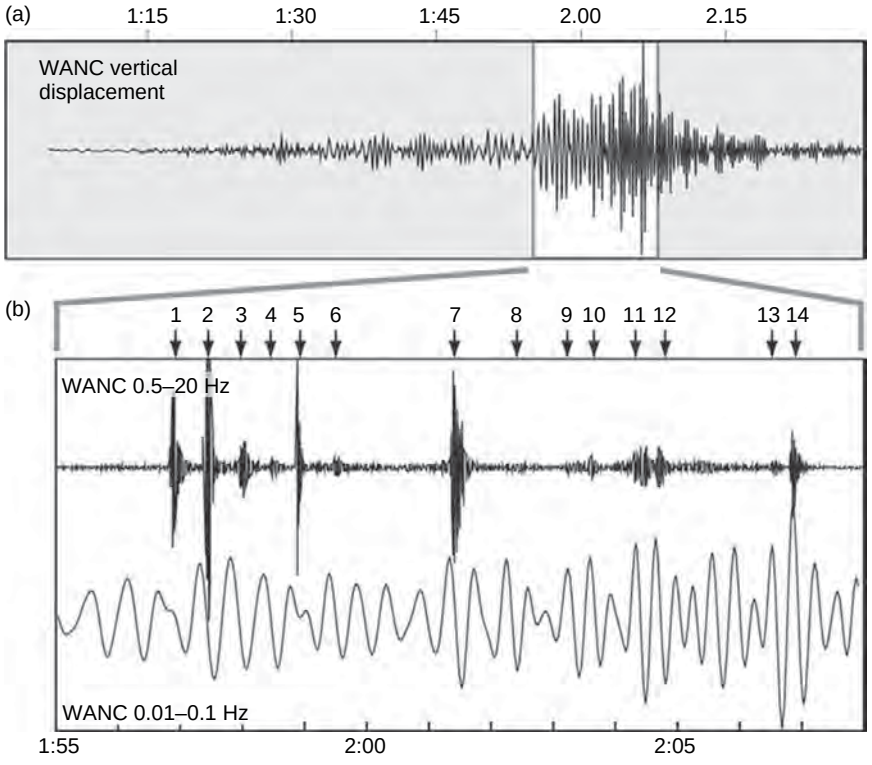


Figure 40 Seismograms showing ground motion from both Sumatra earthquake and local events triggered at Mount Wrangell, Alaska (West *et al.*, 2005). (a) Vertical component displacement at short-period station. Large-amplitude Rayleigh phases arrived in Alaska 1 h after the Sumatra earthquake. (b) Expanded view of surface wave displacement at WANC. Upper record filtered from 0.5 to 20 Hz to emphasize high-frequency local earthquakes (indicated by numbers 1–14). Lower record is filtered from 0.01 to 0.1 Hz to emphasize teleseismic ground motion. Local, triggered earthquake correlate with large-amplitude vertical excursions of the teleseismic signal where stress is expected to be extensional. From West M, Sanchez JJ, and McNutt SR (2005) Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake. *Science* 308: 1144–1146 (doi:10.1126/science.1112462).

4.01.6.6 Temporal Distribution of Earthquakes

Large earthquakes are typically followed by numerous aftershocks. The rate of aftershocks, R, as a function of time, t, following a large earthquake typically decays in time following a relation known as Omori's law, which in its modified form (Utsu, 1961) can be written as

$$R(t) \sim (t + c)^{-p}$$

where c is a constant and the exponent, p, is usually observed to be ~1. Aftershocks are usually thought to occur in response to the stress change imposed by the main shock they follow (Scholz, 1990). The mechanism behind the temporal decay of Omori's law has been attributed to many physical effects. Elastic effects act immediately, and do not explain the gradual decay. Among the mechanisms that have been proposed to explain the temporal decay of aftershock rate are: pore fluid flow (Nur and Booker, 1972), viscoelastic relaxation (Freed and Lin, 2001), stress corrosion (Scholz, 1968), and earthquake nucleation under rate- and state-variable friction (Dieterich, 1994).

Since dynamic stresses can trigger earthquakes at large distance, it stands to reason that they can trigger earthquakes at short distances as well. This might explain the diverse aftershocks of the 1989 Loma Prieta earthquake and the observation that they were not favored by the main-shock-induced stress change (Beroza and Zoback, 1993). Felzer and Brodsky (2006) showed that the distance decay of aftershocks was consistent with dynamic triggering, even for small, local earthquakes.

Whether by static triggering, dynamic triggering, or other processes such as poroelastic effects, it is clear that earthquakes interact with each other. That is, the occurrence of one earthquake affects the probability of other earthquakes. This is true not only for large earthquakes,

but for small earthquakes as well. Models that attempt to document and quantify the consequences of these interactions have come to be known by the acronym ETAS, which stands for epidemic-type aftershock sequence (Ogata, 1988). These and other seismicity models are covered in more detail in Chapter 4.23.

4.01.7 Associated Problems

4.01.7.1 Strong Motion Prediction

From a societal perspective, the prediction of strong ground motion is arguably one of the most important issues that seismologists can address. Much of the risk that earthquakes pose, particularly to property, could not be mitigated by short-term prediction, which is an area of seismological research that is afforded a lot more attention by the general public. Even if short-term earthquake prediction should someday prove possible and reliable, it would not be possible, for example, to retrofit large engineered structures on a timescale of days to weeks.

The wave propagation properties of the Earth's crust are strongly heterogeneous on all scales. As a result, the seismic wavefield becomes strongly distorted as it propagates through the crust. Compounding this wavefield complexity is the possible complexity of the earthquake source itself, which can be substantial for the large damaging earthquakes that are of concern in strong ground motion prediction. Understanding and predicting this variability in strong ground motion poses a major challenge to engineering seismologists and civil engineers who have to design structures based on anticipated ground motions.

So how is strong ground motion predicted? Work in this area bridges the disciplines of seismology and earthquake engineering and thus is termed 'engineering seismology'. Engineering seismology seeks to predict the intensity of ground shaking likely to occur in future earthquakes. The standard procedure for predicting strong ground motion proceeds as follows.

The first step is to decide what intensity measure is to be used, which is an engineering decision based on what information is needed for structural design. Commonly used intensity measures include PGA, peak ground velocity (PGV), spectral acceleration (Sa), spectral velocity (Sv), and spectral displacement (Sd). Then a data set of strong motion recordings is assembled. A great deal of care is taken in the process to ensure that the ground motion parameter of interest is not somehow obscured by particular characteristics of the recording site. Moreover, reliable information such as the distance to the fault plane and the geological characteristics of the recording site are required in order to take full advantage of the data.

Once these data are assembled, and the intensity measure is determined for them, we posit a parametric relationship between relevant aspects of the earthquake of interest and the strong ground motion intensity measure. These relationships are usually called 'attenuation relations' despite the fact that they are dominated by effects other than anelastic attenuation. Examples of the factors that are used to develop these relationships include magnitude, closest distance to fault rupture, earthquake mechanism, and the nature of the soils at the site of interest. Next, a regression between these factors and the intensities observed in earthquakes is carried out in order to determine the coefficients of the attenuation relation.

The current scarcity of strong motion data at short distances from large earthquakes means that there are inadequate data to represent the hazard from the most dangerous events. Ground motion computer simulation provides a way to fill this gap in the data. Another argument in favor of ground motion simulation stems from the need of civil engineers for a more complete description of ground motion (i.e., seismograms) instead of a scalar intensity measurement, such as the spectral acceleration.

Simulation of near-fault ground motion that is accurate enough to be used to take the place of attenuation relations is an emerging technology that is not quite ready for application. A major source of uncertainty for these efforts remains an accurate characterization of the earthquake source and, in particular, the relationship between source parameters implied by rupture dynamics (Guatteri *et al.*, 2004). Also crucial to strong ground motion prediction is the ability to model path effects, that is, the effect of wave propagation through the complex geologic structure of the Earth's crust. This includes not only elastic wave propagation, but also loss of wave energy due to anelasticity, and changes to the wavefield due to nonlinear site effects. The approach of using small earthquakes as empirical Green's functions (eGf's) is a very promising avenue of research for the prediction of high-frequency strong ground motion (Joyner and Boore, 1986). The advantage of the eGf approach is that it includes the true complexity of wave propagation effects. In practice, its application is limited by the availability of suitable eGf events, and the possibility that nonlinearities present during the strong ground motion of the

mainshock will not affect the relatively weak motions of the eGf event.

4.01.7.2 Tsunamis

A tsunami is a long-period wavetrain that propagates as a shallow water wave in the open ocean. By 'shallow', we mean that the wavelength is much longer than the depth of the ocean, such that particle trajectories in the wave are nearly horizontal. This criterion is easily satisfied by earthquake-generated tsunamis, which can have wavelengths hundreds of kilometers long – an order of magnitude greater than the depth of even the deepest ocean trenches. Because they are shallow water waves, tsunamis propagate at high speed across the open ocean – at speeds that approach the speed of travel by commercial jets.

Earthquakes represent one of the primary triggers of tsunamis, but tsunamis can be generated in many ways. The permanent deformation of the seafloor due to an earthquake is the source of some of the largest, most destructive tsunamis, but other important potential tsunami sources include submarine landslides, volcanic eruptions, and bolide impacts. For some earthquakes, such as the 1946 earthquake in the Aleutian Islands, tsunami modeling suggests that landslides triggered by earthquakes may have been the principal tsunami source (Fryer *et al.*, 2004), though Lopez and Okal (2006) model the tsunami generation using a very large, but slow earthquake rupture. Clearly, both triggered landslides and deformation of the seafloor due to earthquake rupture must be considered capable of tsunami generation regardless of whether one particular tsunami arose from one source or the other. For a detailed treatment of tsunamis as they relate to earthquakes, *see* Chapter 4.17.

4.01.7.3 Test-Ban Treaty Verification

Since a nuclear bomb was first tested underground in 1957, seismology has played a key role in nuclear monitoring and test-ban treaty verification. Much of the impetus and funding to establish the WWSSN in the early 1960s and its successor the Global Seismic Network (GSN) nearly three decades later came not from the need to monitor earthquakes, but from the need to monitor nuclear testing. The proliferation of nuclear weapons to additional countries, and the reduction in the target monitoring threshold from 150 kiloton of TNT to 10 kiloton, required a truly global seismological network that spans all the continents and records data over a broad range of frequencies. The multilateral nature of the Comprehensive Test Ban Treaty also meant that the data had to be openly available to all interested parties. This combination has served the purposes of test-ban treaty verification research, but it has also been a bonanza to earthquake seismology in general. It is also fair to say that earthquake monitoring has been an important source of sustenance for these same networks.

There are a number of issues that arise in test-ban treaty verification research. First, and perhaps foremost, is the need to detect and locate small magnitude events that represent possible low-yield nuclear tests. The second task is to discriminate explosions from earthquakes (and, in addition, nuclear tests from other explosions). **Figure 41** shows a comparison of a seismogram from an earthquake with one from an explosion. The relatively efficient generation of P waves relative to S waves is one means for discriminating explosions from earthquakes. A third task is to assess the feasibility of various evasion scenarios in which a clandestine test could be sufficiently decoupled from the Earth, or somehow disguised such that it would go unreported. As the size threshold decreases, all of these tasks become more difficult due to reduced signal-to-noise ratio, a larger number of small earthquakes, and an increasing population of ordinary chemical explosions in the same size range.

4.01.7.4 Solid Earth–Atmospheric Coupling

Seismic waves are generated not only by earthquakes, but also by other sources. Volcanoes, for example, are obviously prodigious sources of earthquakes as well as other seismic sources such as harmonic tremor. Large landslides on volcanoes are another source. All of these represent processes within the solid earth. Vigorous volcanic eruptions disturb the atmosphere sufficiently that the acoustic and gravitational waves in the atmosphere couple to the solid Earth and excite seismic waves (Widmer and Zürn, 1992; Kanamori and Mori, 1992).

The impedance of the atmosphere is much smaller than that of the solid Earth. As a result, a displacement of the Earth's surface will translate more or less directly into a displacement of the atmosphere. This effect is so small that it is usually neglected, but there are circumstances in which it is important and has a detectable effect. The density of the atmosphere decreases exponentially with increasing altitude. As a result, a perturbation that is modest at the Earth's surface, grows dramatically as it propagates upward through the atmosphere. For a

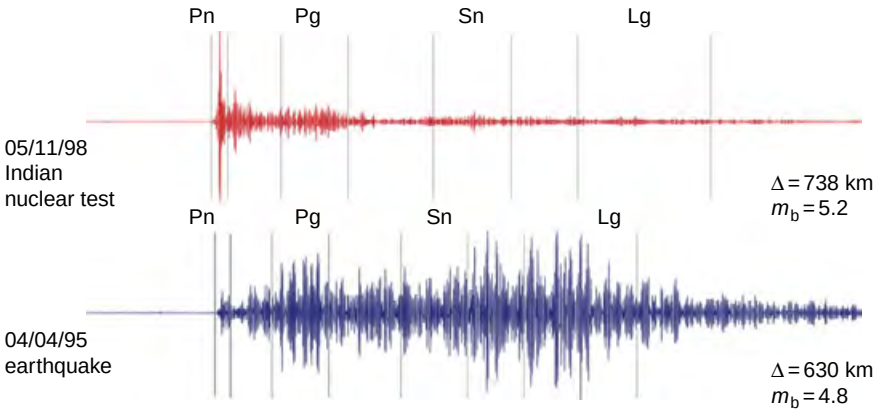


Figure 41 Seismograms recorded on the same instrument at Nilore, Pakistan. Upper seismogram shows ground motion due to an Indian nuclear test. Lower seismogram shows seismogram due to an earthquake at a similar distance. The phases Pn, Pg, Sn, and Lg are shown with vertical lines. Note that the S-wave arrivals from the earthquake are systematically much larger than for the explosion. Clues such as these help to discriminate underground nuclear tests from earthquakes. From Walter WR, Rodgers AJ, Mayeda K, Myers SC, Pasyanos M, and Denny M (1998) Preliminary regional seismic analysis of nuclear explosions and earthquakes in Southwest Asia. In *Proceedings of the 20th Annual Seismic Research Symposium on Monitoring a Comprehensive Test Ban Treaty*, Department of Defense and Department of Energy Report, 442–452.

major earthquake, this leads to a large and easily detectable change in the ionosphere. Because the ionosphere has a strong effect on electromagnetic waves propagating through it, changes in the ionosphere caused by earthquakes can be detected using, for example, transmissions from GPS satellites (Calais and Minster, 1995). The displacements caused by a large tsunami have also generated detectable ionospheric perturbations (Artru *et al.*, 2005).

The continuous excitation of the very-long-period free oscillations (Suda *et al.*, 1998) also appears to result from coupling between the Earth and its atmosphere and/or oceans. The basic observation is that the Earth's very-low-frequency fundamental spheroidal modes of oscillations are excited even during periods where there are no earthquakes to account for them. The excitation has an annual signature that strongly suggests it is not driven by processes within the solid Earth. Early explanations for this excitation focused on atmospheric disturbances as the cause (Kobayashi and Nishida, 1998), but more recent analysis suggests that excitation by oceanic infragravity waves (**Figure 42**) is more likely (Rhie and Romanowicz, 2004). For more on this subject, *see* Chapter 4.14.

4.01.8 Earthquake Risk Mitigation

Efficient mitigation of the risk posed by earthquakes requires as a foundation a quantitative measurement of the hazard that earthquakes pose. Once this is understood, informed policy decisions on building codes and their enforcement and on priorities for retrofitting susceptible structures can be made. Much of the work in earthquake hazard analysis casts strong ground motion prediction as a probabilistic statement in a process called probabilistic seismic hazard analysis (PSHA).

PSHA was developed to integrate multiple factors that control the likelihood and strength of shaking to be expected, each of which can be treated as a stochastic process. PSHA asks, for example, what is the ground motion that has a 2% probability of being exceeded over a 50 year time interval? To determine this number, and to map it spatially, requires combining information on earthquake likelihoods for all relevant earthquakes sources (i.e., active nearby faults) with strong motion attenuation laws that include site effects. The output of PSHA is a map of ground motion intensities at a specified exceedence probability that can be used in building codes to develop design criteria for buildings, and by policymakers to develop priorities for earthquake risk reduction. **Figure 43** shows an example calculation for the entire planet.

Another important area in seismic risk reduction is earthquake early warning. Earthquake early-warning systems rely on the fact that although seismic waves travel at high speed, information that an earthquake has occurred can be transmitted at the speed of light, and that an alert that a significant earthquake is underway can be transmitted at the speed of light,

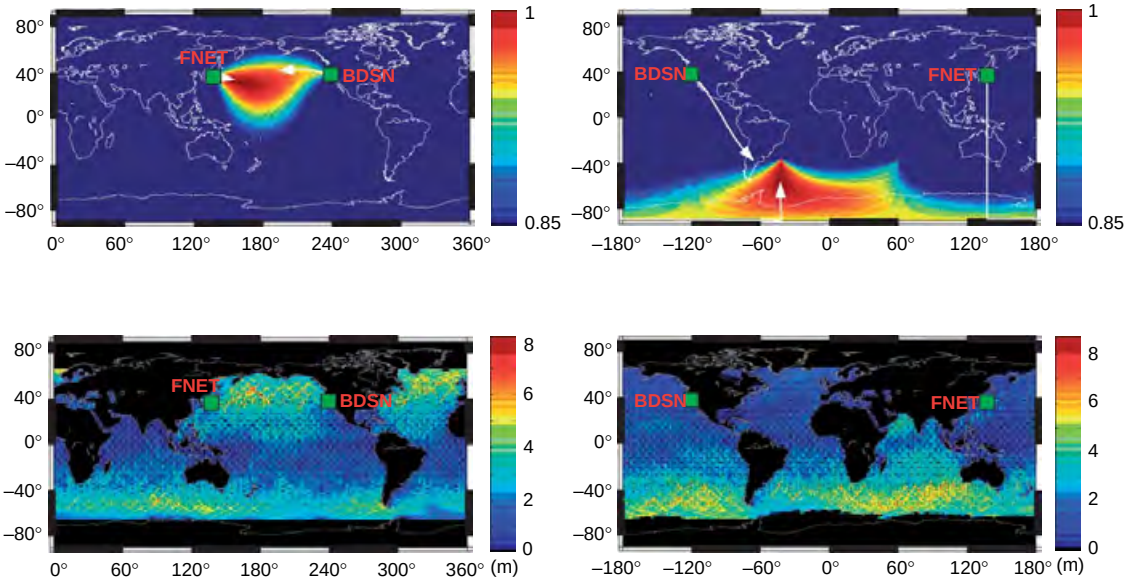


Figure 42 Localization of the source of background excitation of the Earth's free oscillations (Rhie and Romanowicz, 2004) using arrays in Japan and the United States. Arrows in upper panels point from the arrays to the source of excitation (red area) for two different time periods. The excitation on the left is for the northern Pacific and on the right for the southern Atlantic. Sources are more active in boreal and austral winter seasons. Lower panels show month-long significant wave heights for those seasons from *TOPEX/Poseidon* measurements.

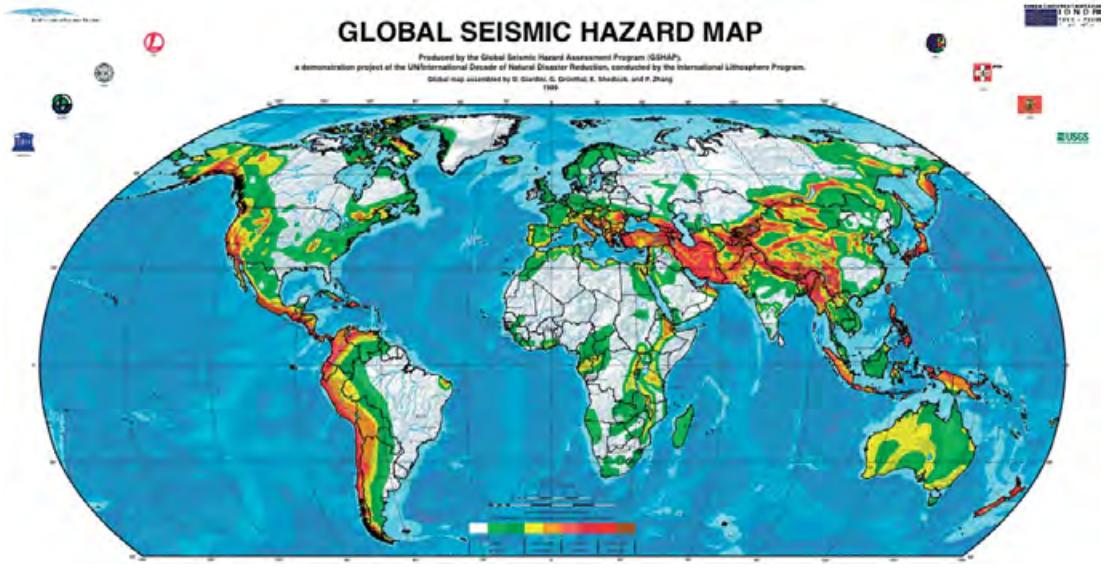


Figure 43 Global seismic hazard map. Areas of red are of highest seismic hazard while clear areas are of lowest seismic hazard. Seismic hazard is dominated by plate boundary seismicity, except in continental collision zones – most notably the Alpide Belt. From Giardini *et al.* (1999) The GSHAP global seismic hazard map. Annali di Geofisica 42: 1225–1230.

and an alert that a significant earthquake is underway can be issued before damaging seismic waves arrive. Earthquake early-warning systems are operational in Japan, Mexico, Taiwan, and are in the planning stages in Romania, Turkey, and the United States. The amount of warning that such systems can provide for large earthquakes under favorable circumstances can be several tens of seconds. For more on seismic hazard and real-time warning systems, *see* Chapter 4.21.

4.01.9 Conclusions

In this chapter, we have taken a synoptic view of earthquake seismology as well as several closely related fields of study. In the chapters that follow, many of the same topics are covered, but in much greater detail. The field of earthquake seismology is rapidly evolving. This evolution is driven by several factors. Perhaps the most important single factor is rapid progress in the quality, density, and wide availability of seismic data. Seismic data centers, which house many terabytes of seismic data, and their easy availability via the Internet, have resulted in access for interested seismologists to literally millions of seismograms. These large data volumes enable fundamentally new ways of looking at earthquakes. They also mean that visual analysis of seismograms, except on a proof-of-concept basis, is rapidly becoming a thing of the past. The computational capabilities of computers have kept pace with the rapid rise in data availability. As data collection and computational resources continue a rapid upward trajectory, imaginative seismologists are presented with an unprecedented opportunity to pose, and to answer, new questions about the earthquake process.

As a result, we can look forward to rapid progress in the fields of earthquake and volcano seismology. This progress will be punctuated by large, devastating earthquakes. As in the past, important aspects of these earthquakes will not be expected, which testifies to the fact that earthquake science remains a young and exciting field of research. The havoc future earthquakes will wreak will also serve to remind seismologists, and those who depend on their research, of the profound societal relevance of earthquake science.

References

Aagaard BT and Heaton TH (2004) Near-source ground motions from simulations of sustained intersonic and supersonic fault ruptures. *Bulletin of the Seismological Society of America* 94: 2064–2078.

Abercrombie RE (1995) Earthquake source scaling relationships from −1 to 5 ML, using seismograms recorded at 2.5 km depth. *Journal of Geophysical Research* 100: 24015–24036.

Abercrombie RE and Rice JR (2005) Can observations of earthquake scaling constrain slip weakening? *Geophysical Journal International* 162: 406–424 (doi: 10.1111/j.1365-246X.2005.02579.x).

Aki K (1966) Generation and propagation of G waves from the Niigata earthquake of June 16, 1964: Part 2. Estimation of earthquake moment, released energy and stress drop from the G wave spectra. *Bulletin of the Earthquake Research Institute, University of Tokyo* 44: 73–88.

Anderson JA and Wood HO (1925) Description and theory of the torsion seismometer. *Bulletin of the Seismological Society of America* 15: 1–72.

Andrews DJ (1976) Rupture velocity of plane strain shear cracks. *Journal of Geophysical Research* 81: 5679–5687.

Andrews DJ (2005) Rupture dynamics with energy loss outside the slip zone. *Journal of Geophysical Research* 110: (doi:01310.01029/02004JB003191).

Andrews DJ and Ben-Zion Y (1997) Wrinkle-like slip pulse on a fault between different materials. *Journal of Geophysical Research* 102: 553–571.

Argus DF and Gordon RG (2001) Present tectonic motion across the Coast Ranges and San Andreas fault system in central California. *Geological Society of America Bulletin* 113: 1580–1592.

Artru J, Ducic V, Kanamori H, Lognonne P, and Murakami M (2005) Ionospheric gravity waves induced by tsunamis. *Geophysical Journal International* 160: 840–848 (doi:10.1111/j.1365-246X.2005.02552.x).

Astiz L, Lay T, and Kanamori H (1988) Large intermediate-depth earthquakes and the subduction process. *Physics of the Earth and Planetary Interiors* 53: 80–166 (doi: 10.1016/0031-9201(88)90138..0).

Arvidsson R (1996) Fennoscandian earthquakes: Whole crustal rupturing related to postglacial rebound. *Science* 274: 744–746.

Backus G and Mulcahy M (1976) Moment tensors and other phenomenological descriptions of seismic sources II. Discontinuous displacements. *Geophysical Journal of the Royal Astronomical Society* 47: 301–329.

Bak P and Tang C (1989) Earthquakes as a self-organized critical phenomenon. *Journal of Geophysical Research* 94: 15635–15637.

Bakun WH and Wentworth CM (1997) Estimating earthquake location and magnitude from intensity data. *Bulletin of the Seismological Society of America* 877: 1502–1521.

Barazangi M and Dorman J (1969) World seismicity maps compiled from ESSA, Coast and Geodetic Survey, epicenter data, 1961–1967. *Bulletin of the Seismological Society of America* 59: 369–380.

Beck ME (1983) On the mechanism of tectonic transport in zones of oblique subduction. *Tectonophysics* 93: 1–11.

Ben-Zion Y, Peng Z, Okaya D, *et al.* (2003) A shallow fault-zone structure illuminated by trapped waves in the Karadere-Duzce branch of the North Anatolian Fault, Turkey. *Geophysical Journal International* 152: 699–717.

Beroza GC and Jordan TH (1990) Searching for slow and silent earthquakes using free oscillations. *Journal of Geophysical Research* 95: 2485–2510.

Beroza GC and Zoback MD (1993) Mechanism diversity of the Loma Prieta aftershocks and the mechanics of mainshock–aftershock interaction. *Science* 259: 210–213.

Bilham R (1995) Global fatalities in the past 2000 years: Prognosis for the next 30. In: Rundle J, Klein F, and Turcotte D (eds.) *Santa Fe Institute Studies in the Sciences of Complexity,* 25: *Reduction and Predictability of Natural Disasters*, pp. 19–31. Boston, MA: Addison-Wesley.

Boatwright J and Bundock H (2005) Modified Mercalli Intensity maps for the 1906 San Francisco earthquake plotted in shake-map format. *US Geological Survey Open-File Report* 2005–1135.

Bolt BA (1993) *Abridged Modified Mercalli Intensity Scale, Earthquakes*, Appendix C, 331 pp. New York: W.H. Freeman and Company.

Booker JR (1974) Time dependent strain following faulting of porous medium. *Journal of Geophysical Research* 79: 2037–2044.

Bouchon M, Bouin MP, Karabulut H, Toksoz MN, and Dietrich M (2001) How fast does rupture propagate during an earthquake? New insights from the 1999 Turkey earthquakes. *Geophysical Research Letters* 28: 2723–2726.

Bouchon M and Vallée M (2003) Observation of long supershear rupture during the magnitude 8.1 Kunlunshan earthquake. *Science* 301: 824–826.

Brune JN (1968) Seismic moment, seismicity, and rate of slip along major fault zones. *Journal of Geophysical Research* 73: 777–784, 1968.

Brune JN, Henyey TL, and Roy RF (1969) Heat flow, stress and rate of slip along San Andreas fault, California. *Journal of Geophysical Research* 74: 3821–3827.

Byerleee JD (1978) Friction of rocks. *Pure and Applied Geophysics* 116: 615–626.

Calais E and Minster JB (1995) GPS detection of ionospheric perturbations following the January 17, 1994, Northridge earthquake. *Geophysical Research Letters* 22: 1045–1048.

Campbell DL (1978) Investigation of the stress-concentration mechanism for intraplate earthquakes. *Geophysical Research Letters* 5: 477–479.

Chapple WM and Forsyth DW (1979) Earthquakes and bending of plates at trenches. *Journal of Geophysical Research* 84: 6729–6749.

Chouet B (1988) Resonance of a fluid-driven crack: Radiation properties and implications for the source of long-period events and harmonic tremor. *Journal of Geophysical Research* 93: 4375–4400.

Chouet B (1996) Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature* 380: 309–316.

Chung W-Y and Kanamori H (1976) Source process and tectonic implications of the Spanish deep-focus earthquake of March 29, 1954. *Physics of the Earth and Planetary Interiors* 13: 85–96.

D'Alessio MA, Blythe AE, and Burgmann R (2003) No frictional heat along the San Gabriel fault, California: Evidence from fission-track thermochronology. *Geology* 31: 541–544 (DOI: 10.1130/0091-7613).

Dieterich JH (1972) Time-dependent friction in rocks. *Journal of Geophysical Research* 77: 3690–3697.

Dieterich JH (1978) Time-dependent friction and the mechanics of stick-slip. *Pure and Applied Geophysics* 116: 790–806.

Dieterich JH (1979) Modeling of rock friction, 1. Experimental results and constitutive equations. *Journal of Geophysical Research* 84: 2161–2168.

Dieterich JH (1979) Modeling of rock friction, 2. Simulation of preseismic slip. *Journal of Geophysical Research* 84: 2169–2175.

Dieterich JH (1994) A constitutive law for rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research* 99: 2601–2618.

Dragert H, Wang K, and James TS (2001) A silent slip event on the deeper Cascadia subduction interface. *Science* 292: 1525–1528.

Dunham EM and Archuleta RJ (2004) Evidence for a supershear transient during the 2002 Denali fault earthquake. *Bulletin of the Seismological Society of America* 94: S256–S268.

Dziewonski AM, Chou T-A, and Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research* 86: 2825–2852.

Ellsworth WL and Beroza GC (1995) Seismic evidence for an earthquake nucleation phase. *Science* 268: 851–855.

Ellsworth WL, Lindh AG, Prescott WH, and Herd DG (1981) The 1906 San Francisco earthquake and the seismic cycle. In: Simpson DW and Richards PG (eds.) *Earthquake Prediction: An International Review*, Maurice Ewing Ser., vol. 4: pp. 126–140. Washington, DC: AGU.

Ellsworth WL, Malin PE, Imanishi K, *et al.* (2007) Seismology inside the fault zone: Applications to fault-zone properties and rupture dynamics. *Scientific Drilling Journal* (in press).

Engdahl ER, van der Hilst R, and Buland R (1998) Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America* 88: 722–743.

Fedotov SA (1965) Regularities of the distribution of strong earthquakes in Kamchatka, the Kurile islands, and northeast Japan. *Trudy Instituta Fiziki Zemli Academiya Nauk SSSR* 36: 66.

Felzer KR and Brodsky EE (2006) Decay of aftershock activity with distance indicates triggering by dynamic stress. *Nature* 441: 735–738.

Fialko Y and Khazan Y (2005) Fusion by earthquake fault friction: Stick or slip?. *Journal of Geophysical Research* 110: B12,407 (doi:10.1029/2005JB003,869).

Fialko Y, Sandwell D, Agnew D, Simons M, Shearer P, and Minster B (2002) Deformation on nearby faults induced by the 1999 Hector Mine earthquake. *Science* 297: 1858–1862.

Fitch TJ (1972) Plate convergenece, transcurrent faults, and internal deformation adjacent to southeast Asia and the western Pacific. *Journal of Geophysical Research* 77: 4432–4460.

Frankel A, Fletcher JB, Vernon FL, *et al.* (1986) Rupture characteristics and tomographic source imaging of $M_L \sim 3$ earthquakes near Anza, Southern California. *Journal of Geophysical Research* 91: 12633–12650.

Freed AM and Lin J (2001) Delayed triggering of the 1999 Hector Mine earthquake by viscoelastic stress transfer. *Nature* 411: 180–183.

Freund LB (1990) *Dynamic Fracture Mechanics*, 563 pp. Cambridge, U.K: Cambridge University Press.

Frohlich C (1989) The nature of deep focus earthquakes. *Annual Review of Earth and Planetary Sciences* 17: 227–254.

Fryer GJ, Watts P, and Pratson LF (2004) Source of the great tsunami of 1 April 1946: A landslide in the upper Aleutian forearc. *Marine Geology* 203: 201–218.

Fukao Y, Hori S, and Ukawa M (1983) A seismological constraint on the depth of the basalt-eclogite transition in a subducting oceanic crust. *Nature* 303: 413–415.

Giardini D, Grünthal G, Shedlock KM, and Zhang P (1999) The GSHAP global seismic hazard map. *Annali di Geofisica* 42: 1225–1230.

Gibowicz SJ, Young RP, Talebi S, and Rawlence DJ (1991) Source parameters of seismic events at the Underground Research Laboratory in Manitoba, Canada: scaling relations for events with moment magnitude smaller than -2. *Bulletin of the Seismological Society of America* 81: 1157–1182.

Gordon RG (1998) The plate tectonic approximation: Plate nonrigidity, diffuse plate boundaries, and global plate reconstructions. *Annual Review of Earth and Planetary Sciences* 26: 615–642.

Guatteri M, Mai PM, and Beroza GC (2004) A pseudo-dynamic approximation to dynamic rupture models for strong ground motion prediction. *Bulletin of the Seismological Society of America* 94(6): 2051–2063 (doi:10.1785/0120040037).

Guatteri M and Spudich P (2000) What can strong-motion data tell us about slip-weakening fault-friction laws. *Bulletin of the Seismological Society of America* 90: 98–116 (doi: 10.1785/0119990053).

Gutenberg B (1955) Magnitude determination for larger Kern County shocks, 1952; Effects of station azimuths and calculation methods. *California Division of Mines and Geology Bulletin* 171: 171–175.

Gutenberg B and Richter CF (1936) Magnitude and energy of earthquakes. *Science* 83: 183–185.

Gutenberg B and Richter CF (1942) Earthquake magnitude, intensity, energy, and acceleration. *Bulletin of the Seismological Society of America* 32: 163–191.

Gutenberg B and Richter CF (1954) *Seismicity of the Earth and Associated Phenomenon,* 2nd edn. Princeton: Princeton University Press.

Gutenberg B and Richter CF (1956) Magnitude and energy of earthquakes. *Annali di Geofisica* 9: 1–15.

Hagerty MT, Schwartz SY, Garces MA, and Protti M (2000) Analysis of seismic and acoustic observations at Arenal volcano, Costa Rica, 1995–1997. *Journal of Volcanology and Geothermal Research* 101: 27–65.

Hanks TC and Kanamori H (1979) A moment magnitude scale. *Journal of Geophysical Research* 84: 2348–2350.

Harris R (1998) Introduction to special section: Stress triggers, stress shadows, and implications for seismic hazard. *Journal of Geophysical Research* 103: 24347–24358 (doi: 10.1029/98JB01576).

Healy JH, Rubey WW, Griggs DT, and Raleigh CB (1968) The Denver Earthquakes. *Science* 161: 1301–1310.

Heaton TH (1990) Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors* 64: 1–20.

Hill DP, Reasenberg PA, Michael A, *et al.* (1993) Seismicity in the western United States remotely triggered by the M 7.4 Landers, California, earthquake of June 28, 1992. *Science* 260: 1617–1623.

Hirose H and Obara K (2005) Repeating short- and long-term slow slip events with deep tremor activity around the Bungo channel region, southwest Japan. *Earth Planets Space* 57: 961–972.

Hubbert MK and Rubey WW (1959) Role of fluid pressure in mechanics of overthrust faulting 1. Mechanics of fluid-filled porous solids and its application to overthrust faulting. *GSA Bulletin* 70: 115–166.

Huixian L, Housner GW, Lili X, and Duxin H (2002) *The Great Tangshan Earthquake of 1976*, Technical Report: CaltechEERL:EERL.2002.001. California Institute of Technology.

Ida Y (1972) Cohesive force across the tip of a longitudinal-shear crack and Griffith's specific surface energy. *Journal of Geophysical Research* 77: 3796–3805.

Ide S and Beroza GC (2001) Does apparent stress vary with earthquake size? *Geophysical Research Letters* 28: 3349–3352.

Ide S, Beroza GC, and McGuire JJ (2005) Imaging earthquake source complexity. In: Levander A and Nolet G (eds.) *Geophysical Monograph Series, 157: Seismic Earth: Array Analysis of Broadband Seismograms*. Washington, DC: AGU.

Ide S and Takeo M (1997) Determination of constitutive relations of fault slip based on seismic wave analysis. *Journal of Geophysical Research* 102: 27379–27392 (10.1029/97JB02675).

Ishii M, Shearer PM, Houston H, and Vidale J (2005) Extent, duration and speed of the 2004 Sumatra–Andaman earthquake imaged by the Hi-Net array. *Nature* 435: 933–936.

Ito Y and Obara K (2005) Very low frequency earthquakes excited by the 2004 off Kii Peninsula earthquakes: A dynamic deformation process in the large accretionary prism. *Earth Planets Space* 57: 321–326.

Ito Y and Obara K (2006) Dynamic deformation of the accretionary prism excites very low frequency earthquakes. *Geophysical Research Letters* 33: (doi:10.1029/2005GL025270).

Jeffreys H (1942) On the mechanics of faulting. *Geological Magazine* 79: 291–295.

Johnston AC and Kanter LR (1990) Earthquakes in stable continental crust. *Scientific American* 262(3): 68–75.

Jost ML, Bußelberg T, Jost Ö, and Harjes H-P (1998) Source parameters of injection-induced microearthquakes at 9 km depth at the KTB deep drilling site, Germany. *Bulletin of the Seismological Society of America* 88: 815–832.

Joyner WB and Boore DM (1986) On simulating large earthquakes by Green's function addition of smaller earthquakes. In: Das S, Boatwright J, and Scholz CH (eds.) *Earthquake Source Mechanics*, pp. 269–274. Washington, DC: American Geophysical Union.

Julian BR (1994) Volcanic tremor: Nonlinear excitation by fluid flow. *Journal of Geophysical Research* 99: 11859–11878.

Julian BR (2002) Seismological detection of slab metamorphism. *Science* 296: 1625–162.

Julian BR, Miller AD, and Foulger GR (1998) Non-double-couple earthquakes 1. Theory. *Reviews of Geophysics* 36: 525–549.

Kagan YY and Jackson DD (1991) Seismic gap hypothesis: Ten years after. *Journal of Geophysical Research* 96: 21419–21431.

Kanamori H (1972) Mechanism of tsunami earthquakes. *Physics of the Earth and Planetary Interiors* 6: 346–359.

Kanamori H (1977) The Energy Release in Great Earthquakes. *Journal of Geophysical Research* 82: 2981–2987.

Kanamori H and Given JW (1982) Analysis of long-period seismic waves excited by the May 18, 1980, eruption of Mount St. Helens: A terrestrial monopole. *Journal of Geophysical Research* 87: 5422–5432.

Kanamori H and Heaton TH (2000) Microscopic and macroscopic physics of earthquakes. In: Rundle JB, Turcotte DL, and Klein W (eds.) *Geophysical Monograph 120: Geophysics and the Complexity of Earthquakes*, pp. 147–163. Washington, DC: AGU.

Kanamori H, Hauksson E, Hutton KL, and Jones LM (1993) Determination of earthquake energy release and ML using TERRAscope. *Bulletin of the Seismological Society of America* 83: 330–346.

Kanamori H and Mori J (1992) Harmonic excitation of mantle Rayleigh waves by the 1991 eruption of Mount Pinatubo, Philippines. *Geophysical Research Letters* 19: 721–724.

Kane MF (1977) Correlation of major eastern earthquake centers with mafic/ultramafic basement masses, *US Geological Survey Professional Paper* 1028-O.

Kano Y, Mori J, Fujio R, *et al.* (2006) Heat Signature on the Chelungpu Fault Associated with the 1999 Chi–Chi, Taiwan Earthquake. *Geophysical Research Letters* 33: L14306 (doi:10.1029/2006GL026733).

Kelleher JA, Sykes LR, and Oliver J (1973) Possible criteria for predicting earthquake locations and their applications to major plate boundaries of the Pacific and Caribbean. *Journal of Geophysical Research* 78: 2547–2585.

Kirby S, Engdahl ER, and Denlinger R (1996) Intermediate-depth intraslab earthquakes and arc volcanism as physical expressions of crustal and uppermost mantle metamorphism in subducting slabs. In: Bebout GE, *et al.* (eds.) *American Geophysical Union Geophysical Monograph 96: Subduction: Top to Bottom*, pp. 195–214, 2043 (doi:10.1029/2001JB000522). Washington, DC: AGU.

Kirby SH, Stein S, Okal EA, and Rubie DC (1996) Metastable mantle phase transformations and deep earthquakes in subducting oceanic lithosphere. *Reviews of Geophysics* 34: 261–306.

Kobayashi N and Nishida K (1998) Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature* 395: 357–360 (doi:10.1038/26427).

Kostoglodov V, Singh SK, Santiago JA, Larson KM, Lowry AR, and Bilham R. A large silent earthquake in the Guerrero seismic gap, Mexico. *Geophysical Research Letters* 30, (doi:10.1029/2003GL017219).

Kostrov BV and Das S (1988) Principles of earthquake source mechanics. In: Batchelor GK (ed.) *Cambridge Monographs on Mechanics and Applied Mathematics*, 286 pp. Cambridge: Cambridge University Press.

Lachenbruch A and Sass J (1980) Heat flow and energetics of the San Andreas fault zone. *Journal of Geophysical Research* 85: 6185–222.

Lapusta N, Rice JR, Ben-Zion Y, and Zheng G (2000) Elastodynamic analysis for long deformation history with spontaneous rupture episodes on faults with rate- and state-dependent friction. *Journal of Geophysical Research* 105: 23765–23789.

Leary P, Li YG, and Aki K (1987) Observations and modeling of fault-zone fracture anisotropy I. P, SV, SH travel times. *Geophysical Journal of the Royal Astronomical Society* 91: 461–484.

Li YG, Leary P, Aki K, and Malin P (1990) Seismic trapped modes in the Oroville and San Andreas fault zones. *Science* 249: 763–766.

Li YG, Vidale JE, Aki K, and Xu F (2000) Depth-dependent structure of the Landers fault zone from trapped waves generated by aftershocks. *Journal of Geophysical Research* 105: 6237–6254.

Li YG, Vidale JE, Aki K, Xu F, and Burdette T (1998) Evidence of shallow fault zone strengthening after the 1992 M 7.5 Landers, California, earthquake. *Science* 279: 217–219.

Linde AT, Gladwin MT, Johnston MJS, and Gwyther RL (1996) A slow earthquake srequence on the San Andreas fault. *Nature* 383: 65–69.

Lopez AM and Okal EA (2006) A seismological reassessment of the source of the 1946 Aleutian 'tsunami'earthquake. *Geophysical Journal International* 165: 835–849.

Mai PM and Beroza GC (2002) A spatial random-field model to characterize complexity in earthquake slip. *Journal of Geophysical Research* 107(B11): 2308 (doi:10.1029/2001JB000588).

Mallet R and Mallet JW (1858) *The earthquake catalogue of the British Association, with the discussion, curves, and maps, etc*. Transaction of the British Association for the advancement of Science, London.

Marone C and Scholz CH (1988) The depth of seismic faulting and the upper transition from stable to unstable slip regimes. *Geophysical Research Letters* 15: 621–624.

Mayeda K and Walter WR (1996) Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes. *Journal of Geophysical Research* 101: 11195–11208.

McCaffrey R (1992) Oblique plate convergence, slip vectors, and forearc deformation. *Journal of Geophysical Research* 97: 8905–8915.

McCann WR, Nishenko SR, Sykes LR, and Krause J (1979) Seismic gaps and plate tectonics: Seismic potential for major plate boundaries. *Pure and Applied Geophysics* 117: 1082–1147.

McGuire JJ, Zhao L, and Jordan TH (2002) Predominance of unilateral rupture for a global catalog of large earthquakes. *Bulletin of the Seismological Society of America* 92: 3309–3317.

McKenzie D and Brune JN (1972) Melting on fault planes during large earthquakes. *Geophysical Journal of the Royal Astronomical Society* 29: 65–78.

McNally KC and McEvilly TV (1977) Velocity contrast across the San Andreas Fault in central California: Small scale variations from P-wave nodal plane distortion. *Bulletin of the Seismological Society of America* 67: 1565–1576.

McNutt SR (2005) Volcanic seismology. *Annual Review of Earth and Planetary Sciences* 32: 461–491 (doi:10.1146/annurev.earth.33.092203.122459).

Michael AJ (1990) Energy constraints on kinematic models of oblique faulting: Loma Prieta versus Parkfield-Coalinga. *Geophysical Research Letters* 17: 1453–1456.

Miller MM, Melbourne T, Johnson DJ, and Sumner WQ (2002) Periodic slow earthquakes from the Cascadia subduction zone. *Science* 295: 2423.

Miller SA, Collettini C, Chiaraluce L, Cocco M, Barchi M, and Kaus BJ (2004) Aftershocks driven by a high-pressure CO2 source at depth. *Nature* 427: 724–727.

Molnar P, Freedman D, and Shih JSF (1979) Lengths of intermediate and deep seismic zones and temperatures in downgoing slabs of lithosphere. *Geophysical Journal International* 56: 41–54.

Montelli R, Nolet G, Dahlen FA, Masters G, Engdahl ER, and Hung S-H (2004) Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* 303: 338–343 (doi: 10.1126/science.1092485).

Nadeau RM and Dolenc D (2004) Nonvolcanic tremors deep beneath the San Andreas fault. *Science* 307: 389.

Nadeau RM and Johnson LR (1998) Seismological studies at Parkfield VI: Moment release rates and estimates of source parameters for small repeating earthquakes. *Bulletin of the Seismological Society of America* 88: 790–814.

Nason RD and Tocher D (1970) Measurement of movement on the San Andreas fault. In: Mansinha L, Smylie DE, and Beck AE (eds.) *Earthquake Displacement Fields and the Rotation of the Earth*, pp. 246–254. Dordrecht, Netherlands: Reidel.

Nur A and Booker JR (1972) Aftershocks caused by pore fluid flow? *Science* 175: 885–887.

Nur A and Ron H (1997) Armageddon's earthquakes. *International Geology Review* 39: 532–541.

Obara K (2002) Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* 296: 1679–1681.

Obara K, Hirose H, Yamamizu F, and Kasahara K (2004) Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone. *Geophysical Research Letters* 31: L23602 (doi:10.1029/2004GL020848).

Okal EA (1992) Use of the mantle magnitude M_m for the reassessment of the moment of historical earthquakes I: Shallow events. *Pure and Applied Geophysics* 139: 17–57.

Okal E and Stewart LM (1982) Slow earthquakes along oceanic fracture zones: Evidence for asthenospheric flow away from hotspots? *Earth and Planetary Science Letters* 57: 75–87.

Ogata Y (1998) Statistical-models for earthquake occurrences and residual analysis for point-processes. *Journal of the American Statistical Association* 83: 9–27.

Ozawa S, Murakami M, Kaidzu M, and Hatanaka Y (2005) Transient crustal deformation in Tokai region, central Japan, until May 2004. *Earth Planets Space* 57: 909–915.

Parsons T, Toda S, Stein RS, Barka A, and Dieterich JH (2000) Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation. *Science* 288: 661–665.

Pérez-Campos X and Beroza GC (2001) Mechanism-dependent scaling of the radiated seismic energy. *Journal of Geophysical Research* 106: 11127–11136.

Perrin G, Rice JR, and Zheng G (1995) Self-healing slip pulse on a frictional surface. *Journal of the Mechanics and Physics of Solids* 43: 1461–1495.

Peyrat S, Olsen KB, and Madariaga R (2001) Dynamic modeling of the 1992 Landers earthquake. *Journal of Geophysical Research* 106: 26467–26482.

Poupinet G and Ellsworth WL (1984) Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California. *Journal of Geophysical Research* 89: 5719–5732.

Quin H (1990) Dynamic stress drop and rupture dynamics of the October 15, 1979 Imperial Valley, California earthquake. *Tectonophysics* 175: 93–117.
Raleigh CB, Healy JH, and Bredehoeft JD (1976) An experiment in earthquake control at Rangely, Colorado. *Science* 191: 1230–1237.
Reinecker J, Heidbach O, Tingay M, Sperner B, and Müller B (2005) The release 2005 of the World Stress Map (available online at http://www.world-stress-map.org).
Rempel AW and Rice JR (2006) Thermal pressurization and onset of melting in fault zones. *Journal of Geophysical Research* 111: B09314 (doi: 10.1029/2006JB004314).
Rhie J and Romanowicz B (2004) Excitation of Earth's continuous free oscillations by atmosphere-ocean–seafloor coupling. *Nature* 431: 552–556 (doi: 10.1038/nature02942).
Richter CF (1935) An instrumental earthquake magnitude scale. *Bulletin of the Seismological Society of America* 25: 1–32.
Richter CF (1958) *Elementary Seismology*, 758 pp. San Francisco: W. H. Freeman and Company.
Rice JR (1980) The mechanics of earthquake rupture. In: Dziewonski AM and Boschi E (eds.) *Physics of the Earth's Interior*, pp. 555–649. North-Holland: Amsterdam.
Rice JR and Cleary MP (1976) Some basic stress diffusion solutions for fluid-saturated elastic porous media with compressible constituents. *Reviews of Geophysics and Space Physics* 14: 227–241.
Risk Management Solutions Inc. (1995) *What if the 1923 Earthquake Strikes Again? A Five Prefecture Tokyo Region Scenario*, 97 pp. Menlo Park: California.
Rodgers AJ and Walter WR (2002) Seismic discrimination of the May 11, 1998 Indian nuclear test with short-period regional data from station NIL (Nilore, Pakistan). *Pure and Applied Geophysics* 159: 679–700.
Rodgers G and Dragert H (2003) Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* 300: 1942–1943.
Rong Y, Jackson DD, and Kagan YY (2003) Seismic gaps and earthquakes. *Journal of Geophysical Research* 108: (doi: 10.1029/2002JB002334).
Rosakis AJ, Samudrala O, and Coker D (1999) Cracks faster than the shear wave speed. *Science* 284: 1337–1340.
Rubin AM, Gillard D, and Got J-L (1999) Streaks of microearthquakes along creeping faults. *Nature* 400: 635–641.
Ruina A (1983) Slip instability and state variable friction laws. *Journal of Geophysical Research* 88: 10359–10370.
Savage JC (1983) A dislocation model of strain accumulation aand release at a subduction zone. *Journal of Geophysical Research* 88: 4984–4996.
Sbar ML and Sykes LR (1973) Contemporary compressive stress and seismicity in eastern North America: An example of intra-plate tectonics. *Geological Society of America Bulletin* 84: 1861–1882.
Schaff DP, Bokelmann GHR, Beroza GC, Waldhauser F, and Ellsworth WL (2002) High resolution image of Calaveras Fault seismicity. *Journal of Geophysical Research* 107(B9): 2186 (doi:10.1029/2001JB000633).
Schaff DP, Bokelmann GHR, Ellsworth WL, Zanzerkia E, Waldhauser F, and Beroza GC (2004) Optimizing correlation techniques for improved earthquake location. *Bulletin of the Seismological Society of America* 94: 705–721.
Scholz CH (1968) Microfractures, aftershocks, and seismicity. *Bulletin of the Seismological Society of America* 58: 1117–1130.
Scholz CH (1989) Mechanics of faulting. *Annual Review of Earth and Planetary Sciences* 17: 309–334.
Scholz CH (1990) *The Mechanics of Earthquakes and Faulting* New York: Cambridge University Press.
Scholz CH (2000) Evidence for a strong San Andreas Fault. *Geology* 28: 163–166 (doi: 10.1130/0091-7613).
Scholz CH, Molnar P, and Johnson T (1972) Frictional sliding of granite and earthquake mechanism implications. *Journal of Geophysical Research* 77: 6392–6406.
Seno T and Yamasaki T (2003) Low-frequency tremors, intraslab and interplate earthquakes in Southwest Japan – from a viewpoint of slab dehydration. *Geophysical Research Letters* 30: (doi:10.1029/2003GL018349).
Shearer PM, Prieto GA, and Hauksson E (2006) Comprehensive analysis of earthquake source spectra in southern California. *Journal of Geophysical Research* 111: B06303 (doi:10.1029/2005JB003979).
Shelly DR, Beroza GC, Ide S, and Nakamula S (2006) Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature* 442: 188–191.
Sibson RH (1973) Interaction between temperature and pore-fluid pressure during earthquake faulting – A mechanism for partial or total stress relief. *Nature* 243: 66–68.
Sibson RH (1981) Fluid flow accompanying faulting: Field evidence and models. In: Simpson DW and Richards PG (eds.) *Maurice Ewing Series 4: Earthquake Prediction: An International Review*, pp. 593–603. Washington, DC: AGU.
Singh SK, Mena E, and Castro R (1988) Some aspects of source characteristics of the 19 September 1985 Michoacan earthquake and ground motion amplification in and near Mexico City from strong motion data. *Bulletin of the Seismological Society of America* 78: 451–477.
Singh SK and Ordaz M (1994) Seismic energy release in Mexican subduction zone earthquakes. *Bulletin of the Seismological Society of America* 84: 1533–1550.
Somerville PG, Smith NF, Graves RW, and Abrahamson NA (1997) Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismological Research Letters* 68: 199–222.
Stein RS (1999) The role of stress transfer in earthquake occurrence. *Nature* 402: 605–609 (doi:10.1038/45144).
Stein RS, Barka AA, and Dieterich JH (1997) Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International* 128: 594–604.
Stirling M, Wesnousky SG, and Shimizaki K (1996) Fault trace complexity, cumulative slip, and the shape of the magnitude-frequency distribution for strike-slip faults – A global survey. *Geophysical Journal International* 124: 833–868.
Stover CW and Coffman JL (1993) *Seismicity of the United States, 1568–1989 (Revised), US Geological Survey Professional Paper 1527*. Washington: United States Government Printing Office.
Suda N, Nawa K, and Fukao Y (1998) Earth's background free oscillations. *Science* 279: 2089–2091 (doi: 10.1126/science.279.5359.2089).
Sykes LR (1967) Mechanism of earthquakes and nature of faulting on the mid-oceanic ridges. *Journal of Geophysical Research* 72: 2131–2153.
Titus SJ, DeMets C, and Tikoff B (2006) Thirty-five-year creep rates for the creeping section of the San Andreas Fault and the effects of the 2004 Parkfield earthquake; constraints from alignment arrays, continuous global positioning system, and creepmeters. *Bulletin of the Seismological Society of America* 96: S250–S268 (doi: 10.1785/0120050811).
Toksöz MN, Shakal AF, and Michael AJ (1979) Space-time migration of earthquakes along the North Anatolian fault zone and seismic gaps. *Pure and Applied Geophysics* 117: 1258–1270.

Toppozada TR (1975) Earthquake magnitude as a function of intensity data in California and western Nevada. *Bulletin of the Seismological Society of America* 65: 1223–1238.
Toppozada TR, Real CR, and Park DL (1981) Preparation of isoseismal maps and summaries of reported effects for pre-1900 California earthquakes: 1–182.
Tse ST and Rice JR (1986) Crustal earthquake instability in relation to the depth variation of frictional slip properties. *Journal of Geophysical Research* 91: 9452–9472.
Unsworth MJ, Malin PE, Egbert GD, and Booker JR (1997) Internal Structure of the San Andreas fault at Parkfield, California. *Geology* 25: 359–362.
Utsu T (1961) A statistical study on the occurrence of aftershocks. *Geological Magazine* 30: 521–605.
Yu G, Wesnousky SG, and Ekstrom G (1993) Slip partitioning along major convergent plate boundaries. *Pure and Applied Geophysics* 140: 183–210.
Vanek J, Zapotek A, Karnik V, *et al.* (1962) Standardization of magnitude scales. *Izvestiya Akademii Nauk SSSR. Seriya Geologicheskaya* 2: 153–158.
Wald DJ and Heaton TH (1994) Spatial and temporal distribution of slip for the 1992 Landers, California earthquake. *Bulletin of the Seismological Society of America* 84: 668–691.
Wald DJ, Quitoriano V, Heaton T, and Kanamori H (1999) Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California. *Earthquake Spectra* 15: 557–564.
Waldhauser F and Ellsworth WL (2000) A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California. *Bulletin of the Seismological Society of America* 90: 1353–1368.
Waldhauser F, Ellsworth WL, Schaff DP, and Cole A (2004) Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity. *Geophysical Research Letters* 31: L18608 (doi:10.1029/2004GL020649).
Walter WR, Rodgers AJ, Mayeda K, Myers SC, Pasyanos M, and Denny M (1998) Preliminary regional seismic analysis of nuclear explosions and earthquakes in Southwest Asia. In *Proceedings of the 20th Annual Seismic Research Symposium on Monitoring a Comprehensive Test Ban Treaty*, Department of Defense and Department of Energy Report, 442–452.
Wesnousky SG (1994) The Gutenberg-Richter or characteristic earthquake distribution, which is it? *Bulletin of the Seismological Society of America* 84: 1940–1959.
Wesson RL (1988) Dynamics of fault creep. *Journal of Geophysical Research* 93: 8929–8951.
West M, Sanchez JJ, and McNutt SR (2005) Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake. *Science* 308: 1144–1146 (doi:10.1126/science.1112462).
Wibberley CAJ and Shimamoto T (2003) Internal structure and permeability of major strike-slip fault zones: The Median Tectonic Line in Mie Prefecture, Southwest Japan. *Journal of Structural Geology* 25: 59–78.
Widmer R and Zürn W (1992) Bichromatic excitation of long-period Rayleigh and air waves by the Mount Pinatubo and El Chichon volcanic eruptions. *Geophysical Research Letters* 19: 765–768.
Wilson B, Dewers T, Reches Z, and Brune J (2005) Particle size and energetics of gouge from earthquake rupture zones. *Nature* 434: 749–752.
Wong T-F (1982) Shear fracture energy of Westerly granite from postfailure behavior. *Journal of Geophysical Research* 87: 990–1000.
Wood HO and Neumann F (1931) Modified Mercalli intensity scale of 1931. *Bulletin of the Seismological Society of America* 21: 277–283.
Working Group on California Earthquake Probabilities (WGCEP), (2002) Earthquake Probabilities in the San Francisco Bay Region: 2002–2031, *US Geological Survey Open-File Report 03-214*.
Wu J, Wallace T, and Beck S, A very broadband study of the 1994 deep Bolivia earthquake sequence, *Geophys. Res. Lett.*, *22*, 2237–2240, 1995.
Yong C, Tsoi K-L, Feibi C, Henhuan GZ, Qijia Z, and Zhangli C (1988) *The Great Tangshan Earthquake of 1976: An Anatomy of Disaster*, 153 pp. Oxford: Pergamon Press.
Zoback MD, Zoback ML, Mount V, *et al.* (1987) New evidence on the state of stress on the San Andreas fault system. *Science* 238: 1105–1111.
Zoback ML, Zoback MD, Adams J, *et al.* (1989) Global patterns of tectonic stress. *Nature* 341: 291–298 (doi:10.1038/341291a0).

Relevant Websites

http://www.earthscope.org – Earthscope Home Page.
http://www.aeic.alaska.edu – The Alaska Earthquake Information Center.
http://www.seismo.ethz.ch/GSHAP – The GHSAP Page.
http://denali.gsfc.nasa.gov – The Planetary Geodynamics Laboratory at NASA's Goddard Space Flight Center.
http://www.data.scec.org – The Southern California Earthquake Data Center.
http://libraryphoto.cr.usgs.gov – US Geological Survey Photographic Library.
http://quake.usgs.gov – USGS Earthquake Hazards Program – Northern California.
http://www.world-stress-map.org – World Stress Map Project Page.

4.02 Seismic Source Theory

R. Madariaga, Ecole Normale Supérieure, Paris, France

4.02.1 Introduction

Earthquake source dynamics provides key elements for the prediction of ground motion, and to understand the physics of earthquake initiation, propagation, and healing. The simplest possible model of seismic source is that of a point source buried in an elastic half-space. The development of a proper model of the seismic source took more than 50 years since the first efforts by Nakano (1923) and colleagues in Japan. Earthquakes were initially modeled as simple explosions, then as the result of the displacement of conical surfaces and finally as the result of fast transformational strains inside a sphere. In the early 1950s it was recognized that P waves radiated by earthquakes presented a spatial distribution similar to that produced by single couples of forces, but it was very soon recognized that this type of source could not explain S wave radiation (Honda, 1962). The next level of complexity was to introduce a double couple source, a source without resultant force or moment. The physical origin of the double couple model was established in the early 1960s, thanks to the observational work of numerous seismologists and the crucial theoretical breakthrough of Maruyama (1963) and Burridge

and Knopoff (1964), who proved that a fault in an elastic model was equivalent to a double couple source.

In this chapter we review what we believe are the essential results obtained in the field of kinematic earthquake rupture to date. In Section 4.02.2 we review the classical point source model of elastic wave radiation and establish some basic general properties of energy radiation by that source. In Section 4.02.3 we discuss the now classical seismic moment tensor source. In Section 4.02.4 we discuss extended kinematic sources including the simple rectangular fault model proposed by Haskell (1964, 1966) and a circular model that tries to capture some essential features of crack models. Section 4.02.5 introduces crack models without friction as models of shear faulting in the earth. This will help to establish some basic results that are useful in the study of dynamic models of the earthquake source.

4.02.2 Seismic Wave Radiation from a Point Force: Green's Function

There are many ways of solving the elastic wave equation for different types of initial conditions, boundary conditions, sources, etc. Each of these methods requires a specific approach so that a complete solution of the wave equation would be necessary for every different problem that we would need to study. Ideally, we would like however to find a general solution method that would allow us to solve any problem by a simple method. The basic building block of such a general solution method is the Green function, the solution of the following elementary problem: find the radiation from a point source in an infinitely extended heterogeneous elastic medium. It is obvious that such a problem can be solved only if we know how to extend the elastic medium beyond its boundaries without producing unwanted reflections and refractions. Thus, constructing Green's functions is generally as difficult as solving a general wave propagation problem in an inhomogeneous medium. For simplicity, we consider first the particular case of a homogeneous elastic isotropic medium, for which we know how to calculate the Green's function. This will let us establish a general framework for studying more elaborate source models.

4.02.2.1 Seismic Radiation from a Point Source

The simplest possible source of elastic waves is a point force of arbitrary orientation located inside an infinite homogeneous, isotropic elastic body of density ρ, and elastic constants λ and μ. Let $\alpha = \sqrt{(\lambda + 2\mu)/\rho}$ and $\beta = \sqrt{\mu/\rho}$ be the P and S wave speeds, respectively. Let us note $\mathbf{u}(\mathbf{x}, t)$, the particle displacement vector. We have to find the solution of the elastodynamic wave equation

$$\rho \frac{\partial^2}{\partial t^2}\mathbf{u}(\mathbf{x},t) = (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}(\mathbf{x},t)) + \mu\nabla^2\mathbf{u}(\mathbf{x},t) + \mathbf{f}(\mathbf{x},t) \qquad [1]$$

under homogeneous initial conditions, that is, $\mathbf{u}(\mathbf{x}, 0) = \dot{\mathbf{u}}(\mathbf{x}, 0) = 0$, and the appropriate radiation conditions at infinity. In [1] $\mathbf{f}$ is a general distribution of force density as a function of position and time. For a point force of arbitrary orientation located at a point $\mathbf{x}_0$, the body force distribution is

$$\mathbf{f}(\mathbf{x},t) = \mathbf{f}\,s(t)\,\delta(\mathbf{x}-\mathbf{x_0}) \qquad [2]$$

where $s(t)$ is the source time function, the variation of the amplitude of the force as a function of time. $\mathbf{f}$ is a unit vector in the direction of the point force.

The solution of eqn [1] is easier to obtain in the Fourier transformed domain. As is usual in seismology, we use the following definition of the Fourier transform and its inverse:

$$\tilde{u}(x,\omega) = \int_{-\infty}^{\infty} u(x,t)\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t$$
$$u(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{u}(x,\omega)\,\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega \qquad [3]$$

Here, and in the following, we will denote Fourier transform with a tilde.

After some lengthy work (see, e.g., Achenbach, 1975), we find the Green function in the Fourier domain:

$$\tilde{\mathbf{u}}(R,\omega) = \frac{1}{4\pi\rho}\left[\mathbf{f}\cdot\nabla\nabla\left(\frac{1}{R}\right)\right]\frac{\tilde{s}(\omega)}{\omega^2} \times\left[-\left(1+\frac{\mathrm{i}\omega R}{\alpha}\right)\mathrm{e}^{-\mathrm{i}\omega R/\alpha}+\left(1+\frac{\mathrm{i}\omega R}{\beta}\right)\mathrm{e}^{-\mathrm{i}\omega R/\beta}\right] + \frac{1}{4\pi\rho\alpha^2}\frac{1}{R}(\mathbf{f}\cdot\nabla R)\nabla R\tilde{s}(\omega)\,\mathrm{e}^{-\mathrm{i}\omega R/\alpha} + \frac{1}{4\pi\rho\beta^2}\frac{1}{R}[\mathbf{f}-(\mathbf{f}\cdot\nabla R)\nabla R]\tilde{s}(\omega)\,\mathrm{e}^{-\mathrm{i}\omega R/\beta} \qquad [4]$$

where $R = \|\mathbf{x} - \mathbf{x}_0\|$ is the distance from the source to the observation point. Using the following Fourier transform

$$-\frac{1}{\omega^2}\left[1 + \frac{\mathrm{i}\omega R}{\alpha}\right]\mathrm{e}^{-\mathrm{i}\omega R/\alpha} \leftrightarrow t\mathrm{H}(t - R/\alpha)$$

we can transform eqn [4] to the time domain in order to obtain the final result:

$$\mathbf{u}(R,t) = \frac{1}{4\pi\rho}\left[\mathbf{f}\cdot\nabla\nabla\left(\frac{1}{R}\right)\right]\int_{R/\alpha}^{\min(t,R/\beta)} \tau s(t-\tau)\,\mathrm{d}\tau$$

$$+\frac{1}{4\pi\rho\alpha^2}\frac{1}{R}(\mathbf{f}\cdot\nabla R)\nabla R s(t - R/\alpha)$$

$$+\frac{1}{4\pi\rho\beta^2}\frac{1}{R}[\mathbf{f} - (\mathbf{f}\cdot\nabla R)\nabla R]\, s(t - R/\beta) \qquad [5]$$

This complicated looking expression can be better understood considering each of its terms separately. The first line is the near field which comprises all the terms that decrease with distance faster than R^{-1}. The last two lines are the far field that decreases with distance like R^{-1} as for classical spherical waves.

4.02.2.2 Far-Field Body Waves Radiated by a Point Force

Much of the practical work of seismology is done in the far field, at distances of several wavelengths from the source. In that region it is not necessary to use the complete elastic field as detailed by eqn [5]. When the distance R is large only the last two terms are important. There has been always been some confusion in the seismological literature with respect to the exact meaning of the term "far field". For a point force, which by definition has no length scale, what is exactly the distance beyond which we are in the far field? This problem has important practical consequences for the numerical solution of the wave equation, for the computation of "near-source" accelerograms, etc. In order to clarify this, we examine the frequency domain expression for the Green function [4]. Under what conditions can we neglect the first term of that expression with respect to the last two? For that purpose we notice that R appears always in the non-dimensional combination $\omega R/\alpha$ or $\omega R/\beta$. Clearly, these two ratios determine the far-field conditions. Since $\alpha > \beta$, we conclude that the far field is defined by

$$\frac{\omega R}{\alpha} >> 1 \quad \text{or} \quad \frac{R}{\lambda} >> 1$$

where $\lambda = 2\pi\alpha/\omega$ is the wavelength of a P wave of circular frequency ω. The condition for the far field depends therefore on the characteristic frequency or wavelength of the radiation. Thus, depending on the frequency content of the signal $\tilde{s}(\omega)$, we will be in the far field for high-frequency waves, but we may be in the near field for the low-frequency components. In other words, for every frequency component there is a distance of several wavelengths for which we are in the far field. In particular for zero-frequency waves, the static approximation, all points in the earth are in the near field of the source, while at high frequencies higher than 1 Hz say we are in the far field 10 km away from the source.

The far-field radiation from a point force is usually written in the following, shorter form:

$$\mathbf{u}^{\mathrm{P}}_{\mathrm{FF}}(R,t) = \frac{1}{4\pi\rho\alpha^2}\frac{1}{R}\mathcal{R}^P s(t - R/\alpha)$$
$$\mathbf{u}^{\mathrm{S}}_{\mathrm{FF}}(R,t) = \frac{1}{4\pi\rho\beta^2}\frac{1}{R}\mathcal{R}^S s(t - R/\beta) \qquad [6]$$

where $\mathcal{R}^{\mathrm{P}}$ and $\mathcal{R}^{\mathrm{S}}$ are the radiation patterns of P and S waves, respectively. Noting that $\nabla R = \mathbf{e}_{\mathrm{R}}$, the unit vector in the radial direction, we can write the radiation patterns in the following simplified form, $\mathcal{R}^{\mathcal{P}} = f_{\mathrm{R}}\mathbf{e}_{\mathrm{R}}$ and $\mathcal{R}^{\mathcal{S}} = \mathbf{f}_{\mathrm{T}} = \mathbf{f} - f_{\mathrm{R}}\mathbf{e}_{\mathrm{R}}$ where f_{R} is the radial component of the point force $\mathbf{f}$, and $\mathbf{f}_{\mathrm{T}}$, its transverse component.

Thus, in the far field of a point force, P waves propagate the radial component of the point force, whereas the S waves propagate information about the transverse component of the point force. Expressing the amplitude of the radial and transverse component of $\mathbf{f}$ in terms of the azimuth θ of the ray with respect to the applied force, we can rewrite the radiation patterns in the simpler form

$$\mathcal{R}^{\mathrm{P}} = \cos\theta\,\mathbf{e}_{\mathrm{R}}, \qquad \mathcal{R}^{\mathrm{S}} = \sin\theta\,\mathbf{e}_{\mathrm{T}} \qquad [7]$$

As we could expect from the natural symmetry of the problem, the radiation patterns are axially symmetric about the axis of the point force. P waves from a point force have a typical dipolar radiation pattern, while S waves have a toroidal (doughnut-shaped) distribution of amplitudes.

4.02.2.3 The Near Field of a Point Force

When $\omega R/\alpha$ is not large compared to one, all the terms in eqns [5] and [4] are of equal importance. In fact, both far-and near-field terms are of the same order of magnitude near the point source. In order to

calculate the small R behavior it is preferable to go back to the frequency domain expression [4]. When $R \to 0$ the term in brackets in the first line tends to zero. In order to calculate the near-field behavior we have to expand the exponentials to order R^2, that is,

$$\exp(-i\omega R/\alpha) \simeq 1 - i\omega R/\alpha - \omega^2 R^2/\alpha^2 + \mathcal{O}(\omega^3 R^3)$$

and a similar expression for the exponential that depends on the S wave speed. After some algebra we find

$$\tilde{\mathbf{u}}(R,\omega) = \frac{1}{8\pi\rho}\frac{1}{R}\left[(\mathbf{f}\cdot\nabla R)\nabla R\left(\frac{1}{\beta^2}-\frac{1}{\alpha^2}\right) + \mathbf{f}\left(\frac{1}{\beta^2}+\frac{1}{\alpha^2}\right)\right]\tilde{s}(\omega) \quad [8]$$

or in the time domain

$$\mathbf{u}(R,t) = \frac{1}{8\pi\rho}\frac{1}{R}\left[(\mathbf{f}\cdot\nabla R)\nabla R\left(\frac{1}{\beta^2}-\frac{1}{\alpha^2}\right) + \mathbf{f}\left(\frac{1}{\beta^2}+\frac{1}{\alpha^2}\right)\right]s(t) \quad [9]$$

This is the product of the source time function $s(t)$ and the static displacement produced by a point force of orientation $\mathbf{f}$:

$$\mathbf{u}(R) = \frac{1}{8\pi\rho}\frac{1}{R}\left[(\mathbf{f}\cdot\nabla R)\nabla R\left(\frac{1}{\beta^2}-\frac{1}{\alpha^2}\right) + \mathbf{f}\left(\frac{1}{\beta^2}+\frac{1}{\alpha^2}\right)\right] \quad [10]$$

This is one of the most important results of static elasticity and is frequently referred to as the Kelvin solution.

The result [9] is quite interesting and somewhat unexpected. The radiation from a point source decays like R^{-1} in the near field, exactly like the far field terms. This result has been remarked and extensively used in the formulation of regularized boundary integral equations for elastodynamics (Hirose and Achenbach, 1989; Fukuyama and Madariaga, 1995).

4.02.2.4 Energy Flow from Point Force Sources

A very important issue in seismology is the amount of energy radiated by seismic sources. Traditionally, seismologists call seismic energy the total amount of energy that flows across a surface that encloses the force far way from it. The flow of energy across any surface that encloses the point source must be the same, so that seismic energy is defined for any arbitrary surface.

Let us take the scalar product of eqn [1] by the particle velocity $\dot{\mathbf{u}}$ and integrate on a volume V that encloses all the sources, in our case the single point source located at $\mathbf{x}_0$:

$$\int_V \rho \dot{u}_i \ddot{u}_i \, dV = \int_V \sigma_{ij,j}\dot{u}_i \, dV + \int_V f_i \dot{u}_i \, dV \quad [11]$$

where we use dots to indicate time derivatives and the summation convention on repeated indices. In order to facilitate the calculations, in [11] we have rewritten the left-hand side of [1] in terms of the stresses $\sigma_{ij} = \lambda\epsilon_{ij}\delta_{ij} + 2\mu\epsilon_{ij}$, where $\epsilon_{ij} = 1/2\,(\partial_j u_i + \partial_j u_i)$.

Using $\sigma_{ij,j}\dot{u}_i = (\sigma_{ij}\dot{u}_i)_{,j} - \sigma_{ij}\dot{\epsilon}_{ij}$ and Gauss' theorem we get the energy flow identity:

$$\frac{d}{dt}(K(t) + U(t)) = \int_S \sigma_{ij}\dot{u}_i n_j \, dS + \int_V f_i \dot{u}_i \, dV \quad [12]$$

where $\mathbf{n}$ is the outward normal to the surface S (see **Figure 1**). In [12] K is the kinetic energy contained in volume V at time t:

$$K(t) = \frac{1}{2}\int_V \dot{\mathbf{u}}^2 \, dV \quad [13]$$

while

$$U(t) = \frac{1}{2}\int_V \left[\lambda(\nabla\cdot\mathbf{u})^2 + 2\mu\epsilon_{ij}\epsilon_{ij}\right] dV \quad [14]$$

is the strain energy change inside the same volume. The last term is the rate of work of the force against elastic displacement. Equation [12] is the basic energy conservation statement for elastic sources. It says that the rate of energy change inside the body V

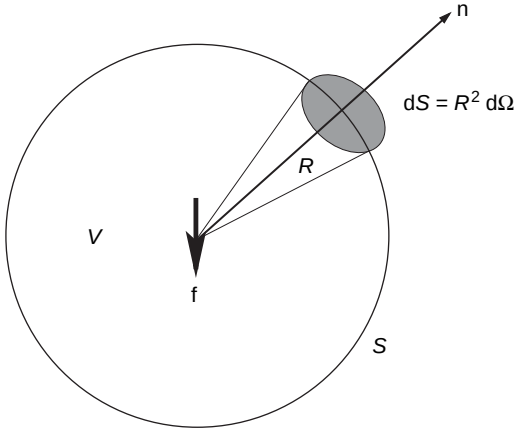


Figure 1 Geometry for computing radiated energy from a point source. The source is at the origin and the observer at a position defined by the spherical coordinates R, θ, ϕ, distance, polar angle, and azimuth.

is equal to the rate of work of the sources **f** plus the energy flow across the boundary S.

Let us note that in [12] energy flows into the body. In seismology, however, we are interested in the seismic energy that flows out of the elastic body; thus, the total seismic energy flow until a certain time t is

$$\begin{aligned} E_s(t) &= -\int_0^t \mathrm{d}t \int_S \sigma_{ij}\dot{u}_i n_j \,\mathrm{d}S \\ &= -K(t) - \Delta U(t) + \int_0^t \mathrm{d}t \int_V f_i \dot{u}_i \,\mathrm{d}V \end{aligned} \quad [15]$$

where $\Delta U(t)$ is the strain energy change inside the elastic body since time $t=0$. If t is sufficiently long, so that all motion inside the body has ceased, $K(t) \to 0$ and we get the simplest possible expression

$$E_s = -\Delta U + \int_0^\infty \mathrm{d}t \int_V f_i \dot{u}_i \,\mathrm{d}V \quad [16]$$

Thus, total energy radiation is equal to the decrease in internal energy plus the work of the sources against the elastic deformation.

Although we can use [16] to compute the seismic energy, it is easier to evaluate the energy directly from the first line of [15] using the far field [6]. Consider as shown in **Figure 1** a cone of rays of cross section $\mathrm{d}\Omega$ issued from the source around the direction θ, ϕ. The energy crossing a section of this ray beam at distance R from the source per unit time is given by the energy flow per unit solid angle:

$$\dot{e}_s \,\mathrm{d}\Omega = \sigma_{ij}\dot{u}_i n_j R^2 \,\mathrm{d}\Omega$$

where σ_{ij} is stress, $\dot{u}_i$ the particle velocity, and **n** the normal to the surface $\mathrm{d}S = R^2\,\mathrm{d}\Omega$. We now use [6] in order to compute σ_{ij} and $\dot{u}$. By straightforward differentiation and keeping only terms of order $1/R$ with distance, we get $\sigma_{ij}n_j \simeq \rho c \dot{u}_i$, where ρc is the wave impedance and c the appropriate wave speed. The energy flow rate per unit solid angle for each type of wave is then

$$\dot{e}_s(t) = \begin{cases} \rho\alpha \;\; R^2\dot{u}^2(R,t) & \text{for P waves} \\ \rho\beta \;\; R^2\dot{u}^2(R,t) & \text{for S waves} \end{cases} \quad [17]$$

Inserting [6] and integrating around the source for the complete duration of the source, we get the total energy flow associated with P and S waves:

$$\begin{aligned} E_p &= \frac{1}{4\pi\rho\alpha^3} < \mathcal{R}^P >_2 \int_0^\infty \dot{s}^2(t)\,\mathrm{d}t \quad \text{for P waves} \\ E_s &= \frac{1}{4\pi\rho\beta^3} < \mathcal{R}^S >_2 \int_0^\infty \dot{s}^2(t)\,\mathrm{d}t \quad \text{for S waves} \end{aligned} \quad [18]$$

where $< \mathcal{R}^c >_2 = (1/4\pi)\int_\Omega (\mathcal{R}^c)^2 \,\mathrm{d}\Omega$ is the mean squared radiation pattern for wave $c = \{P, S\}$. Since the radiation patterns are the simple sinusoidal functions listed in [7], the mean square radiation patterns are $1/3$ for P waves and $2/3$ for the sum of the two components of S waves. In [18] we assumed that $\dot{s}(t) = 0$ for $t < 0$. Finally, it not difficult to verify that, since $\dot{s}$ has units of force rate, E_s and E_p have units of energy. Noting that in the earth, α is roughly $\sqrt{3}\beta$ so that $\alpha^3 \simeq 5\beta^3$, the amount of energy carried by S waves is close to 10 times that carried by P waves.

4.02.2.5 The Green Tensor for a Point Force

The Green function is a tensor formed by the waves radiated from a set of three point forces aligned in the direction of each coordinate axis. For an arbitrary force of direction **f**, located at point $\mathbf{x}_0$ and source function $s(t)$, we define the Green tensor for elastic waves by

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{G}(\mathbf{x}, t|\mathbf{x}_0, 0) \cdot \mathbf{f} * s(t)$$

where the star indicates time-domain convolution.

We can also write this expression in the usual index notation

$$u_i(\mathbf{x}, t) = \sum_j G_{ij}(\mathbf{x}, t\big| \cdot \mathbf{x}_0, 0) f_j * s(t)$$

in the time domain or

$$\tilde{u_i}(\mathbf{x}, \omega) = \sum_j \tilde{G_{ij}}(\mathbf{x}\big|\mathbf{x}_0, \omega) f_j \tilde{s}(\omega)$$

in the frequency domain.

The Green function itself can easily be obtained from the radiation from a point force [5]

$$\begin{aligned} G_{ij}(\mathbf{x}, t|\mathbf{x}_0, 0) = &\frac{1}{4\pi\rho}\left(\frac{1}{R}\right)_{,ij} \\ &\times t[H(t - R/\alpha) - H(t - R/\beta)] \\ &+ \frac{1}{4\pi\rho\alpha^2}\frac{1}{R}(R_{,i}R_{,j})\delta(t - R/\alpha) \\ &+ \frac{1}{4\pi\rho\beta^2}\frac{1}{R}[\delta_{ij} - R_{,i}R_j]\delta(t - R/\beta) \end{aligned} \quad [19]$$

Here $\delta(t)$ is Dirac's delta, δ_{ij} is Kronecker's delta, and the comma indicates derivative with respect the component that follows it.

Similarly, in the frequency domain

$$\tilde{G_{ij}}(\mathbf{x}|\boldsymbol{x}_0,\omega) = \frac{1}{4\pi\rho}\left(\frac{1}{R}\right)_{,ij}\frac{1}{\omega^2} \times\left[-\left(1+\frac{\mathrm{i}\omega R}{\alpha}\right)\mathrm{e}^{-\mathrm{i}\omega R/\alpha}+\left(1+\frac{\mathrm{i}\omega R}{\beta}\right)\mathrm{e}^{-\mathrm{i}\omega R/\beta}\right] +\frac{1}{4\pi\rho\alpha^2}\frac{1}{R}(R_{,i}R_{,j})\mathrm{e}^{-\mathrm{i}\omega R/\alpha} +\frac{1}{4\pi\rho\beta^2}\frac{1}{R}[\delta_{ij}-R_{,i}R_{,j}]\mathrm{e}^{-\mathrm{i}\omega R/\beta} \quad [20]$$

For the calculation of radiation from a moment tensor seismic source, or for the calculation of strain and stress radiated by the point source, we need the space derivatives of [20]. In the following, we list separately the near-field (NF) terms, the intermediate-field (IF), and the far-field (FF) terms. The separation into intermediate and near field is somewhat arbitrary but it facilitates the computations of Fourier transforms.

Let us write first the gradient of displacement:

$$G_{ij,k} = \frac{\partial G_{ij}}{\partial x_k} = G^{\mathrm{NF}}_{ij,k} + G^{\mathrm{IF}}_{ij,k} + G^{\mathrm{FF}}_{ij,k}$$

After a relatively long, but straightforward work we get

$$\tilde{G}^{\mathrm{FF}}_{ij,k}(\mathbf{x}|\mathbf{x}_0,\omega) = \frac{1}{4\pi\rho\alpha^3}\frac{1}{R}\mathcal{R}^{\mathrm{P}}_{ijk}(-\mathrm{i}\omega)\mathrm{e}^{-\mathrm{i}\omega R/\alpha} +\frac{1}{4\pi\rho\beta^3}\frac{1}{R}\mathcal{R}^{\mathrm{S}}_{ijk}(-\mathrm{i}\omega)\mathrm{e}^{-\mathrm{i}\omega R/\beta}$$

$$\tilde{G}^{\mathrm{IF}}_{ij,k}(\mathbf{x}|\mathbf{x}_0,\omega) = \frac{1}{4\pi\rho\alpha^2}\frac{1}{R^2}\mathcal{I}^{\mathrm{P}}_{ijk}\mathrm{e}^{-\mathrm{i}\omega R/\alpha} +\frac{1}{4\pi\rho\beta^2}\frac{1}{R^2}\mathcal{I}^{\mathrm{S}}_{ijk}\mathrm{e}^{-\mathrm{i}\omega R/\beta}$$

$$\tilde{G}^{\mathrm{NF}}_{ij,k}(\mathbf{x}|\mathbf{x}_0,\omega) = \frac{1}{4\pi\rho}\frac{1}{R^4}\mathcal{N}_{ijk}\frac{1}{\omega^2} \times\left[-\left(1+\frac{\mathrm{i}\omega R}{\alpha}\right)\mathrm{e}^{-\mathrm{i}\omega R/\alpha} +\left(1+\frac{\mathrm{i}\omega R}{\beta}\right)\mathrm{e}^{-\mathrm{i}\omega R/\beta}\right] \quad [21]$$

where $R = |\mathbf{x}-\mathbf{x_0}|$ and the coefficients $\mathcal{R}$, $\mathcal{I}$, and $\mathcal{N}$ are listed on **Table 1**.

We observe that the frequency dependence and distance decay is quite different for the various terms.

Table 1 Radiation patterns for a point force in a homogeneous elastic medium

Coefficient	P waves	S waves
$\mathcal{R}_{ijk}$	$R_{,i}R_{,j}R_{,k}$	$\delta_{ij}R_{,k}-R_{,i}R_jR_{,k}$
$\mathcal{I}_{ijk}$	$-R^2R_{,ijk}-3R_{,i}R_jR_{,k}$	$R^2R_{,ijk}+3R_{,i}R_jR_{,k}-\delta_{ij}R_{,k}$
$\mathcal{N}_{ijk}$	$R^4\,(R^{-1})_{,ijk}$	

The most commonly used terms, the far field, decay like $1/R$ and have a time dependence dominated by the time derivative of the source time function $\dot{s}(t)$.

4.02.3 Moment Tensor Sources

The Green function for a point force is the fundamental solution of the equation of elastodynamics and it will find extensive use in this book. However, except for a few rare exceptions seismic sources are due to fast internal deformation in the earth, for instance, faulting or fast phase changes on localized volumes inside the earth. For a seismic source to be of internal origin, it has to have zero net force and zero net moment. It is not difficult to imagine seismic sources that satisfy these two conditions:

$$\sum \mathbf{f} = 0 \qquad \sum \mathbf{f}\times\mathbf{r} = 0 \quad [22]$$

The simplest such sources are dipoles and quadrupoles. For instance, the so-called linear dipole is made of two point sources that act in opposite directions at two points separated by a very small distance h along the axes of the forces. The strength, or seismic moment, of this linear dipole is $\mathbf{M} = \mathbf{f}h$. Experimental observation has shown that linear dipoles of this sort are not good models of seismic sources and, furthermore, there does not seem to be any simple internal deformation mechanism that corresponds to a pure linear dipole. It is possible to combine three orthogonal linear dipoles in order to form a general seismic source; any dipolar seismic source can be simulated by adjusting the strength of these three dipoles. It is obvious, as we will show later, that these three dipoles represent the principal directions of a symmetric tensor of rank 2 that we call the seismic moment tensor:

$$\mathbf{M} = \begin{pmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{xy} & M_{yy} & M_{yz} \\ M_{xz} & M_{yz} & M_{zz} \end{pmatrix}$$

This moment tensor has a structure that is identical to that of a stress tensor, but it is not of elastic origin as we shall see promptly.

What do the off-diagonal elements of the moment tensor represent? Let us consider a moment tensor such that all elements are zero except M_{xy}. This moment tensor represents a double couple, a pair of two couples of forces that turn in opposite directions.

The first of these couples consists in two forces of direction $\mathbf{e}_x$ separated by a very small distance h in the direction y. The other couple consists in two forces of direction $\mathbf{e}_y$ with a small arm in the direction x. The moment of each of the couple is M_{xy}, the first pair has positive moment, the second has a negative one. The conditions of conservation of total force and moment [22] are satisfied so that this source model is fully acceptable from a mechanical point of view. In fact, as shown by Burridge and Knopoff (1964), the double couple is the natural representation of a fault. One of the pair of forces is aligned with the fault; the forces indicate the directions of slip and the arm is in the direction of the fault thickness.

4.02.3.1 Radiation from a Point Moment Tensor Source

Let us now use the Green functions obtained for a point force in order to calculate the radiation from a point moment tensor source located at point $\mathbf{x}_0$:

$$\mathbf{M}_0(\mathbf{r}, t) = \mathbf{M}_0(t)\,\delta(\mathbf{x} - \mathbf{x}_0) \qquad [23]$$

$\mathbf{M}_0$ is the moment tensor, a symmetric tensor whose components are independent functions of time.

We consider one of the components of the moment tensor, for instance, M_{ij}. This represents two point forces of direction i separated by an infinitesimal distance h_j in the direction j. The radiation of each of the point forces is given by the Green function G_{ij} computed in [19]. The radiation from the M_{ij} moment is then just

$$u_k(\mathbf{x}, t) = (G_{ki}(\mathbf{x}, t|\mathbf{x}_0 + h_j\mathbf{e}_j, t) * f_i(t) - G_{ki}(\mathbf{x}, t|\mathbf{x}_0, t) * f_i(t))$$

When $h \to 0$ we get

$$u_k(\mathbf{x}, t) = (\partial_j G_{ki}(\mathbf{x}, t|\mathbf{x}_0, 0) * M_{ij}(t))$$

where $M_{ij} = f_i h_j$. For a general moment tensor source, the radiation is then simply

$$u_k(\mathbf{x}, t) = \sum_{ij} G_{ki,j}(\mathbf{x}, t|\mathbf{x}_0, t) * M_{ij}(t) \qquad [24]$$

The complete expression of the radiation from a point moment tensor source can then be obtained from [24] and the entries in Table 1. We will be interested only on the FF terms since the near field is too complex to discuss here.

We get, for the FF waves,

$$u_i^{\mathrm{P}}(R, t) = \frac{1}{4\pi\rho\alpha^3}\frac{1}{R}\sum_{jk}\mathcal{R}_{ijk}^{\mathrm{P}}\dot{M}_{jk}(t - R/\alpha)$$
$$u_i^{\mathrm{S}}(R, t) = \frac{1}{4\pi\rho\beta^3}\frac{1}{R}\sum_{jk}\mathcal{R}_{ijk}^{\mathrm{S}}\dot{M}_{jk}(t - R/\beta) \qquad [25]$$

where $\mathcal{R}_{ijk}^{\mathrm{P}}$ and $\mathcal{R}_{ijk}^{\mathrm{S}}$, listed in Table 1, are the radiation patterns of P and S waves, respectively. We observe that the radiation pattern is different for every element of the moment tensor. Each orientation of the moment has its own characteristic set of symmetries and nodal planes. As shown by [25] the FF signal carried by both P and S waves is the time derivative of the seismic moment components, so that far field seismic waves are proportional to the moment rate of the source. This may be explained as follows. If slip on a fault occurs very slowly, no seismic waves will be generated by this process. For seismic waves to be generated, fault slip has to be rather fast so that waves are generated by the time variation of the moment tensor, not by the moment itself.

Very often in seismology it is assumed that the geometry of the source can be separated from its time variation, so that the moment tensor can be written in the simpler form:

$$\mathbf{M}_0(t) = \mathbf{M}s(t)$$

where $\mathbf{M}$ is a time-invariant tensor that describes the geometry of the source and $s(t)$ is the time variation of the moment, the source time function determined by seismologists. Using Figure 2 we can now write a simpler form of [25]:

$$u_c(\mathbf{x}, t) = \frac{1}{4\pi\rho c^3}\frac{\mathcal{R}_c(\theta, \phi)}{R}\Omega(t - R/c) \qquad [26]$$

where R is the distance from the source to the observer. c stands for either α for P waves or β for shear waves (SH and SV). For P waves u_c is the radial component; for S waves it is the appropriate transverse component for SH or SV waves. In [26] we have introduced the standard notation $\Omega(t) = \dot{s}(t)$ for the source time function, the signal emitted by the source in the far field.

The term $\mathcal{R}_c(\theta, \phi)$ is the radiation pattern, a function of the takeoff angle of the ray at the source. Let (R, θ, ϕ) be the radius, co-latitude, and azimuth of a system of spherical coordinates centered at the source. It is not difficult to show that the radiation pattern is given by

$$\mathcal{R}_P(\theta, \phi) = \mathbf{e}_R \cdot \mathbf{M} \cdot \mathbf{e}_R \qquad [27]$$

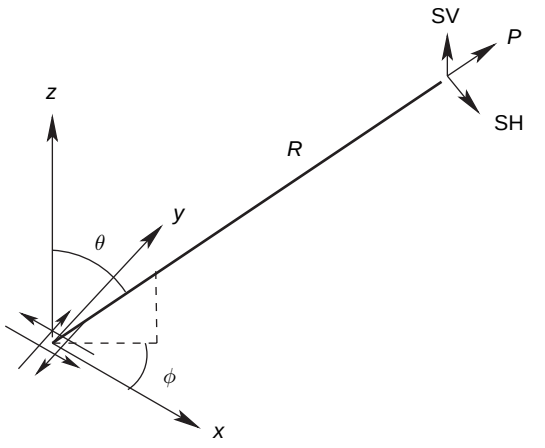


Figure 2 Radiation from a point double source. The source is at the origin and the observer at a position defined by the spherical coordinates R, θ, ϕ, distance, polar angle, and azimuth.

for P waves, where $\mathbf{e}_R$ is the radial unit vector at the source. Assuming that the z-axis at the source is vertical, so that θ is measured from that axis, S waves are given by

$$\begin{aligned}\mathcal{R}_{SV}(\theta,\phi) &= \mathbf{e}_\theta \cdot \mathbf{M} \cdot \mathbf{e}_R \\ \mathcal{R}_{SH}(\theta,\phi) &= \mathbf{e}_\phi \cdot \mathbf{M} \cdot \mathbf{e}_R\end{aligned} \qquad [28]$$

where $\mathbf{e}_\phi$ and $\mathbf{e}_\theta$ are unit vectors in spherical coordinates. Thus, the radiation patterns are the radial componets of the moment tensor projected on spherical coordinates.

With minor changes to take into account smooth variations of elastic wave speeds in the earth, these expressions are widely used to generate synthetic seismograms in the so-called FF approximation. The main changes that are needed are the use of travel time $T_c(r, r_o)$ instead of R/c in the waveform $\Omega(t - T_c)$, and a more accurate geometrical spreading $g(\Delta, H)/a$ to replace $1/R$, where a is the radius of the earth and $g(\Delta, H)$ is a tabulated function that depends on the angular distance Δ between hypocenter and observer and the source depth H. In most work with local earthquakes, the approximation [26] is frequently used with a simple correction for free surface response.

4.02.3.2 A More General View of Moment Tensors

What does a seismic moment represent? A number of mechanical interpretations are possible. In the previous sections we introduced it as a simple mechanical model of double couples and linear dipoles. Other authors (Backus and Mulcahy, 1976) have explained them in terms of the distribution of inelastic stresses (some times called stress 'glut').

Let us first notice that a very general distribution of force that satisfies the two conditions [22] necessarily derives from a symmetrical seismic moment density of the form

$$\mathbf{f}(\mathbf{x}, t) = \nabla \cdot \mathbf{M}(\mathbf{x}, t) \qquad [29]$$

where $\mathbf{M}(\mathbf{x}, t)$ is the moment tensor density per unit volume. Gauss' theorem can be used to prove that such a force distribution, derived from a moment tensor field, has no net force nor moment. In many areas of applied mathematics, the seismic moment distribution is often termed a 'double layer potential'.

We can now use [29] in order to rewrite the elastodynamic eqn [1] as a system of first-order partial differential equations:

$$\begin{aligned}\rho \frac{\partial}{\partial t}\mathbf{v} &= \nabla \cdot \sigma \\ \frac{\partial}{\partial t}\sigma &= \lambda \nabla \cdot \mathrm{v}\mathbf{I} + \mu\left[(\nabla \mathbf{v}) + (\nabla \mathbf{v})^{\mathrm{T}}\right] + \dot{\mathbf{M}}_0\end{aligned} \qquad [30]$$

where $\mathbf{v}$ is the particle velocity and σ is the corresponding elastic stress tensor. We observe that the moment tensor density source appears as an addition to the elastic stress rate $\dot{\sigma}$. This is probably the reason that Backus and Mukahy adopted the term 'glut'. In many other areas of mechanics, the moment tensor is considered to represent the stresses produced by inelastic processes. A full theory of these stresses was proposed by Eshelby (1956). Incidentally, the equation of motion written in this form is the basis of some very successful numerical methods for the computation of seismic wave propagation (see, e.g., Madariaga, 1976; Virieux, 1986; Madariaga *et al.*, 1998).

We can get an even clearer view of the origin of the moment tensor density by considering it as defining an "inelastic" strain tensor ϵ^{I} defined implicitly by

$$(m_0)_{ij} = \lambda \delta_{ij} \epsilon^{\mathrm{I}}_{kk} + 2\mu \epsilon^{\mathrm{I}}_{ij} \qquad [31]$$

Many seismologists have tried to use ϵ_{I} in order to represent seismic sources. Sometimes termed "potency" (Ben Menahem and Singh, 1981), the inelastic strain has not been widely adopted even if it is a more natural way of introducing seismic source in bimaterial interfaces and other heterogeneous media. For a recent discussion, see Ampuero and Dahlen (2005).

The meaning of ϵ^{I} can be clarified by reference to **Figure 3**. Let us make the following 'gedanken'

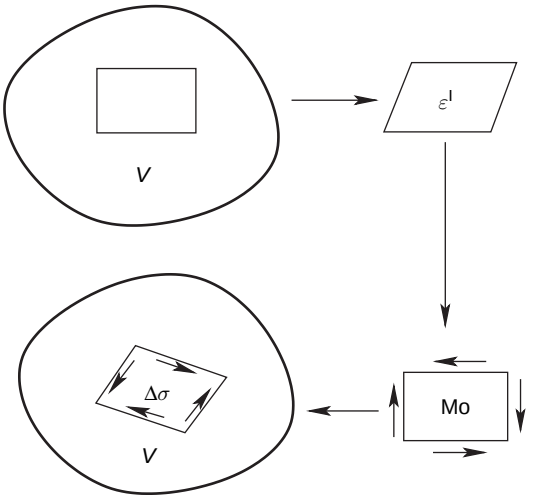


Figure 3 Inelastic stresses or stress glut at the origin of the concept of seismic moment tensor. We consider a small rectangular zone that undergoes an spontaneous internal deformation ϵ^{I} (top row). The elastic stresses needed to bring it back to a rectangular shape are the moment-tensor or stress glut (bottom row right). Once stresses are relaxed by interaction with the surrounding elastic medium, the stress change is $\Delta\sigma$ (bottom left).

(mental) experiment. Let us cut an infinitesimal volume V from the source region. Next, we let it undergo some inelastic strain ϵ^{I}, for instance, a shear strain due to the development of internal dislocations as shown in the figure. Let us now apply stresses on the borders of the internally deformed volume V so as to bring it back to its original shape. If the elastic constants of the internally deformed volume V have not changed, the stresses needed to bring V back to its original shape are exactly given by the moment tensor components defined in [31]. This is the definition of seismic moment tensor: it is the stress produced by the inelastic deformation of a body that is elastic everywhere. It should be clear that the moment tensor is not the same thing as the stress tensor acting in the fault zone. The latter includes the elastic response to the introduction of internal stresses as shown in the last row of **Figure 3**. The difference between the initial stresses before the internal deformation, and those that prevail after the deformed body has been reinserted in the elastic medium is the stress change (or stress drop as originally introduced in seismology in the late 1960s). If the internal strain is produced in the sense of reducing applied stress – and reducing internal strain energy – then stresses inside the source will decrease in a certain average sense. It must be understood, however, that a source of internal origin like faulting can only redistribute internal stresses. During faulting stresses reduce in the immediate vicinity of slip zones, but increase almost everywhere else.

4.02.3.3 Moment Tensor Equivalent of a Fault

For a point moment tensor of type [23], we can write

$$(M_0)_{ij} = (\lambda\delta_{ij}\epsilon^{I}_{kk} + 2\mu\epsilon^{I}_{ij})V\delta(\mathbf{x}-\mathbf{x}_0) \qquad [32]$$

where V is the elementary source volume on which acts the source. Let us now consider that the source is a very thin cylinder of surface S, thickness h, and unit normal $\mathbf{n}$, then,

$$V = Sh \qquad [33]$$

Now, letting the thickness of the cylinder tend to zero, the mean inelastic strain inside the volume V can be computed as follows:

$$\lim_{h\to 0}\epsilon^{I}_{ij}h = \frac{1}{2}[\Delta u_i n_j + \Delta u_j n_i] \qquad [34]$$

where $\Delta\mathbf{u}$ is the displacement discontinuity (or simply the "slip") across the fault volume. The seismic moment for the flat fault is then

$$(M_0)_{ij} = [\lambda\delta_{ij}\Delta u_k n_k + \mu(\Delta u_i n_j + \Delta u_j n_i)]S \qquad [35]$$

so that the seismic moment can be defined for a fault as the product of an elastic constant by the displacement discontinuity and the source area. Actually, this is the way the seismic moment was originally determined by Burridge and Knopoff (1964). If the slip discontinuity is written in terms of a direction of slip ν and a scalar slip D, $\Delta u_i = D\nu_i$, we get

$$(M_0)_{ij} = \delta_{ij}\nu_k n_k \lambda DS + (\nu_i n_j + \nu_j n_i)\mu DS \qquad [36]$$

Most seismic sources do not produce normal displacement discontinuities (fault opening) so that $\nu \cdot \mathbf{n} = 0$ and the first term in [36] is equal to zero. In that case the seismic moment tensor can be written as the product of a tensor with the scalar seismic moment $M_0 = \mu DS$:

$$(M_0)_{ij} = (\nu_i n_j + \nu_j n_i)\mu DS \qquad [37]$$

This is the form originally derived from dislocation theory by Burridge and Knopoff (1964). The first practical determination of the scalar seismic moment

$$M_0 = \mu DS$$

is due to Aki (1966), who estimated M_0 from seismic data recorded after the Niigata earthquake of 1966 in Japan. Determination of seismic moment has become the standard way in which earthquakes are measured. All sort of seismological, geodetic, and geological techniques have been used to determine M_0. A worldwide catalog of seismic moment tensors is made available online by Harvard University (Dziewonski and Woodhouse, 1983). Initially, moments were determined by Harvard for the limited form [37], but since the 1990s Harvard computes the full six components of the moment tensor without reference to a particular source model.

Let us remark that the restricted form of the moment tensor [37] reduces the number of independent parameters of the moment tensor. For a general source representation there are six parameters, whereas for the restricted case there are only four: the moment, two components of the slip vector ν, and one component of the normal vector $\mathbf{n}$, which is perpendicular to ν. Very often seismologists use the simple fault model of the source moment tensor. The fault is parametrized by the seismic moment plus the three Euler angles for the fault plane. Following the convention adopted by Aki and Richards, these angles are defined as δ the dip of the fault, ϕ the strike of the fault with respect to the North, and λ the rake of the fault, that is the angle of the slip vector with respect to the horizontal.

4.02.3.4 Eigenvalues and Eigenvectors of the Moment Tensor

Since the moment tensor is a symmetric tensor of order 3, it has three orthogonal eigenvectors with real eigenvalues, just like any stress tensor. These eigenvalues and eigenvectors are the three solutions of

$$\mathbf{M}_0\mathbf{v} = m\mathbf{v} \qquad [38]$$

Let the eigenvalues and eigenvector be m_i, $\mathbf{v}_i$, then the moment tensor can be rewritten as

$$\mathbf{M}_0 = \sum_i m_i \mathbf{v}_i^{\mathrm{T}} \mathbf{v}_i \qquad [39]$$

Each eigenvalue–eigenvector pair represents a linear dipole, that is, two collinear forces acting in opposite directions at two points situated a small distance h away from each other. The eigenvalue represents the moment of these forces that is the product of the force by the distance h. From extensive studies of moment tensor sources, it appears that many seismic sources are very well represented by an almost pure-double couple model with $m_1 = -m_3$ and $m_2 \simeq 0$.

A great effort for calculating moment tensors for deeper sources has been made by several authors. It appears that the non-double couple part is larger for these sources but that it does not dominate the radiation. For deep sources, Knopoff and Randall (1970) proposed the so-called compensated linear vector dipole (CLVD). This is a simple linear dipole from which we subtract the volumetric part so that $m_1 + m_2 + m_3 = 0$. Thus, a CLVD is a source model where $m_2 = m_3 = -1/2\,m_1$. The linear dipole along the x-axis dominates the source but is compensated by two other linear dipoles along the other two perpendicular directions. Radiation from a CLVD is very different from that from a double couple model and many seismologists have tried to separate a double couple from a CLVD component from the moment tensor. In fact, moment tensors are better represented by their eigenvalues, separation into a fault, and a CLVD part is generally ambiguous.

Seismic moments are measured in units of Nm. Small earthquakes that produce no damage have seismic moments less than 10^{12} Nm, while the largest subduction events (such as those of Chile in 1960, Alaska in 1964, or Sumatra in 2004) have moments of the order of 10^{22}–10^{23} N m. Large destructive events (such as Izmit, Turkey 1999, Chichi, Taiwan 1999, or Landers, California 1992) have moments of the order of 10^{20} N m.

Since the late 1930s it became commonplace to measure earthquakes by their magnitude, a logarithmic measure of the total energy radiated by the earthquake. Methods for measuring radiated energy were developed by Gutenberg and Richter using well-calibrated seismic stations. At the time, the general properties of the radiated spectrum were not known and the concept of moment tensor had not yet been developed. Since at present time earthquakes are systematically measured using seismic moments, it has become standard to use the following empirical relation defined by Kanamori (1977) to convert moment tensors into a magnitude scale:

$$\log_{10} M_0(\text{in N m}) = 1.5M_w + 9.3 \qquad [40]$$

Magnitudes are easier to retain and have a clearer meaning for the general public than the more difficult concept of moment tensor.

4.02.3.5 Seismic Radiation from Moment-Tensor Sources in the Spectral Domain

In actual applications, the NF signals radiated by earthquakes may become quite complex because of multipathing, scattering, etc., so that the actually observed seismogram, say, $u(t)$ resembles the source time function $\Omega(t)$ only at long periods. It is usually verified that complexities in the wave propagation affect much less the spectral amplitudes in the Fourier transformed domain. Radiation from a simple point moment-tensor source can be obtained from [24] by straightforward Fourier transformation. Radiation from a point moment tensor in the Fourier transformed domain is then

$$\tilde{u}_c(\mathbf{x},\omega) = \frac{1}{4\pi\rho c^3}\frac{\mathcal{R}_c(\theta_0,\phi_0)}{R}\tilde{\Omega}(\omega)\mathrm{e}^{-\mathrm{i}\omega R/c} \qquad [41]$$

where $\tilde{\Omega}(\omega)$ is the Fourier transform of the source time function $\Omega(t)$. A straightforward property of any time domain Fourier transform is that the low-frequency limit of the Fourier transform is the integral of the source time function, that is,

$$\lim_{\omega\to 0}\tilde{\Omega}(\omega) = \int_0^\infty \dot{M}_0(t)\,\mathrm{d}t = M_0$$

So that in fact, the low-frequency limit of the transform of the displacement yields the total moment of the source. Unfortunately, the same notation is used to designate the total moment release by an earthquake, M_0, and the time-dependent moment $M_0(t)$.

From the observation of many earthquake spectra, and from the scaling of moment with earthquake size, Aki (1967) and Brune (1970) concluded that the seismic spectra decayed as ω^{-2} at high frequencies. Although, in general, spectra are more complex for individual earthquakes, a simple source model can be written as follows:

$$\Omega(\omega) = \frac{M_0}{1+(\omega/\omega_0)^2} \qquad [42]$$

where ω_0 is the so-called corner frequency. In this simple 'omega-squared model', seismic sources are characterized by only two independent scalar parameters: the seismic moment M_0 and the corner frequency ω_0. As mentioned earlier, not all earthquakes have displacement spectra as simple as [42], but the omega-squared model is a simple starting point for understanding seismic radiation.

From [42], it is possible to compute the spectra predicted for ground velocity:

$$\dot{\Omega}(\omega) = \frac{iM_0\omega}{1+(\omega/\omega_0)^2} \qquad [43]$$

Ground velocity spectra present a peak situated roughly at the corner frequency ω_0. In actual earthquake ground velocity spectra, this peak is usually broadened and contains oscillations and secondary peaks, but [43] is a good general representation of the spectra of ground velocity for frequencies lower than 6–7 Hz. At higher frequencies, attenuation, propagation scattering, and source effects reduce the velocity spectrum.

Finally, by an additional differentiation we get the acceleration spectra:

$$\ddot{\Omega}(\omega) = -\frac{M_0\omega^2}{1+(\omega/\omega_0)^2} \qquad [44]$$

This spectrum has an obvious problem: it predicts that ground acceleration is flat for arbitrarily high frequencies. In practice this is not the case: acceleration spectra systematically differ from [44] at high frequencies. The acceleration spectrum usually decays after a high-frequency corner identified as $f_{\max}$. The origin of this high-frequency cutoff was a subject of discussion in the 1990s, that was settled by the implicit agreement that $f_{\max}$ reflects the dissipation of high-frequency waves due to propagation in a strongly scattering medium, like the crust and near surface sediments.

It is interesting to observe that [42] is the Fourier transform of

$$\Omega(t) = \frac{M_0\omega_0}{2}\mathrm{e}^{-|\omega_0 t|} \qquad [45]$$

This is a noncausal strictly positive function, symmetric about the origin, and has an approximate width of $1/\omega_0$. By definition, the integral of the function is exactly equal to M_0. Even if this function is noncausal it shows that $1/\omega_0$ controls the width or duration of the seismic signal. At high frequencies the function behaves like ω^{-2}. This is due to the slope discontinuity of [45] at the origin, where slope changes abruptly from $M_0\omega_0^2/2t$ for $t<0$ to $-M_0\omega_0^2/2t$, that is, a total jump in slope is $-M_0\omega^2 t$. Thus, the high-frequency behavior of [45] is controlled by slope discontinuities in the source time function.

We can also interpret [42] as the absolute spectral amplitude of a causal function. There are many such functions, one of them proposed by Brune (1970) is

$$\Omega(t) = M_0\omega_0^2 t\mathrm{e}^{-\omega_0 t}H(t) \qquad [46]$$

As for [45], the width of the function is roughly $1/\omega_0$ and the high frequencies are due to the slope break of $\Omega(t)$ at the origin. This slope break has the same amplitude as that of [45] but with the opposite sign.

4.02.3.6 Seismic Energy Radiated by Point Moment-Tensor Sources

As we have already discussed for a point force, at any position sufficiently far from the source, energy flow per unit solid angle is proportional to the square of local velocity (see [17]):

$$e_c = \rho c R^2 \int_0^\infty \dot{u}_i^2(t)\,\mathrm{d}t \qquad [47]$$

where c is the P or S wave speed. Inserting the far field, or ray approximation, we can express the radiated energy density in terms of the seismic source time function using [26]:

$$e_c(\theta,\phi) = \frac{1}{16\pi^2\rho c^5}\mathcal{R}_c^2(\theta,\phi)\int_0^\infty \dot{\Omega}(t)^2\,\mathrm{d}t$$

where c stands again for P or S waves. By Parseval's theorem

$$\int_0^\infty \dot{\Omega}(t)^2\,\mathrm{d}t = \frac{1}{\pi}\int_0^\infty \omega^2\tilde{\Omega}(\omega)^2\,\mathrm{d}\omega$$

we can express the radiated energy density in terms of the seismic spectrum [42] as

$$e_c(\theta,\phi) = \frac{1}{8\pi^3\rho c^5}\mathcal{R}_c^2(\theta,\phi)\int_0^\infty \omega^2\Omega^2(\omega)\,\mathrm{d}\omega$$

where we have limited the integral over ω only to positive frequencies.

From the energy flow per unit solid angle, we can estimate the total radiated energy, or simply the seismic energy (see Boatwright, 1980):

$$E_c = \int_0^{2\pi}\int_0^{\pi} e_c(\theta,\phi)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \qquad [48]$$

so that

$$E_p = \frac{1}{2\pi^2\rho\alpha^5} < \mathcal{R}^{\mathrm{P}} >_2 \int_0^\infty \omega^2\Omega^2(\omega)\,\mathrm{d}\omega \quad \text{for P waves}$$

$$E_s = \frac{1}{2\pi^2\rho\beta^5} < \mathcal{R}^{\mathrm{S}} >_2 \int_0^\infty \omega^2\Omega^2(\omega)\,\mathrm{d}\omega \quad \text{for S waves} \qquad [49]$$

As in [18] $< \mathcal{R}^i >_2 = (1/4\pi)\int_\Omega (\mathcal{R}^i)^2\,\mathrm{d}\Omega$ is the mean square radiation pattern. It is easy to verify that, since Ω has units of moment, E_s and E_p have units of energy. For Brune's spectrum [42] the integral in [49] is

$$\int_0^\infty \omega^2\Omega^2(\omega)\,\mathrm{d}\omega = \frac{\pi}{2}M_0^2\omega_0^3$$

so that radiated energy is proportional to the square of moment. We can finally write

$$\frac{E_c}{M_0} = \frac{1}{4\pi\rho c^5} < \mathcal{R}^c >^2 M_0\omega_0^3 \qquad [50]$$

This non-dimensional relation makes no assumptions about the rupture process at the source except that the spectrum of the form [42], yet it does not seem to have been used in practical work.

Since the energy flow e_c can usually be determined in only a few directions (θ,ϕ) of the focal sphere, [48] can only be estimated, never computed very precisely. This problem still persists; in spite of the deployment of increasingly denser instrumental networks there will always be large areas of the focal sphere that remain out of the domain of seismic observations because the waves in those directions are refracted away from the station networks, energy is dissipated due to long trajectories, etc.

4.02.3.7 More Realistic Radiation Model

In reality earthquakes occur in a complex medium that is usually scattering and dissipative. Seismic waves become diffracted, reflected, and in general the suffer multipathing in those structures. Accurate seismic modeling would require perfect knowledge of those structures. It is well known and understood that those complexities dominate signals at certain frequency bands. For this reason the simple model presented here can be used to understand many features of earthquakes and the more sophisticated approaches that attempt to model every detail of the wave form are reserved only for more advanced studies. Here, like in many other areas of geophysics, a balance between simplicity and concepts must be kept against numerical complexity that may not always be warranted by lack of knowledge of the details of the structures. If the simple approach were not possible, then many standard methods to study earthquakes would be impossible to use. For instance, source mechanism, the determination of the fault angles δ, ϕ, and λ would be impossible. These essential parameters are determined by back projection of the displacement directions from the observer to a virtual unit sphere around the point source.

A good balance between simple, but robust concepts, and the sophisticated reproduction of the complex details of real wave propagation is a permanent challenge for seismologists. As we enter the

twenty-first century, numerical techniques become more and more common. Our simple models detailed above are not to be easily neglected, in any case they should always serve as test models for fully numerical methods.

4.02.4 Finite Source Models

The point source model we just discussed provides a simple approach to the simulation of seismic radiation. It is probably quite sufficient for the purpose of modeling small sources situated sufficiently far from the observer so that the source looks like a single point source. Details of the rupture process are then hidden inside the moment-tensor source time function $M_0(t)$. For larger earthquakes, and specially for earthquakes observed at distances close to the source, the point source model is not sufficient and one has to take into account the geometry of the source and the propagation of rupture across the fault. Although the first finite models of the source are quite ancient, their widespread use to model earthquakes is relatively recent and has been more extensively developed as the need to understand rupture in detail has been more pressing. The first models of a finite fault were developed simultaneously by Maruyama (1963), and Burridge and Knopoff (1964) in the general case, Ben Menahem (1961, 1962) for surface and body waves, and by Haskell (1964, 1966) who provided a very simple solution for the far field of a rectangular fault. Haskell's model became the *de facto* earthquake fault model in the late 1960s and early 1970s and was used to model many earthquakes. In the following we review the available finite source models, focusing on the two main models: the rectangular fault and the circular fault.

4.02.4.1 The Kinematic Dislocation Model

In spite of much recent progress in understanding the dynamics of earthquake ruptures, the most widely used models for interpreting seismic radiation are the so-called dislocation models. In these models the earthquake is simulated as the kinematic spreading of a displacement discontinuity (slip or dislocation in seismological usage) along a fault plane. As long as the thickness of the fault zone h is negligible with respect to the other length scales of the fault (width W and length L), the fault may be idealized as a surface of displacement discontinuity or slip. Slip is very often called dislocation by seismologists, although this is not the same as the concept of a dislocation in solid mechanics.

In its most general version, slip as a function of time and position in a dislocation model is completely arbitrary and rupture propagation may be as general as wanted. In this version the dislocation model is a perfectly legitimate description of an earthquake as the propagation of a slip episode on a fault plane. It must be remarked, however, that not all slip distributions are physically acceptable. Madariaga (1978) showed that the Haskell model, by far the most used dislocation modes, presents unacceptable features like inter-penetration of matter, releases unbounded amounts of energy, etc., that make it very difficult to use at high frequencies without important modifications. For these reasons dislocation models must be considered as a very useful intermediate step in the formulation of a physically acceptable description of rupture but examined critically when converted into dynamic models. From this perspective, dislocation models are very useful in the inversion of NF accelerograms (see, e.g., Wald and Heaton, 1994).

A finite source model can be described as a distribution of moment-tensor sources. Since we are interested in radiation from faults, we use the approximation [37] for the moment of a fault element. Each of these elementary sources produces a seismic radiation that can be computed using [24]. The total displacement seismogram observed at an arbitrary position $\mathbf{x}$ is the sum:

$$u_i(\mathbf{x},t) = \int_0^t \int_{S_{x_0}} \mu(\mathbf{x}_0)\Delta u_j(\mathbf{x}_0,\tau) \times G_{ij,k}(\mathbf{x},t|\mathbf{x}_0,\tau)n_k(\mathbf{x}_0)\,\mathrm{d}^2\mathbf{x}_0\,\mathrm{d}\tau \qquad [51]$$

where $\Delta u(\mathbf{x}_0, t)$ is the slip across the fault as a function of position on the fault ($\mathbf{x}_0$) and time t, $\mathbf{n}$ is the normal to the fault, and $\mathbf{G}(x, t)$ is the elastodynamic Green tensor that may be computed using simple layered models of the crustal structure, or more complex finite difference simulations.

In a first, simple analysis, we can use the ray approximation [26] that often provides a very good approximation in the far field. Inserting [26] into [51] and after some simplification, we get

$$u_c(\mathbf{x},t) = \frac{1}{4\pi\rho c^3}\int_0^t \int_{S_{x_0}} \frac{\mathcal{R}^c_{ijk}(\theta,\phi)}{R}\mu\Delta\dot{u}_j \times \left[\mathbf{x}_0, t-\tau-\frac{R(\mathbf{x}-\mathbf{x}_0)}{c}\right] n_k\,\mathrm{d}^2\mathbf{x}_0\,\mathrm{d}\tau \qquad [52]$$

where $R(\mathbf{x}-\mathbf{x}_0)$ is the distance between the observer and a source point located at $\mathbf{x}_0$. In almost all applications the reference point is the hypocenter, the point where rupture initiates.

In [52] both the radiation pattern $\mathcal{R}_c$ and the geometrical decay $1/R$ change with position on the fault. In the far field, according to ray theory, we can make the approximation that only phase changes are important so that we can approximate the integral [52] assuming that both radiation pattern and geometrical spreading do not change significantly across the fault. In the far field we can also make the Fraunhoffer approximation:

$$R(\mathbf{x}-\mathbf{x}_0) \simeq R(\mathbf{x}-\mathbf{x}_H) - \mathbf{e}_r \cdot (\mathbf{x}_0 - \mathbf{x}_H)$$

where $\mathbf{x}_H$ is a reference point on the fault, usually the hypocenter, and $\mathbf{e}_r$ is the unit vector in the radial direction from the reference point to the observer. With these approximations, FF radiation from a finite source is again given by the generic expression [26] where the source time function Ω is replaced by

$$\Omega(t,\theta,\phi) = \mu \int_0^t \int_{S_{x_0}} \Delta\dot{u}_j \times \left[\xi_1, \xi_2, t-\tau+\frac{\mathbf{e}_r\cdot\xi)}{c}\right] n_k \, \mathrm{d}\xi_1 \, \mathrm{d}\xi_2 \, \mathrm{d}\tau \qquad [53]$$

where ξ is a vector of components (ξ_1, ξ_2) that measures position on the fault with respect to the hypocenter $\mathbf{x}_H$. The main difference between a point and a finite source as observed from the far field is that in the finite case the source time function Ω depends on the direction of radiation through the term $\mathbf{e}_r \cdot \xi$. This directivity of seismic radiation can be very large when ruptures propagate at high subshear or intersonic speeds.

The functional [53] is linear in slip rate amplitude but very nonlinear with respect to rupture propagation which is implicit in the time dependence of $\Delta\dot{u}$. For this reason, in most inversions, the kinematics of the rupture process (position of rupture front as a function of time) is simplified. The most common assumption is to assume that rupture propagates at constant speed away from the hypocenter. Different approaches have been proposed in the literature in order to approximately invert for variations in rupture speed about the assumed constant rupture velocity (see, e.g., Cotton and Campillo, 1995; Wald and Heaton, 1994).

4.02.4.1.1 Haskell's model

One of the most widely used dislocation models was introduced by Haskell (1964, 1966). In this model, shown in **Figure 4**, a uniform displacement discontinuity spreads at constant rupture velocity inside a rectangular-shaped fault. At low frequencies, or wavelengths much longer than the size of the fault, this model is a reasonable approximation to a simple seismic rupture propagating along a strike slip fault.

In Haskell's model at time $t=0$ a line of dislocation of width W appears suddenly and propagates along the fault length at a constant rupture velocity until a region of length L of the fault has been broken. As the dislocation moves it leaves behind a zone of constant slip D. Assuming that the fault lies on a plane of coordinates (ξ_1, ξ_2), the slip function can be written (see also **Figure 4**) as

$$\Delta\dot{u}_1(\xi_1,\xi_2,t) = D\,\dot{s}(t-\xi_1/v_r)H(\xi_1)H(L-\xi_1) \quad \text{for} \quad -W/2 < \xi_2 < W/2 \qquad [54]$$

where $\dot{s}(t)$ is the sliprate time function that, in the simplest version of Haskell's model, is invariant with position on the fault. The most important feature of this model is the propagation of rupture implicit in the time delay of rupture ξ_1/v_r. v_r is the rupture velocity, the speed with which the rupture front propagates along the fault in the ξ_1-direction. An obvious unphysical feature of this model is that rupture appears instantaneously in the ξ_2-direction; this is of course impossible for a spontaneous seismic rupture. The other inadmissible feature of the Haskell model is the fact that on its borders slip

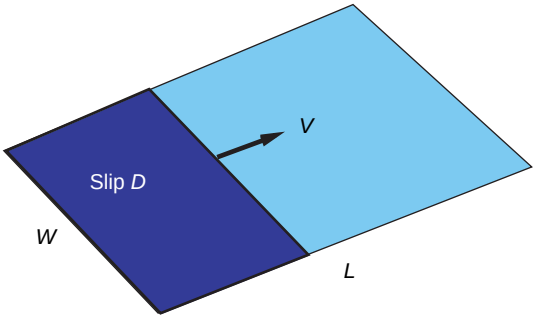


Figure 4 Figure Haskell's kinematic model, one of the simplest possible earthquake models. The fault has a rectangular shape and a linear rupture front propagates from one end of the fault to the other at constant rupture speed *v*. Slip in the broken part of the fault is uniform and equal to *D*.

suddenly jumps from the average slip D to zero. This violates material continuity so that the most basic equation of motion [1] is no longer valid near the edges of the fault. In spite of these two obvious shortcomings, Haskell's model gives a simple, first-order approximation to seismic slip, fault finiteness, and finite rupture speed. The seismic moment of Haskell's model is easy to compute, the fault area is $L \times W$, and slip D is constant over the fault, so that the seismic moment is $M_0 = \mu DLW$. Using the far field, or ray approximation, we can compute the radiated field from Haskell's model. It is given by the ray expression [26] where, using [53], the source time function Ω is now a function not only of time but also of the direction of radiation:

$$\Omega_{\mathrm{H}}(t,\theta,\phi) = \mu \int_{-W/2}^{W/2} \mathrm{d}\xi_2 \times \int_0^L D\dot{s}\left(t - \frac{\xi_1}{v_{\mathrm{r}}} + \frac{\xi_1}{c}\cos\phi\sin\theta + \frac{\xi_2}{c}\sin\phi\sin\theta\right)\mathrm{d}\xi_1 \qquad [55]$$

where we used the index H to indicate that this is the Haskell model. The two integrals can be evaluated very easily for an observer situated along the axis of the fault, that is, when $\phi = 0$. Integrating we get

$$\Omega_{\mathrm{H}}(\theta,0,t) = M_0 \frac{1}{T_M}\int_0^{\min(t,T_M)} \dot{s}(t-\tau)\,\mathrm{d}\tau \qquad [56]$$

where $T_M = L/c(1 - v_{\mathrm{r}}/c\sin\theta)$. Thus, the FF signal is a simple integral over time of the source slip rate function. In other directions the source time function Ω_{H} is more complex but it can be easily computed by the method of isochrones that is explained later. Radiation from Haskell's model shows two very fundamental properties of seismic radiation: finite duration, given by T_{M}; and directivity, that is, the duration and amplitude of seismic waves depends on the azimuthal angle of radiation/theta.

A similar computation in the frequency domain was made by Haskell (1966). In our notation the result is

$$\tilde{\Omega}_{\mathrm{H}}(\theta,0,\omega) = M_0\,\mathrm{sinc}(\omega T_M/2)\mathrm{e}^{-\mathrm{i}\omega T_M/2}\tilde{\dot{s}}(\omega) \qquad [57]$$

where $\mathrm{sinc}(x) = \sin(x)/x$.

It is often assumed that the slip rate time function $\dot{s}(t)$ is a boxcar function of amplitude $1/\tau_{\mathrm{r}}$ and duration τ_{r}, the rise time. In that case the spectrum, in the frequency domain, $\Omega(\omega)$, becomes

$$\tilde{\Omega}_{\mathrm{H}}(\theta,0,\omega) = M_0\,\mathrm{sinc}(\omega T_M/2) \times \mathrm{sinc}(\omega\tau_{\mathrm{r}}/2)\mathrm{e}^{-\mathrm{i}\omega(T_M+\tau_{\mathrm{r}})/2} \qquad [58]$$

or, in the time domain,

$$\Omega_{\mathrm{H}}(\theta,0,t) = M_0\,\mathrm{boxcar}[t, T_M] * \mathrm{boxcar}[t,\tau_{\mathrm{r}}]$$

where the star '$*$' means time convolution and boxcar is a function of unit area that is zero everywhere except that in the time interval from 0 to τ_r where it is equal to $1/\tau_r$. Thus, Ω_{H} is a simple trapezoidal pulse of area M_0 and duration $T_d = T_M + \tau_r$. This surprisingly simple source time function matches the ω-squared model for the FF spectrum since Ω_{H} is flat at low frequencies and decays like ω^{-2} at high frequencies. The spectrum has two corners associated with the pulse duration T_M and the other with rise time τ_{r}. This result is however only valid for radiation along the plane $\phi = 0$ or $\phi = \pi$. In other directions with $\phi \neq 0$, radiation is more complex and the high-frequency decay is of order ω^{-3}, faster than in the classical Brune model.

In spite of some obvious mechanical shortcomings, Haskell's model captures some of the most important features of an earthquake and has been extensively used to invert for seismic source parameters both in the near and far field from seismic and geodetic data. The complete seismic radiation for Haskell's model was computed by Madariaga (1978) who showed that, because of the stress singularities around the edges, the Haskell model can only be considered as a rough low-frequency approximation of fault slip.

4.02.4.2 The Circular Fault Model

The other simple source model that has been widely used in earthquake source seismology is a circular crack model. This model was introduced by several authors including Savage (1966), Brune (1970), and Keilis-Borok (1959) to quantify a simple source model that was mechanically acceptable and to relate slip on a fault to stress changes. As already mentioned, dislocation models such as Haskel's produce nonintegrable stress changes due to the violation of material continuity at the edges of the fault. A natural approach to model earthquakes was to assume that the earthquake fault was circular from the beginning, with rupture starting from a point and then propagating self-similarly until it finally stopped at a certain source radius. This model was carefully studied in the 1970s and a complete understanding of it is available without getting into the details of dynamic models.

4.02.4.2.1 Kostrov's Self-Similar Circular Crack

The simplest possible crack model is that of a circular rupture that starts form a point and then spreads self-similarly at constant rupture speed v_r without ever stopping. Slip on this fault is driven by stress drop inside the fault. The solution of this problem is somewhat difficult to obtain because it requires very advanced use of self-similar solutions to the wave equation and its complete solution for displacements and stresses must be computed using the Cagniard de Hoop method (Richards, 1976). Fortunately, the solution for slip across the fault found by Kostrov (1964) is surprisingly simple:

$$\Delta u_x(r,t) = C(v_r)\frac{\Delta\sigma}{\mu}\sqrt{v_r^2 t^2 - r^2} \qquad [59]$$

where r is the radius in a cylindrical coordinate system centered on the point of rupture initiation. $v_r t$ is the instantaneous radius of the rupture at time t. $\Delta\sigma$ is the constant stress drop inside the rupture zone, μ is the elastic rigidity, and $C(v_r)$ is a very slowly varying function of the rupture velocity. For most practical purposes $C \approx 1$. As shown by Kostrov (1964), inside the fault, the stress change produced by the slip function [59] is constant and equal to $\Delta\sigma$. This simple solution provides a very basic result that is one of the most important properties of circular cracks. Slip in the fault scales with the 'ratio of stress drop over rigidity "times" the instantaneous radius of the fault'. As rupture develops, all the displacements around the fault scale with the size of the rupture zone.

The circular self-similar rupture model produces FF seismic radiation with a very peculiar signature. Inserting the slip function into the expression for FF radiation [52], we get

$$\Omega_K(t,\theta) = A(v_r,\theta)t^2 H(t)$$

where we used an index K to indicate Kostrov's model. The amplitude coefficient A is

$$A(v_r,\theta) = C(v_r)\frac{2\pi}{1 - v_r^2/c^2\sin^2\theta}\Delta\sigma v_r^3$$

(see Richards, 1976, Boatwright, 1980). Thus, the initial rise of the FF source time function is proportional to t^2 for Kostrov's model. The rate of growth is affected by a directivity factor that is different from that of the Haskell model, directivity for a circular crack being generally weaker. The corresponding spectral behavior of the source time function is $\tilde{\Omega}(\omega,\theta,\phi) \simeq \omega^{-3}$, which is steeper than Brune's (1970) inverse omega-squared decay model.

4.02.4.2.2 The Kinematic Circular Source Model of Sato and Hirasawa

The simple Kostrov self-similar crack is not a good seismic source model for two reasons: (1) rupture never stops so that the seismic waves emitted by this source increase like t^2 without limit, and (2) it does not explain the high-frequency radiation from seismic sources. Sato and Hirasawa (1973) proposed a modification of the Kostrov model that retained its initial rupture behavior [59] but added the stopping of rupture. They assumed that the Kostrov-like growth of the fault was suddenly stopped across the fault when rupture reached a final radius a (see **Figure 5**). In mathematical terms the slip function is

$$\Delta u_x(r,t) = \begin{cases} C(v_r)\dfrac{\Delta\sigma}{\mu}\sqrt{v_r^2 t^2 - r^2}\,H(v_r t - r) & \text{for } t < a/v_r \\ C(v_r)\dfrac{\Delta\sigma}{\mu}\sqrt{a^2 - r^2}\,H(a - r) & \text{for } t > a/v_r \end{cases} \qquad [60]$$

Thus, at $t = a/v_r$ the slip on the fault becomes frozen and no motion occurs thereafter. This mode of healing is noncausal but the solution is mechanically acceptable because slip near the borders of the fault always tapers like a square root of the distance to the fault tip. Using the FF radiation approximation [52], Sato and Hirasawa found that the source time function for this model could be computed exactly

$$\Omega_{SH}(t,\theta) = C(v_r)\frac{2\pi}{1 - v_r^2/c^2\sin^2\theta}\Delta\sigma v_r^3 t^2 \qquad [61]$$

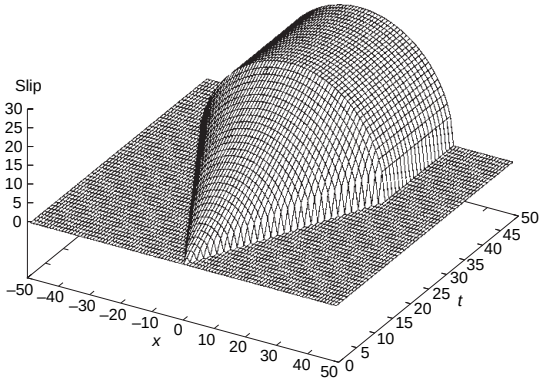


Figure 5 Slip distribution as a function of time on Sato and Hirasawa circular dislocation model.

for $t < L/v_r(1 - v_r/c \sin\theta)$, where θ is the polar angle of the observer. As should have been expected the initial rise of the radiated field is the same as in the Kostrov model, the initial phase of the source time function increases very fast like t^2. After the rupture stops the radiated field is

$$\Omega_{SH}(t,\theta) = C(v_r)\frac{\pi}{2}\frac{1}{v_r/c\sin\theta} \times \left[1 - \frac{v_r^2 t^2}{a^2(1 + v_r^2/c^2\cos^2\theta)}\right]\Delta\sigma a^2 v_r \quad [62]$$

for times between $t_{s1} = a/v_r(1 - v_r/c\sin\theta)$ and $t_{s2} = a/v_r(1 + v_r/c\sin\theta)$, radiation from the stopping process is spread in the time interval between the two stopping phases emitted from the closest (t_{s1}) and the farthest (t_{s2}) points of the fault. These stopping phases contain the directivity factors $(1 \pm v_r/c\sin\theta)$ which appear because, as seen from different observation angles θ, the waves from the edges of the fault put more or less time to cross the fault. The last term in both ([61] and [62]) has the dimensions of moment rate as expected.

It is also possible to compute the spectrum of the FF signal ([61] and [62]) analytically. This was done by Sato and Hirasawa (1973). The important feature of the spectrum is that it is dominated by the stopping phases at times t_{s1} and t_{s2}. The stopping phases are both associated with a slope discontinuity of the source time function. This simple model explains one of the most universal features of seismic sources: the high frequencies radiated by seismic sources are dominated by stopping phases not by the energy radiated from the initiation of seismic rupture (Savage, 1966). These stopping phases appear also in the quasi-dynamic model by Madariaga (1976) although they are somewhat more complex that in the present kinematic model.

4.02.4.3 Generalization of Kinematic Models and the Isochrone Method

A simple yet powerful method for understanding the general properties of seismic radiation from classical dislocation models was proposed by Bernard and Madariaga (1984). The method was recently extended to study radiation from supershear ruptures by Bernard and Baumont (2005). The idea is that since most of the energy radiated from the fault comes from the rupture front, it should be possible to find where this energy is coming from at a given station and at a given time. Bernard and Madariaga (1984) originally derived the isochrone method by inserting the ray theoretical expression [26] into the representation theorem, a technique that is applicable not only in the far field but also in the immediate vicinity of the fault at high frequencies. Here, for the purpose of simplicity, we will derive isochrones only in the far field. For that purpose we study the FF source time function for a finite fault derived in [53]. We assume that the slip rate distribution has the general form

$$\Delta\dot{u}_i(\xi_1,\xi_2,t) = D_i(t - \tau(\xi_1,\xi_2)) = D_i(t) * \delta(t - \tau(\xi_1,\xi_2)) \quad [63]$$

where $\tau(\xi_1,\xi_2)$ is the rupture delay time at a point of coordinates ξ_1, ξ_2 on the fault. This is the time it takes for rupture to arrive at that point. The star indicates time domain convolution. We rewrite [63] as a convolution in order to distinguish between the slip time function $D(t)$ and its propagation along the fault described by the argument to the delta function. While we assume here that $D(t)$ is strictly the same everywhere on the fault, in the spirit of ray theory our result can be immediately applied to a problem where $D(\xi_1, \xi_2, t)$ is a slowly variable function of position. Inserting the slip rate field [63] in the source time function [53], we get

$$\Omega(t,\theta,\phi) = \mu D_i(t) * \int_0^t \int_{S_0} \delta\left[t - \tau(\xi_1,\xi_2) - \frac{\mathbf{e}\cdot\mathbf{x}_0}{c}\right] d^2\boldsymbol{x}_0\, d\tau \quad [64]$$

where the star indicates time domain convolution. Using the sifting property of the delta function, the integral over the fault surface S_0 reduces to an integral over a line defined implicitly by

$$t = \tau(\xi_1,\xi_2) + \frac{\mathbf{e}\cdot\mathbf{x}_0}{c} \quad [65]$$

the solutions of this equation define one or more curves on the fault surface (see **Figure 6**). For every value of time, eqn [65] defines a curve on the fault that we call an isochrone.

The integral over the surface in [64] can now be reduced to an integral over the isochrone using standard properties of the delta function

$$\Omega(t,\theta,\phi) = \mu D_i(t) * \int_{l(t)} \frac{dt}{dn} dl \quad [66]$$

$$= \mu D_i(t) * \int_{l(t)} \frac{v_r}{(1 - v_r/c\cos\psi)} dl \quad [67]$$

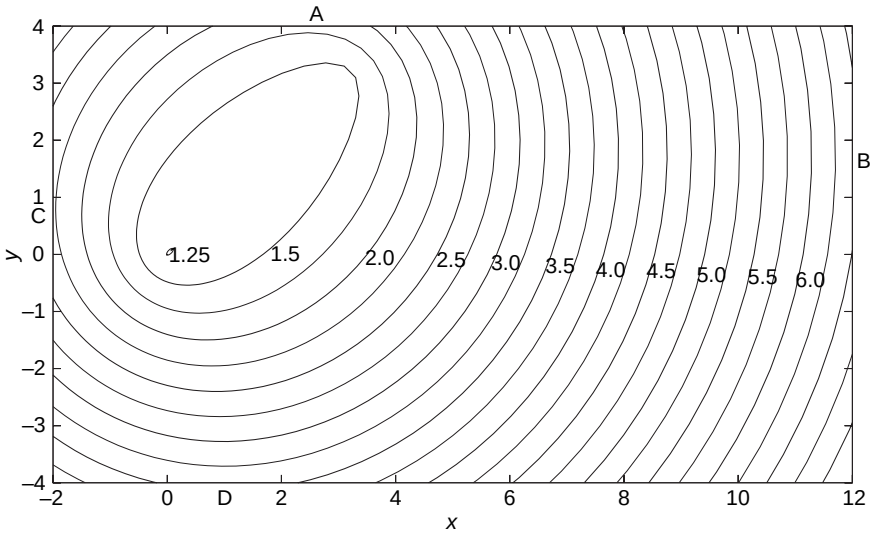


Figure 6 Example of an isochrone. The isochrone was computed for an observer situated at a point of coordinates (3, 3, 1) in a coordinate system with origin at the rupture initiation point (0, 0). The vertical axis is out of the fault plane. Rupture starts at $t = 0$ at the origin and propagates outwards at a speed of 90% the shear wave speed that is 3.5 km s^{-1} in this computation. The signal from the origin arrives at $t = 1.25$ s at the observation point. Points A–D denote the location on the border of the fault where isochrones break, producing strong stopping phases.

where $\ell(t)$ is the isochrone, and $\mathrm{d}t/\mathrm{d}n = \mathbf{n}\cdot\nabla_{x_0} t = v_r/(1 - v_r/c\cos\psi)$ is the derivative of t in the direction perpendicular to the isochrone. Actually, as shown by Bernard and Madariaga (1984), $\mathrm{d}t/\mathrm{d}n$ is the local directivity of the radiation from the isochrone. In general, both the isochrone and the normal derivative $\mathrm{d}t/\mathrm{d}n$ have to be evaluated numerically. The meaning of [66] is simple; the source time function at any instant of time is an integral of the directivity over the isochrone.

The isochrone summation method has been presented in the simpler case here, using the far field (or Fraunhofer approximation). The method can be used to compute synthetics in the near field too; in that case changes in the radiation pattern and distance from the source and observer may be included in the computation of the integral [66] without any trouble. The results are excellent as shown by Bernard and Madariaga (1984) who computed synthetic seismograms for a buried circular fault in a half-space and compared them to full numerical synthetics computed by Bouchon (1982). With improvements in computer speed the use of isochrones to compute synthetics is less attractive and, although the method can be extended to complex media within the ray approximation, most modern computations of synthetics require the appropriate modeling of multipathing, channeled waves, etc., that are difficult to integrate into the isochrone method. Isochrones are still very useful to understand many features of the radiated field and its connection to the rupture process (see, e.g., Bernard and Baumont, 2005).

4.02.5 Crack Models of Seismic Sources

As mentioned several times dislocation models capture some of the most basic geometrical properties of seismic sources, but have several unphysical features that require careful consideration. For small earthquakes, the kinematic models are generally sufficient, while for larger events – specially in the near field – dislocation models are inadequate because they may not be used to predict high-frequency radiation. A better model of seismic rupture is of course a crack model like Kostrov's self-similar crack. In crack models slip and stresses are related near the crack tips in a very precise way, so that a finite amount of energy is stored in the vicinity of the crack. Griffith (1920) introduced crack theory using the only requirement that the appearance of a crack in a body does two things: (1) it relaxes stresses and (2) it releases a finite amount of energy. This simple requirement is enough to define

many of the properties of cracks in particular energy balance (see, e.g., Rice, 1980; Kostrov and Das, 1988; Freund, 1989).

Let us consider the main features of a crack model. Using **Figure 7**, we consider a planar fault lying on the plane x, y with normal z. Although the rupture front may have any shape, it is simpler to consider a linear rupture front perpendicular to the x-axis and moving at speed v_r in the positive x-direction. Three modes of fracture can be defined with respect to the configuration of **Figure 7**:

- Antiplane, mode III or SH, when slip is in the y-direction and stress drops also in this direction, that is, stress σ_{zy} is relaxed by slip.
- Plane, or mode II, when slip is in the x-direction and stress drops also in this direction, that is, stress σ_{zx} is relaxed by this mode.
- Opening, or mode I, when the fault opens with a displacement discontinuity in the z-direction. In this case stress σ_{zz} drops to zero.

In natural earthquakes, the opening mode is unlikely to occur on large scales although it is perfectly possible for very small cracks to appear due to stress concentrations, geometrical discontinuities of the fault, etc.

For real ruptures, when the rupture front is a curve (or several disjoint ruptures if the source is complex), modes II and III will occur simultaneously on the periphery of the crack. This occurs even in the simple self-similar circular crack model studied earlier. Fortunately, in homogeneous media, except near sharp corners and strong discontinuities, the two modes are locally uncoupled, so that most features determined in 2D carry over to 3D cracks with little change.

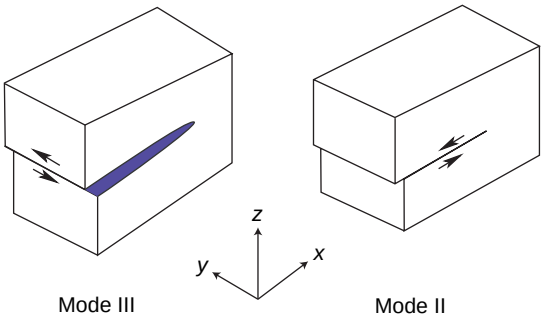


Figure 7 Modes of rupture for shear faulting. Mode III or antiplane mode and mode II or inplane mode may occur at different places on fault boundaries. For general faulting models both modes occur simultaneously.

In order to study a two-dimensional crack model we solve the elastodynamic wave equation together with the following boundary conditions on the $z=0$ plane; for antiplane cracks, mode III,

$$\begin{aligned} \sigma_{zy}(x,0) &= \Delta\sigma \quad \text{for} \quad x < l(t) \\ u_y(x,0) &= 0 \qquad \text{for} \quad x < l(t) \end{aligned} \tag{68}$$

For plane cracks, mode II,

$$\begin{aligned} \sigma_{zx}(x,0) &= \Delta\sigma \quad \text{for} \quad x < l(t) \\ u_x(x,0) &= 0 \quad \text{for} \quad x < l(t) \end{aligned} \tag{69}$$

where $\ell(t)$ is the current position of the rupture front on the x-axis. These boundary conditions constitute a mixed boundary-value problem that can be solved using complex variable techniques. The solution for arbitrary time variation of $\ell(t)$ was found for mode III by Kostrov (1964). For plane ruptures, the solution for arbitrary $\ell(t)$ was found by Freund (1972) (see also Kostrov and Das (1988)). Eshelby (1969) showed that the crack problems have a number of universal features which are independent of the history of crack propagation and depend only on the instantaneous rupture speed. Like electrical charges in an electromagnetic field, in the crack approximation a rupture front has no memory, stresses, and particle velocities around the crack front have some universal features.

4.02.5.1 Rupture Front Mechanics

Since stresses and velocities around a rupture front have universal properties, we can determine them by studying the simpler crack that propagates indefinitely at constant speed. This can be done using a Lorentz transformation of the static elasticity. We are not going to enter into details of the determination of the solution of the wave equation in moving coordinates; very succinctly the stress and velocity fields around the crack tip are related by a nonlinear eigenvalue problem determined by Muskhelishvili in his classical work on complex potentials. There are an infinite number of solutions of the problem, but only one of them insures finite energy flow into the crack tip. All other produce no flow or infinite flow.

Along the fault stress and particle velocities have the universal forms (see **Figure 8**):

$$\begin{aligned} \sigma(x) &= \frac{K}{\sqrt{2\pi}} \frac{1}{[x-l(t)]^{1/2}} \qquad \text{for} \quad x > l(t) \\ \Delta\dot{u}(x) &= \frac{V}{\sqrt{2\pi}} \frac{1}{[x-l(t)]^{1/2}} \qquad \text{for} \quad x < l(t) \end{aligned} \tag{70}$$

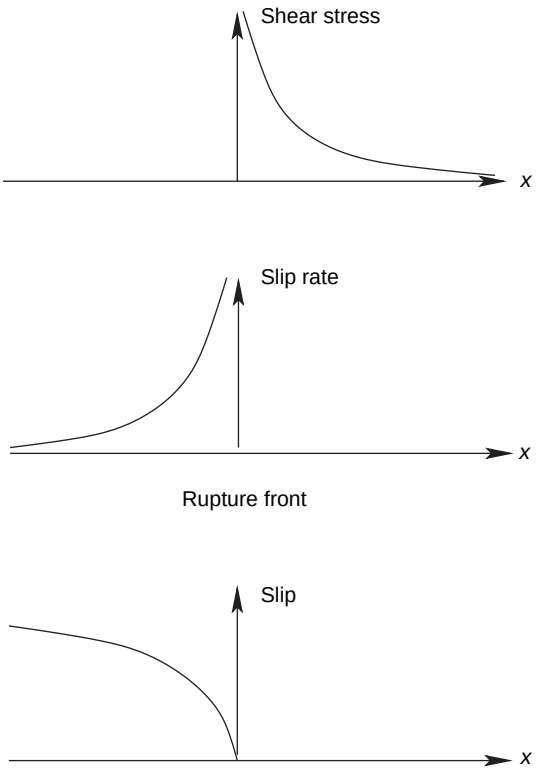


Figure 8 State of stress and slip velocity near the tip of a fracture propagating at subsonic rupture speed.

and the relations [69] or [68] on the rest of the line. Here σ stands for either σ_{yz} or σ_{xz}, and $\Delta\dot{u}$ for the corresponding slip velocity component in either antiplane or plane fracture modes. In [70] K is the stress concentration, a quantity with units of Pa $m^{1/2}$ that represents the strength of the stress field near the rupture front. V is the dynamic slip rate intensity which is related to K by

$$K = \frac{\mu}{2v_r}\sqrt{1 - v_r^2/\beta^2}\,V \qquad [71]$$

in antiplane cracks, where μ and β are the elastic rigidity and shear wave speed, respectively. For plane cracks the relation is more complicated:

$$K = \frac{\mu}{2v_r}\frac{\beta^2}{v_r^2}\frac{R(v_r)}{\sqrt{1 - v_r^2/\beta^2}}V \qquad [72]$$

where $R(v_r)$ is Rayleigh's function

$$R(v_r) = 4\sqrt{1 - v_r^2/\beta^2}\sqrt{1 - v_r^2/\alpha^2} - (2 - v_r^2/\beta^2)^2$$

The complete angular dependence of the stress and particle velocity fields is given by Freund (1989). The inverse-squared-root singularities of the form [70] only occur if the rupture velocity is less than the classical limiting rupture speeds, shear wave velocity β, for antiplane ruptures and the Rayleigh wave speed $c_R \approx 0.91\beta$ for inplane cracks. At the terminal speed the coefficient relating K and V reduces to zero, which means that at the terminal speed there is no stress concentration. A crack running at the terminal speed releases no energy; this speed is thus only possible if there is no rupture resistance and no friction on the fault.

Both experimental and observational evidence cited by Rosakis *et al.* (1999) showed that it is possible for mode II shear cracks to propagate at speeds faster than the shear wave speed in a so-called supershear mode first put in evidence numerically by Andrews (1976). For supershear speeds the stress and velocity concentration have different dependence with distance than the squared root for subshear faults except for a very particular rupture speed $\sqrt{2}\beta$.

The stress field in [70] is infinite at the rupture front; this is a consequence of the idealization that was made to obtain these results: it was assumed that the material around the fault remained elastic even in the immediate vicinity of the rupture front. If more realistic frictional boundary conditions are used instead of the abrupt discontinuities implied by [69] and [68], then the singularity disappears inside a small zone called the rupture process, or slip weakening zone. Many global features of the crack model can be derived with the simple elastic model studied here; more detailed studies involving a finite slip weakening zone are only needed to study crack growth in detail.

4.02.5.2 Stress and Velocity Intensity

K and V in [70] have a more simpler fundamental structure that was discovered by Eshelby (1969) for the antiplane case and later extended to plane shear cracks by many authors as reviewed by Freund (1989). Both the dynamic stress intensity factor and the velocity intensity factors can be written in the form

$$\begin{aligned} K &= k(v_r)K^* \\ V &= b(v_r)V^* \end{aligned} \qquad [73]$$

where, for an antiplane crack,

$$k_{III}(v_r) = \sqrt{1 - v_r/\beta}$$

and, for an inplane crack,

$$k_{\mathrm{II}}(v_{\mathrm{r}}) \approx 1 - v_{\mathrm{r}}/c_{\mathrm{r}}$$

is a very good approximation, the exact expression for k_{II} can be found in the books by Freund (1989) and Kostrov and Das (1988).

The factors K^* and V^* in [73] depend only on the load applied to the fault. In the case of an earthquake, they are determined by the stress drop inside those segments of the fault that have already slipped. Since $k(v_{\mathrm{r}}) \to 0$ as the rupture velocity tends to zero, K^* is simply the stress intensity that would prevail if the rupture velocity dropped instantaneously to zero. Thus, K^* is often called the zero-speed stress intensity factor; it depends only on the history of rupture and stress drop inside the broken parts of the fault. Some authors interpret this property of faults mechanics as meaning that rupture fronts have no inertia, their rupture speed can change instantaneously if rupture resistance increases or if the rupture front encounters some geometrical discontinuity like a fault jog or a kink.

4.02.5.3 Energy Flow into the Rupture Front

We already mentioned that the stress and slip singularities are a consequence of the requirement that there is a finite energy flow into the rupture front. This energy is used to create new fault surface and is spent in overcoming frictional resistance of the fault. The energy flow into the crack tip was first computed by Kostrov and Nikitin (1970) who provided a very complete discussion of the problem.

The energy flow per unit area of crack advance is

$$G = \frac{1}{4v_{\mathrm{r}}} KV \qquad [74]$$

for all the modes, so that for mode III,

$$G_{\mathrm{III}} = \frac{1}{2\mu} \frac{1}{\sqrt{1 - v_{\mathrm{r}}^2/\beta^2}} K^2$$

and for mode II,

$$G_{\mathrm{II}} = \frac{1}{2\mu} \frac{v_{\mathrm{r}}^2}{\beta^2} \frac{\sqrt{1 - v_{\mathrm{r}}^2/\beta^2}}{R(v_{\mathrm{r}})} K^2$$

Let us note that K^2 tends to zero as the rupture velocity v_{r} approaches the terminal speed faster than the Rayleigh function, so that G_{II} vanishes at the terminal speed.

The crack models are mostly concerned with the local conditions near the edge of the fault as it propagates inside the elastic medium. This is the principal subject of fracture mechanics. In seismology we are interested not just on the growth of ruptures, but also on the generation of seismic waves. Earthquakes are three dimensional and the finiteness of the source plays a fundamental role in the generation of seismic waves.

4.02.5.4 The Circular Crack

The simplest fault model that can be imagined is a simple circular crack that grows from a point at a constant, or variable rupture speed and then stops on the rim of the fault arrested by the presence of unbreakable barriers. The first such simple model was proposed by Madariaga (1976). Although this model is unlikely to represent any actual earthquake, it does quite a good job in explaining many features that are an intrinsic part of seismic sources, most notably the scaling of different measurable quantities such as slip, slip rate, stress change, and energy release. The circular crack problem is posed as a crack problem, that is, in terms of stresses not of slip, but the rupture process is fixed in advance so that rupture does not develop spontaneously. This is the only unrealistic feature of this model and it is the reason called quasidynamic, that is, rupture is kinematically defined, but slip is computed solving the elastodynamic equations.

4.02.5.4.1 The static circular crack

We start by a quick study of a simple circular crack from which we derive some of the most fundamental properties of dynamic source models. Let us consider a static circular ("penny shaped") crack of radius a lying on the x, y plane. Assuming that the fault is loaded by an initial shear stress σ_{xz}^0 and that the stress drop

$$\Delta\sigma = \sigma_{xz}^0 - \sigma_{xz}^f$$

is uniform, where σ_{xz}^f is the final, residual stress in the fault zone, the slip on the fault is given by

$$\Delta u_x(r) = D(r) = \frac{24}{7\pi} \frac{\Delta\sigma}{\mu} \sqrt{a^2 - r^2} \qquad [75]$$

where r is the radial distance from the center of the crack on the (x, y) plane, a is the radius of the crack, and μ is the elastic rigidity of the medium surrounding the crack. Slip in this model has the typical ellipsoidal shape that we associate with cracks and is very different from the constant slip inside the fault

assumed in Haskell's model. The taper of the slip near the edges of the crack is of course in agreement with what we discussed about the properties of the elastic fields near the edge of the fault. From [75] we can determine the scalar seismic moment for this circular fault:

$$M_0 = \frac{16}{7}\Delta\sigma a^3$$

so that the moment is the product of the stress drop times the cube of the fault size. This simple relation is the basis of the seismic scaling law proposed by Aki (1967). The circular crack model has been used to quantify numerous small earthquakes for which the moment was estimated from the amplitude of seismic waves, and the source radius was estimated from corner frequencies, aftershock distribution, etc., the result is that for shallow earthquakes in crustal seismogenic zones like the San Andreas fault, or the North Anatolian fault in Turkey, stress drops are of the order of 1–10 MPa. For deeper events in subduction zones, stress drops can reach several tens of MPa. Thus, stress drop of earthquakes does not change much, at most a couple of order of magnitudes, while source radius varies over several orders of magnitudes from meters to 100 km or more. It is only in this sense that should be taken the usual assertion "stress drop in earthquakes is constant"; it actually changes but much less than the other parameters in the scaling law.

Finally, let us take a brief view of the stress field in the vicinity of the fault radius. As expected for crack models the stress presents stress concentrations of the type [70], that is,

$$\sigma_{xz}(r,\phi) = \frac{(K_{\mathrm{II}}\cos\phi + K_{\mathrm{III}}\sin\phi)}{\sqrt{r-a}}$$

where (r, ϕ) are polar coordinates on the plane of the circular fault with ϕ being measured from the x-axis. The stress intensity factors are

$$K_{\mathrm{II}} = 0.515\Delta\sigma\sqrt{a}$$
$$K_{\mathrm{III}} = 0.385\Delta\sigma\sqrt{a}$$

the numerical coefficients were computed using a Poisson ratio of 1/4. It is interesting to note that even if the slip distribution [75] was radially symmetrical, the stress distribution is not. Stress concentration in the mode II direction is larger than in the antiplane one. As a consequence, if rupture resistance is the same in plane and antiplane modes, a circular crack has an unstable shape. This is clearly observed in fully dynamic simulations where the faults become invariably elongated in the inplane direction.

4.02.5.4.2 The quasidynamic circular crack

There are no simple analytical solutions equivalent to that of Sato and Hirasawa (1973) for quasidynamic cracks. We are forced to use numerical solutions that are actually very simple to obtain using either finite difference or boundary integral equation techniques. The full solution to the circular crack problem is shown in **Figure 9**. Initially, until the sudden arrest of rupture at the final radius a, the slip distribution can be accurately computed using Kostrov's self-similar solution [59]. The stopping of rupture generates strong healing waves that propagate inwards from the rim of the fault. These waves are of three types: P, S, and Rayleigh waves. Soon after the passage of the Rayleigh waves, slip rate inside the fault decreases to zero and the fault heals. After healing, we assume that frictional forces are sufficiently strong that no slip will occur until the fault is reloaded. As observed in **Figure 9** it is clear that slip and rise time are functions of position on the fault, the rise time being much longer near the center where slip is also larger than near the edges of the fault where slip is clamped. Finally, let us note that the slip after healing is very similar to that of a static circular crack, except that there is an overshoot of slip with respect to the static solution [75]. The overshoot is of course a function of the rupture speed, but its maximum value is of the order of 15% for a rupture speed of 0.75β.

4.02.6 Conclusions

The study of seismic radiation from realistic source models has reached now its maturity. Seismologists have been able to invert the rupture process of a number of earthquakes and many of the features predicted by simple dynamic source models have been quantified and observed. Foremost among these is the shape of the FF spectrum, the basic scaling laws relating particle velocity and acceleration to properties of the fault, such as size, stress drop and rupture velocity. The frontier today is the accurate estimation and interpretation of seismic energy and, therefore, the quantification of radiation in terms of the energy balance of seismic sources.

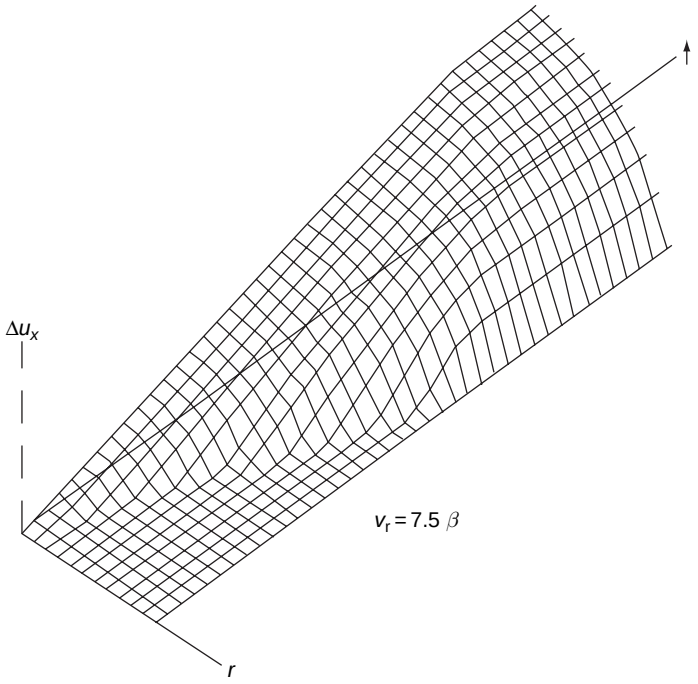


Figure 9 Slip distribution as a function of time and position for Madariaga's (1976) quasi-dynamic circular crack model.

Recent inversions of earthquake slip distributions using kinematic source models have found very complex source distributions that require an extensive reappraisal of classical source models that were mostly based on Kostrov's model of self-similar circular crack. Ruptures in a fault with a very heterogeneous load follow very tortuous paths. While on the average the rupture propagates at a subsonic speed from one end of the fault to another, in detail the rupture front can wander in all directions following the areas of strong stress concentration and avoiding those with low stress or high rupture resistance. If this view of earthquake rupture was to be confirmed by future observations (we believe it will be), then many current arguments about earthquake complexity, narrow rupture pulses, and earthquake distributions will be solved and we may concentrate on the truly interesting problem of determining which features of friction determine that fault stress is always complex under all circumstances.

Acknowledgments

This research was supported by CNRS (Centre National de la Recherche Scientifique) under several contracts from the PNRN program and by the European Union under the SPICE research network.

References

Achenbach JD (1975) *Wave Propagation in Elastic Solids*. Amsterdam: North-Holland.

Aki K (1966) Generation and propagation of G waves from the Niigata earthquake of June 16, 1964. Estimation of earthquake movement, released energy and stress–strain drop from G-wave spectrum. *Bulletin of the Earthquake Research Institute* 44: 23–88.

Aki K (1967) Scaling law of seismic spectrum. *Journal of Geophysical Research* 73: 5359–5376.

Aki K and Richards PG (2002) *Quantitative Seismology,* 2nd edn. Sausalito, California: University Science books.

Ampuero J-P and Dahlen FA (2005) Ambiguity of the Moment Tensor. *Bulletin of the Seismological Society of America* 95: 390–400.

Andrews J (1976) Rupture velocity of plane strain shear cracks. *Journal of Geophysical Research* 81: 5679–5687.

Archuleta R (1984) A faulting model for the 1979 Imperial Valley earthquake. *Journal of Geophysical Research* 89: 4559–4585.

Backus G and Mulcahy M (1976) Moment tensors and other phenomenological descriptions of seismic sources. I: Continuous displacements. *The Geophysical Journal of the Royal Astronomical Society* 46: 321–361.

Ben Menahem A (1961) Radiation of seismic surface waves from finite moving sources. *Bulletin of the Seismological Society of America* 51: 401–435.

Ben Menahem A (1962) Radiation of seismic body waves from finite moving sources. *Journal of Geophysical Research* 67: 345–350.

Ben Menahem A and Singh SJ (1981) *Seismic Waves and Sources*. New York: Springer.

Bernard P and Baumont D (2005) Shear Mach wave characterization for kinematic fault rupture models with constant supershear rupture velocity. *Geophysical Journal International* 162: 431–447.

Bernard P and Madariaga R (1984) A new asymptotic method for the modeling of near-field accelerograms. *Bulletin of the Seismological Society of America* 74: 539–557.

Boatwright J (1980) A spectral theory for circular sources: Simple estimates of source dimension, dynamic stress drop, and related seismic energy. *Bulletin of the Seismological Society of America*.

Bouchon M (1982) The rupture mechanism of the Coyote Lake earthquake of August 6, 1979 inferred from near field data. *Bulletin of the Seismological Society of America* 72: 745–759.

Brune J (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research* 75: 4997–5009.

Burridge R and Knopoff L (1964) Body force equivalents for seismic dislocations. *Bulletin of the Seismological Society of America* 54: 1875–1878.

Cohee B and Beroza G (1994) Slip distribution of the 1992 Landers earthquake and its implications for earthquake source mechanics. *Bulletin of the Seismological Society of America* 84: 692–712.

Cotton F and Campillo M (1995) Frequency domain inversion of strong motions: Application to the 1992 Landers earthquake. *Journal of Geophysical Research* 100: 3961–3975.

Dziewonski AM and Woodhouse JH (1983) Studies of the seismic source using normal-mode theory. In: Kanamori H and Boschi E (eds.) *Earthquakes: Observation, Theory, and Interpretation: Notes from the International School of Physics 'Enrico Fermi' (1982: Varenna, Italy)*, pp. 45–137. Amsterdam: North-Holland.

Eshelby JD (1956) The continuum theory of lattice defect. In: Seitz F and Turnbull D (eds.) *Progress in Solid State Physics*, vol 4. New York: Academic Press.

Eshelby JD (1969) The elastic field of a crack extending non-uniformly under general antiplane loading. *Journal of Mechanics and Physics of Solids* 17: 177–199.

Freund LB (1972) Crack propagation in an elastic solid subjected to general loading. II: Nonuniform rate of extension. *Journal of Mechanics and Physics of Solids* 20: 141–152.

Freund LB (1989) *Fracture Dynamics*. Cambridge: Cambridge University Press.

Fukuyama E and Madariaga R (1995) Integral equation method for plane crack with arbitrary shape in 3D elastic medium. *Bulletin of the Seismological Society of America* 85: 614–628.

Griffith AA (1920) The phenomenon of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London* A221: 163–198.

Haskell NA (1964) Total energy spectral density of elastic wave radiation from propagating faults. *Bulletin of the Seismological Society of America* 54: 1811–1841.

Haskell NA (1966) Total energy spectral density of elastic wave radiation from propagating faults, Part II. *Bulletin of the Seismological Society of America* 56: 1811–1841.

Hirose S and Achenbach JD (1989) Time-domain boundary element analysis of elastic wave interaction with a crack. *International Journal for Numerical Methods in Engineering* 28: 629–644.

Honda H (1962) Earthquake mechanism and seismic waves. *Journal of Physics of the Earth* 10: 1–98.

Kanamori H (1977) The energy release in great earthquakes. *Journal of Geophysical Research* 82: 2921–2987.

Keilis-Borok VI (1959) On the estimation of the displacement in an earthquake source and of source dimensions. *Annales de Geophysique* 12: 205–214.

Knopoff L and Randall MJ (1970) The compensated linear-vector dipole: A possible mechanism for deep earthquakes. *Journal of Geophysical Research* 75: 4957–4963.

Kostrov B (1964) Self-similar problems of propagation of shear cracks. *Journal of Applied Mathematics and Mechanics* 28: 1077–1087.

Kostrov B and Das S (1988) *Principles of Earthquake Source Mechanics*. Cambridge: Cambridge University Press.

Kostrov B and Nikitin L (1970) Some general problems of mechanics of brittle fracture. *Archiwum Mechaniki Stosowanej* 22: 749–776.

Madariaga R (1976) Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America* 66: 639–667.

Madariaga R (1978) The dynamic field of Kaskell's rectangular dislocation fault model. *Bulletin of the Seismological Society of America* 68: 869–887.

Madariaga R, Olsen KB, and Archuleta RJ (1998) Modeling dynamic rupture in a 3D earthquake fault model. *Bulletin of the Seismological Society of America* 88: 1182–1197.

Maruyama T (1963) On the force equivalents of dynamical elastic dislocations with reference to the earthquake mechanism. *Bulletin of the Earthquake Research Institute* 41: 467–486.

Mikumo T and Miyatake T (1979) Earthquake sequences on a frictional fault model with non-uniform strength and relaxation times. *Geophysical Journal of the Royal Astronomical Society* 59: 497–522.

Nakano H (1923) Notes on the nature of the forces which give rise to the earthquake motions. *Seismological Bulletin of Central METROLOGICAL Observatory of Japan* 1: 92–120.

Rice JR (1980) The mechanics of earthquake rupture. In: Dziewonski AM and Boschi E (eds.) *Physics of the Earth's Interior (Proceedings of the International School of Physics 'Enrico Fermi', Course 78, 1979)*, pp. 555–569. Amsterdam: North Holland Publ Co.

Richards PG (1976) Dynamic motions near an earthquake fault: A three dimensional solution. *Bulletin of the Seismological Society of America* 66: 1–32.

Rosakis AJ, Samudrala O, and Coker D (1999) Cracks faster than the shear wave speed. *Science* 284: 1337–1340.

Sato H and Hirasawa T (1973) Body wave spectra from propagating shear cracks. *Journal of the Physics of the Earth* 84: 829–841.

Savage JC (1966) Radiation from a realistic model of faulting. *Bulletin of the Seismological Society of America* 56: 577–592.

Spudich P and Hartzell LN (1984) Use of ray theory to calculate high-frequency radiation from earthquake sources having spatially variable rupture velocity and stress drop. *Bulletin of the Seismological Society of America* 74: 2061–2082.

Virieux J (1986) P–SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics* 51: 889–901.

Wald D and Heaton T (1994) Spatial and temporal distribution of slip for the 1992 Landers, California earthquake. *Bulletin of the Seismological Society of America* 84: 668–691.

4.03 Fracture and Frictional Mechanics – Theory

Y. Fialko, University of California San Diego, La Jolla, CA, USA

Nomenclature

A	rupture area
c	specific heat
$\bar{D}$	average fault slip
D_c	critical weakening displacement
D_m	maximum coseismic displacement
$D(x)$	fault slip
f_i	body force
$f_{ij}(\vartheta)$	azimuthal dependence of a near-tip stress field
F	half-length of a developed part of a crack on which friction has dropped to a residual level σ_d
G	release rate of mechanical energy (per unit crack length)
G_c	effective fracture energy (per unit crack length)
K	stress intensity factor
K_c	critical stress intensity factor, fracture toughness
L	crack half-length, or full length of a self-healing pulse
L^*	dynamic pulse length
L_c	critical crack length
M_0	scalar seismic moment
Q	rate of heat generation
r	distance to the crack tip
r_0	effective radius of the crack tip
R	length of the process zone at the crack tip
R^*	dynamic length of the process zone at the crack tip
S	minor axis of an elliptical cavity
S_f	boundary between elastically and inelastically deforming material
t	time
T	temperature
$\hat{T}$	temperature scale
u_i	displacement vector
U_e	elastic strain energy
U_f	frictional losses
U_G	fracture energy
U_l	potential energy of remotely applied stresses
U_r	radiated energy
V_l	limiting rupture velocity
V_p	P wave velocity
V_r	rupture velocity
V_s	S wave velocity
w	half-width of the fault slip zone
$\bar{w}$	nondimensional width of the fault slip zone
x, y, z (x_1, x_2, x_3)	spatial coordinates
α	nondimensional factor
γ	nondimensional factor
δ_{ij}	Kronecker delta function
δW	work done by external forces
ΔS_1	surface bounding a prospective increment of the process zone at the crack tip in the result of crack growth
ΔS_2	surface bounding a prospective decrement of the process zone at the crack tip in the result of crack growth

ΔU_p	change in the potential energy
$\Delta\sigma$	stress drop
ε	small distance
ε_{ij}	strain tensor
ζ	complex variable
θ	nondimensional temperature
κ	thermal diffusivity
μ	shear modulus
ν	Poisson's ratio
ξ	dummy variable
ρ	density
σ_0	remotely applied stress
σ_d	residual stress on the developed part of a crack
σ_{ij}	stress tensor
σ_{ij}^0	stress tensor prior to crack propagation
σ_{ij}^1	stress tensor after crack propagation
σ_s	static strength, or the yield stress in the crack tip process zone
ϕ	analytic function of a complex argument
χ	nondimensional along-crack coordinate
ψ	nondimensional half-length of the developed part of a crack
$\Re$	real part of a complex argument
$\Im$	imaginary part of a complex argument

4.03.1 Introduction

Seismic and geodetic observations indicate that most earthquakes are a result of unstable localized shear on quasi-planar faults (Gutenberg and Richter, 1949; Dahlen and Tromp, 1998). Because the thickness of earthquake rupture zones that accommodate slip is much smaller than their characteristic in-plane dimensions, it is natural to idealize earthquake ruptures as shear cracks. Development, propagation, and arrest of shear cracks are subject of the earthquake fracture mechanics. Unlike the engineering fracture mechanics that mainly concerns itself with criteria and details of failure at the tip of tensile cracks propagating through 'intact' solids (Lawn, 1993; Freund, 1998), the earthquake fracture mechanics must consider both the inelastic yielding at the rupture fronts, as well as the evolution of friction (in general, rate and slip dependent) on the rest of the slipping surface (*see* Chapters 4.04 and 4.05). Although a distinction is sometimes made between the crack models and friction models of an earthquake source (e.g., Kanamori and Brodsky, 2004), the two processes are intrinsically coupled and should be considered jointly. Note that the shear crack propagation does not necessarily imply creation of a new fault in intact rocks, but also refers to slip on a pre-existing (e.g., previously ruptured) interface. Mathematically, the crack growth in unbroken media and on pre-existing faults is very similar, provided that the slip surface is planar. While shear cracks in unconfined intact media tend to propagate out of their initial planes, such tendency is suppressed at high confining pressures (e.g., Melin, 1986; Broberg, 1987; Lockner *et al.*, 1992), and below we limit our attention to planar ruptures.

The redistribution of stress and strain due to a spatially heterogeneous fault slip can be described using either kinematic (displacement-controlled boundary conditions) or dynamic (stress-controlled boundary conditions) approach. Kinematic (e.g., dislocation) models are useful if the fault slip is known or can be inferred with sufficient accuracy, for instance, from seismologic, geodetic, or geologic observations (Steketee, 1958; Vvedenskaya, 1959; Bilby and Eshelby, 1968; Okada, 1985; Savage, 1998). Dynamic (fracture mechanics) models potentially have a greater predictive power, as they solve for, rather than stipulate, the details of fault slip for given initial stress conditions (Burridge and Halliday, 1971; Andrews, 1976a; Madariaga, 1976; Freund, 1979; Day, 1982; Ben-Zion and Rice, 1997). The time-dependent boundary conditions on the fault are usually deduced by using constitutive laws that relate kinetic friction to the magnitude and velocity of fault slip, preseismic stress, temperature, and other state variables pertinent to the evolution of stress on a slipping interface. The kinematic and dynamic approaches give rise to identical results for the same slip distribution. Dislocation models are well understood, and are widely employed in inversions of seismic and geodetic data for the rupture geometry and spatiotemporal details of slip (e.g., Hartzell and Heaton, 1983; Delouis *et al.*, 2002;

Fialko *et al.*, 2005). Fracture mechanics models are intrinsically more complex and less constrained, but nonetheless increasingly used to interpret high-quality near-field seismic data (Peyrat *et al.*, 2001; Oglesby and Archuleta, 2001; Aochi and Madariaga, 2003). This chapter focuses on the fracture mechanics approach, and kinematic models are be discussed. Consequently, the term 'dynamic' will be used to describe time-dependent aspects of rupture for which inertial effects are important, that is, in a meaning opposite to 'static' (rather than 'kinematic') descriptions. There exist a number of excellent texts covering the fundamentals of fracture mechanics, with applications to the earthquake source seismology, including Cherepanov (1979), Rice (1980), Rudnicki (1980), Li (1987), Segall (1991), Scholz (2002), and Ben-Zion (2003), among others. This chapter reviews the essential aspects of fracture mechanics, with an emphasis on theoretical developments made over the last decade.

4.03.2 Linear Elastic Fracture Mechanics

It is well known that structural flaws and discontinuities such as cracks, voids, and inclusions of dissimilar materials give rise to local stress perturbations that may significantly amplify the average applied stress. A classic example is an elliptical cavity in an elastic plate subject to a uniform extension (Inglis, 1913; Lawn, 1993). Provided that the major and minor axes of the cavity are L and S, respectively, the orientation of the remote tensile stress σ_0 is orthogonal to the major axis, and the cavity walls are stress free, the component of stress parallel to the remote tension assumes a value of $\sigma_0(1+2L/S)=\sigma_0\left(1+2\sqrt{L/r}\right)$ at the cavity tip, where $r=S^2/L$ is the radius of the cavity tip. For an extreme case of a sharp slit, $S/L\to 0$, the stress at the tip scales as $\sigma_0\sqrt{L}/\sqrt{r}$, that is, becomes unbounded for however small remote loading σ_0. Full analytic solutions for the stress distribution around sharp cracks (see Section 4.03.3; also, Rice, 1968a; Freund, 1998; Lawn, 1993) indicate that the stress field indeed has a characteristic square root singularity,

$$\sigma_{ij}|_{r\to 0}\approx\frac{K}{\sqrt{2\pi r}}f_{ij}(\vartheta) \qquad [1]$$

where $K\sim O\left(\Delta\sigma\sqrt{L}\right)$ is the stress intensity factor that depends on the crack geometry and loading configuration, $\Delta\sigma$ being the difference between the far-field stress and stress resolved on the crack walls (hereafter referred to as the stress drop; for an empty crack under remote tension, $\Delta\sigma=\sigma_0$), r is now the distance to the crack tip measured from the crack exterior, and $f_{ij}(\vartheta)$ is a function characterizing the azimuthal dependence of the near-tip stress field (Rice, 1980; Lawn, 1993). Because the governing equations and mathematical structure of solutions are identical for the tensile (mode I), in-plane shear (mode II), and anti-plane shear (Mode III) cracks (see Section 4.03.4), we will use examples of both tensile and shear cracks to highlight universal features of, as well as important differences between the shear and tensile rupture modes. Such a comparison is instructive because many concepts of fracture mechanics have been developed for tensile failure that is most common in engineering applications, and subsequently borrowed for seismologic applications that mostly deal with shear failure. Significant differences in the ambient conditions warrant a careful evaluation of the range of applicabilities of basic assumptions behind the fracture mechanics models (e.g., Rubin, 1993).

4.03.2.1 Singular Crack Models

The square root singularity in the stress field (eqn [1]) is a common feature of all crack models assuming a constant stress drop, and elastic deformation of the ambient solid, regardless of the mode of failure, and rupture velocity (as long as the latter remains subsonic; see Section 4.03.5). While this stress singularity is admissible on thermodynamic grounds (in particular, it is integrable, so that the elastic strain energy is finite in any closed volume containing the crack tip), it is clearly unrealistic as no material is able to support infinite stresses. Theoretical arguments supported by a large number of observations suggest that the assumption of a perfect brittle behavior (i.e., elastic deformation of the unbroken host up to the onset of fracture) is violated in some zone around the crack tip (Irwin, 1957; Atkinson, 1987; Lawn, 1993; **Figure 1**).

In this zone (commonly referred to as the process, or breakdown zone), inelastic yielding prevents stress increases in excess of a certain peak value σ_s that presumably represents the microscopic strength of a material. The size of the process zone r_0 that corresponds to equilibrium (i.e., a crack on a verge of propagating) may be interpreted as the effective radius of the crack tip. Outside of the process zone

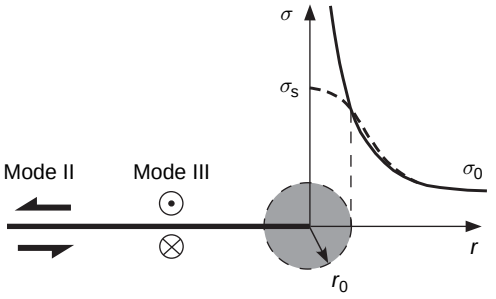


Figure 1 An idealized view of stress variations and yielding at the crack tip. r_0 is the characteristic dimension of an inelastic zone in which stresses exceed the yield threshold σ_s. Solid curve shows the theoretically predicted stress increase with a characteristic $1/\sqrt{r}$ singularity. The theoretical prediction breaks down at distances $r < r_0$, but may be adequate for $r > r_0$.

(at distances $r > r_0$) stresses are below the failure envelope, and the material deforms elastically. If the size of the equilibrium process zone is negligible compared to the crack length, as well as any other characteristic dimensions (e.g., those of an encompassing body), a condition termed small-scale yielding (Rice, 1968a) is achieved, such that the stress and strain fields in the intact material are not appreciably different from those predicted in the absence of a process zone. This is the realm of the linear elastic fracture mechanics (LEFM). According to LEFM, the near-tip ($r_0 < r \ll L$) stress field is completely specified by the scalar multiplier on the singular stress field, the stress intensity factor K (eqn [1]), which can be found by solving an elastic problem for a prescribed crack geometry and loading conditions. The crack propagation occurs when the stress intensity factor exceeds a critical value, $K > K_c$. The critical stress intensity factor, or the fracture toughness K_c, is believed to be a material property, independent of the crack length and loading configuration (although fracture properties may vary for different modes of failure, e.g., K_{Ic}, K_{IIc}, etc., similar to differences between macroscopic tensile and shear strengths). To the extent that the microscopic yield strength σ_s, and the effective equilibrium curvature of the crack tip, r_0, may be deemed physical properties, the fracture toughness K_c may be interpreted as a product $\sigma_s\sqrt{r_0} = \text{const}$.

The critical stress intensity factor is a local fracture criterion, as it quantifies the magnitude of the near-tip stress field on the verge of failure. However, it can be readily related to global parameters characterizing changes in the elastic strain energy ∂U_e and potential energy of applied stresses ∂U_ℓ in the entire body due to a crack extension ∂L, such as the energy release rate $G_c = -(\partial U_e + \partial U_\ell)/\partial L$,

$$G_c = \frac{\alpha K_c^2}{2\mu} \qquad [2]$$

where μ is the shear modulus of an intact material, and α assumes values of $(1-\nu)$ and unity for mode II and III loading, respectively (Irwin, 1957). For ideally brittle materials, the energy release rate may be in turn associated with the specific surface energy spent on breaking the intermolecular bonds (Griffith, 1920; Lawn, 1993). Further analysis of the breakdown process at the crack tip requires explicit consideration of the details of stress concentration in the tip region.

4.03.2.2 Planar Breakdown Zone Models

A simple yet powerful extension of the LEFM formulation is the displacement-weakening model which postulates that (1) the breakdown process is confined to the crack plane, (2) inelastic deformation begins when stresses at the crack tip reach some critical level σ_s, and (3) yielding is complete when the crack wall displacement exceeds some critical value D_c (Leonov and Panasyuk, 1959; Barenblatt, 1959; Dugdale, 1960). In case of tensile (model I) cracks, σ_s represents the local tensile strength in the breakdown zone, and D_c is the critical opening displacement beyond which there is no cohesion between the crack walls. In case of shear (model II and III) cracks (**Figure 2**).

σ_s represents either the shear strength (for ruptures propagating through an intact solid), or the peak static friction (for ruptures propagating along a pre-existing fault), and D_c is the slip-weakening distance at which a transition to the kinetic friction is complete (Ida, 1972; Palmer and Rice, 1973). Under these assumptions, the fracture energy may be defined as work required to evolve stresses acting on the crack plane from σ_s to the residual value σ_d, and is of the order of $(\sigma_s - \sigma_d)D_c$, depending on the details of the displacement-weakening relation (**Figure 3**).

For an ideal brittle fracture that involves severing of intermolecular bonds ahead of the crack tip, σ_s may approach theoretical strength, $\mu/10 \sim O(10^9 - 10^{10}\,\text{Pa})$, and D_c may be of the order of the crystal lattice spacing $(10^{-10} - 10^{-9}\,\text{m})$, yielding $G_c \sim O(1\,\text{J m}^{-2})$. This is close to the experimentally measured fracture energies of highly brittle crystals and glasses (Griffith, 1920; Lawn, 1993). At the same time, laboratory measurements of polycrystalline aggregates (such as rocks, ceramics, and metals) reveal much higher fracture energies ranging from $10–10^2\,\text{J m}^{-2}$ (for tensile failure)

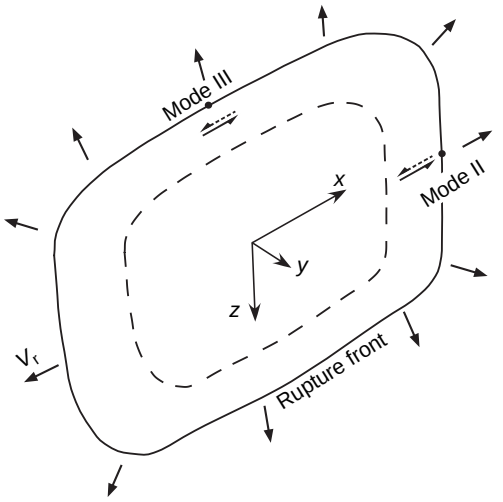


Figure 2 Schematic view of a shear rupture expanding at a constant velocity V_r. Conditions corresponding to the two-dimensional mode II and III loading are approximately satisfied at the rupture fronts orthogonal and parallel to the local slip vector, respectively. For a crack-like rupture, slip occurs on the entire area bounded by the rupture front. For a pulse-like rupture, slip is confined to an area between the rupture front (solid line) and a trailing healing front (dashed line).

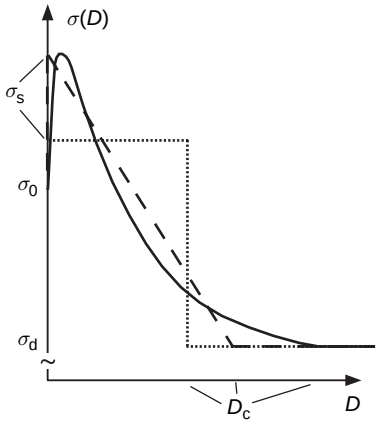


Figure 3 Possible dependence of the breakdown stress σ on the crack wall displacement D within the crack tip process zone. Solid line: a schematic representation of the experimentally measured displacement-weakening relations for rocks (Hashida *et al.*, 1993; Li, 1987). Dashed and dotted lines: approximations of the displacement-weakening relation assuming no dependence of the yield stress on D (dotted line), and linear weakening (dashed line). The area beneath all curves is the same, and equals to the fracture energy $G_c = \alpha\sigma_s D_c$, where α is a numerical factor that depends on the particular form of the displacement-weakening relationship. For example, $\alpha = 1$ for the constant yield stress model, and $\alpha = 0.5$ for the linear weakening model.

to $10^3\,\mathrm{J\,m^{-2}}$ (for shear failure), presumably reflecting dependence of the effective D_c on the material microstructure (e.g., texture, grain size, and distribution of inhomogeneities), and the breakdown mechanism (e.g., grain unlocking, ligamentary bridging, plasticity, off-plane yielding, or some other deviation from the ideally brittle behavior) (Li, 1987; Hashida *et al.*, 1993; Fialko and Rubin, 1997). Indirect inferences of fracture energies associated with *in situ* propagation of tensile cracks such as man-made hydrofractures and magmatic dikes (Rubin, 1993; Sandwell and Fialko, 2004) and earthquake ruptures (Husseini, 1977; Wilson *et al.*, 2004; Chester *et al.*, 2005; Abercrombie and Rice, 2005) reveal significantly larger fracture energies compared to those measured in the laboratory experiments. This disagreement may be indicative of some scaling of fracture energy with the rupture size or rupture history, and perhaps the significance of dynamic effects on cracking and damage that are not captured in laboratory tests. In the context of an in-plane breakdown zone model, the inferred scale dependence of fracture energy is usually interpreted in terms of increases in the effective weakening displacement D_c with the rupture length L. Possible mechanisms of scale dependence of D_c include continued degradation of the dynamic fault strength with slip (e.g., due to thermal or rheologic weakening) (Lachenbruch, 1980; Brodsky and Kanamori, 2001; Di Toro *et al.*, 2004; Abercrombie and Rice, 2005), or fractal nature of the fault roughness (e.g., Barton, 1971; Power and Tullis, 1995). It should be pointed out that the common interpretation of the scale dependence of G_c in terms of the scale dependence of D_c may be too simplistic, as the assumption of a thin process zone unlikely holds under *in situ* stress conditions for neither tensile nor shear cracks, as suggested by both theoretical arguments (Rubin, 1993; Fialko and Rubin, 1997; Poliakov *et al.*, 2002; Rice *et al.*, 2005) and observations (Manighetti *et al.*, 2001; Fialko *et al.*, 2002; Fialko, 2004b; Wilson *et al.*, 2004; Chester *et al.*, 2005). A more detailed quantitative treatment of the energetics of fracture and breakdown processes is presented in Section 4.03.4.

Nonsingular crack models postulating a thin in-plane process zone are mathematically appealing because they allow one to treat the fracture problem as an elastic one by removing nonlinearities associated with failure from the governing equations, and incorporating them into boundary conditions on a crack plane. For example, elastic solutions can be readily used to determine the size of the process zone at the crack tip. Dimensional arguments indicate that the strain associated with the material

breakdown is of the order of D_c/R, where R is the characteristic size of the breakdown zone. This strain must be of the order of the elastic strain associated with the strength drop, $(\sigma_s - \sigma_d)/\mu$, implying

$$R = \gamma \frac{\mu}{\sigma_s - \sigma_d} D_c \qquad [3]$$

where γ is a nondimensional factor of the order of unity that depends on the displacement-weakening relation (e.g., as shown in **Figure 3**). The small-scale yielding condition requires $R \ll L$. Parallels may be drawn between the size of the breakdown zone R in the displacement-weakening model, and the critical radius of the crack tip r_0 in the LEFM model (**Figure 1**; also, see Khazan and Fialko, 1995), in that both parameters demarcate the near-tip region that undergoes failure from the rest of the material in which the deformation is essentially elastic. On a larger scale, the assumption of elasticity postulates a direct proportionality between the stress drop $\Delta\sigma$ and the average fault slip $\bar{D}$,

$$\bar{D} \sim \frac{\Delta\sigma}{\mu} L \qquad [4]$$

where L is the characteristic fault dimension (for isometric ruptures), or the least dimension of the slip area (in case of ruptures having irregular shape or high aspect ratios). The assumption of a predominantly elastic behavior of the Earth's crust appears to be in a good agreement with seismic (Gutenberg and Richter, 1949; Vvedenskaya, 1959; Aki and Richards, 1980) and geodetic (Reid, 1910; Fialko, 2004b) observations of the instantaneous response of crustal rocks to major earthquakes. The brittle–elastic model gives rise to several fundamental scaling relationships used in earthquake seismology. For example, the scalar seismic moment M_0 is a measure of the earthquake size,

$$M_0 = \bar{D} A \mu \qquad [5]$$

where A is the rupture area (Aki, 1967; Kostrov, 1974). The seismic moment is related to a coseismic change in the potential energy of elastic deformation ΔU_e (e.g., Kanamori, 2004),

$$\Delta U_e = \frac{\Delta\sigma}{2\mu} M_0 \qquad [6]$$

For isometric ruptures, $A \sim L^2$, and by combining eqns [4] and [5] one obtains a well-known scaling between the scalar seismic moment, the rupture size, and the stress drop (e.g., Kanamori and Anderson, 1975; Dahlen and Tromp, 1998),

$$M_0 \sim \Delta\sigma L^3 \qquad [7]$$

4.03.3 The Governing Equations

Consider a three-dimensional medium whose points are uniquely characterized by Cartesian coordinates x_i $(i = 1, 2, 3)$ in some reference state prior to the fault-induced deformation. Fault slip gives rise to a displacement field u_i. The displacements u_i are in general continuous (differentiable), except across the slipped part of the fault. The strain tensor is related to the displacement gradients as follows,

$$\varepsilon_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) \qquad [8]$$

where the comma operator as usual denotes differentiation with respect to spatial coordinates, $a_{,i} = \partial a/\partial x_i$. Equation [8] assumes that strains are small ($\epsilon_{ij} \ll 1$), so that terms that are quadratic in displacement gradients can be safely neglected (e.g., Malvern, 1969; Landau and Lifshitz, 1986). The assumption of an infinitesimal strain implies no difference between the material (Lagrangian) and spatial (Eulerian) reference frames. Typical strains associated with earthquake ruptures are of the order of 10^{-4}–10^{-5} (Kanamori and Anderson, 1975; Scholz, 2002) so that the infinitesimal strain approximation is likely justified. For sufficiently small strain changes (with respect to some reference state), laboratory data and theoretical arguments suggest a linear dependence of strain perturbation on the causal stress change. For an isotropic homogeneous elastic material, the corresponding relationship between stresses and strains is given by the Hooke's law (e.g., Timoshenko and Goodier, 1970; Landau and Lifshitz, 1986):

$$\varepsilon_{ij} = \frac{1}{2(1+\nu)\mu}\left[(1+\nu)\sigma_{ij} - \nu\delta_{ij}\sigma_{kk}\right] \qquad [9]$$

where μ and ν are the shear modulus and the Poisson's ratio, respectively, δ_{ij} is the Kronecker delta function, and repeating indexes imply summation.

Conservation of a linear momentum in continuous media gives rise to the Navier–Cauchy equations of equilibrium (Malvern, 1969):

$$\sigma_{ij,j} + f_i = \frac{\partial^2 u_i}{\partial t^2} \qquad [10]$$

where f_i is the body force (e.g., due to gravity), and t is time. Due to the linearity of equilibrium eqn [10], it is possible to represent the full stress tensor as a superposition of some background (e.g., lithostatic, regional tectonic) stress, and a perturbation due to fault slip, such that the latter satisfies a homogeneous

case of eqn [10]. Unless noted otherwise, we assume below that the stress tensor σ_{ij} denotes only stress perturbations due to fault displacements. For convenience, the indicial nomenclature for spatial coordinates (x_1, x_2, x_3) is used interchangeably with the traditional component notation (x, y, z) throughout the rest of the text.

In order to close the problem formulation, some constitutive law relating slip D, slip velocity $\partial D/\partial t$, resolved shear and normal stresses, σ_τ and σ_n, pore fluid pressure p, temperature T, etc., needs to be prescribed on surfaces that violate assumptions of continuity, for example, faults and cracks. Examples are the Mohr–Coulomb (Byerlee, 1978) and rate-and-state friction (*see* Chapter 4.04), flash melting (Molinari *et al.*, 1999; Rice, 2006), thermal pressurization (Sibson, 1973; Segall and Rice, 1995; Andrews, 2002), viscous rheology (Fialko and Khazan, 2005), etc.

4.03.4 Exact Solutions for Quasistatic Two-Dimensional Planar Cracks

In case of two-dimensional deformation (e.g., plane or antiplane strain, or plane stress), elastic solutions for stresses and displacements can be generally expressed in terms of two analytic functions of a complex variable $\zeta = x + iy$ (Kolosoff, 1909; Muskhelishvili, 1953). For simplicity, here we consider loading that is symmetric about the center of the crack $x = 0$. In this case, both stresses σ_{ij} and displacements u_i can be expressed through a single analytic function $\phi(\zeta)$ (Westergaard, 1939). First, we demonstrate that the mathematical structure of solutions for stresses and displacements in the crack plane ($y = 0$) is identical for tensile (mode I), in-plane shear (mode II), and antiplane shear (mode III) cracks. In particular, for tensile (mode I) cracks, stresses and displacements can be found from $\phi(\zeta)$ as (e.g., Muskhelishvili, 1953; Khazan and Fialko, 2001)

$$\sigma_{yy} = 2(\Re\phi' + y\Im\phi'') \quad [11]$$

$$\sigma_{xy} = -2y\Re\phi'' \quad [12]$$

$$u_y = \frac{1}{\mu}(2(1-\nu)\Im\phi - y\Re\phi') \quad [13]$$

where $\Re$ and $\Im$ denote the real and imaginary parts of a complex argument. For plane strain shear (mode II) cracks, conditions of symmetry imply that $\sigma_{yy} = 0$ on the crack plane ($y = 0$), and the corresponding equilibrium equations are

$$\sigma_{xy} = -2(\Im\phi' + y\Re\phi'') \quad [14]$$

$$u_x = \frac{1}{\mu}(2(1-\nu)\Re\phi + \Im\phi) \quad [15]$$

Upon making a substitution $\phi = -i\phi_1$, one can see that the unknown shear stress σ_{xy} and displacement u_x on the crack plane may be expressed through a new function ϕ_1 in the same manner as the normal stress σ_{yy} and displacement u_y are expressed through ϕ eqns [11] and [13]):

$$\sigma_{xy} = 2\left(\Re\phi_1' - y\Im\phi_1''\right) \quad [16]$$

$$u_x = \frac{1}{\mu}\left(2(1-\nu)\Im\phi_1 - y\Re\phi_1'\right) \quad [17]$$

For antiplane shear (mode III) cracks, expressions for the relevant stress and displacement components are

$$\sigma_{yz} = -\Im\phi' \quad [18]$$

$$u_z = \frac{1}{\mu}\Re\phi \quad [19]$$

Upon making a substitution $\phi = -2i\phi_2$, one can see that the dependence of the unknown quantities σ_{yz} and $(1-\nu)u_z$ on the analytic function ϕ_2 is analogous to the dependence of σ_{xy} and u_x on ϕ_1 obtained for the in-plane shear crack for $y = 0$ (eqns [16] and [17]). This analogy mandates that the mathematical structure of solutions for the tensile, in-plane, and antiplane components of stress for the corresponding crack modes is identical, provided that the boundary conditions on the crack plane (the along-crack distribution of the driving stress, the displacement-weakening relationship, etc.) are analogous. Solutions for the crack wall displacements are also identical for different modes, although expressions for displacements for mode III cracks will differ from those for mode I and II cracks by a factor of $(1-\nu)$. Hence we focus on a particular case of an in-plane shear (mode II) crack. The boundary condition for the potential function ϕ_1 are as follows:

$$\Re\phi_1' = \sigma(x)/2 \text{ for } |x| < L \quad [20]$$

$$\Im\phi_1' = 0 \text{ for } |x| > L \quad [21]$$

$$\phi_1' \to \sigma_0/2 \text{ for } |\zeta| \to \infty \quad [22]$$

where $\sigma(x)$ is the distribution of shear stress on the crack surface, and σ_0 is the applied shear stress at infinity (hereafter referred to as prestress). The boundary condition [21] postulates no slip beyond the

rupture front (see eqn [17]). An explicit solution for the function ϕ_1 that is analytic in the upper halfplane and satisfies the boundary conditions [20]–[22] is given by the Keldysh–Sedov formula (Muskhelishvili, 1953; Gakhov, 1966; Khazan and Fialko, 1995, 2001):

$$\phi_1'(\zeta) = \frac{1}{2\pi i}\left(\frac{\zeta+L}{\zeta-L}\right)^{1/2}\int_{-L}^{L} d\xi \left(\frac{\xi-L}{\xi+L}\right)^{1/2}\frac{\sigma(\xi)}{\xi-\zeta} + \frac{\sigma_0}{2}\left(\frac{\zeta+L}{\zeta-L}\right)^{1/2} + \frac{C}{(\zeta^2-L^2)^{1/2}} \quad [23]$$

where C is an arbitrary constant. From eqn [16], an asymptotic behavior of ϕ_1' at infinity is

$$\lim_{|\zeta|\to\infty}\phi_1'(\zeta) = -\frac{1}{2\pi|\zeta|}\int_{-L}^{L} d\xi\left(\frac{\xi-L}{\xi+L}\right)^{1/2}\sigma(\xi) + \frac{\sigma_0}{2} + \frac{L\sigma_0 + C}{|\zeta|} \quad [24]$$

One can readily determine the unknown constant C from eqn [16] by satisfying the boundary condition [22]. The final expression for the derivative of analytic function ϕ_1 is

$$\phi_1'(\zeta) = \frac{1}{2\pi i}\left(\frac{\zeta+L}{\zeta-L}\right)^{1/2}\int_{-L}^{L} d\xi \left(\frac{\xi-L}{\xi+L}\right)^{1/2}\frac{\sigma(\xi)}{\xi-\zeta} + \frac{\sigma_0}{2}\left(\frac{\zeta+L}{\zeta-L}\right)^{1/2}\frac{1}{2\pi(\zeta^2-L^2)^{1/2}}\int_{-L}^{L} d\xi\frac{\sigma(\xi)-\sigma_0}{(L^2-\xi^2)^{1/2}} \quad [25]$$

For any physically admissible failure model, the maximum stress within the slipped region is bounded by the yield strength, $\sigma_{xy} < \sigma_s$ for $|x| < L$. Furthermore, it is reasonable to expect that $\sigma_{xy} \to \sigma_s$ as $|x| \to L$. The shear stress $\sigma_{xy}(\zeta)$ in the 'locked' region ahead of the rupture front is then given by the real part of ϕ_1 (see eqn [16]). Sufficiently close to the crack tip (i.e., for $\zeta = L + \epsilon$, such that $\Im\epsilon = 0; 0 < \epsilon \ll L$), we obtain

$$\sigma_{xy}(L+\epsilon) = \frac{1}{\pi}\left(\frac{L}{2\epsilon}\right)^{1/2}\int_{-L}^{L} d\xi\frac{\sigma(\xi)-\sigma_0}{(L^2-\xi^2)^{1/2}} + \sigma_s + O\left(\epsilon^{1/2}\right) \quad [26]$$

The first term on the right-hand side of eqn [26] is of the order of $1/\sqrt{\epsilon}$ and represents the LEFM approximation (see Section 4.03.2). By comparing eqns [26] and [1], one can formally introduce the stress intensity factor:

$$K = \frac{1}{\pi}\left(\frac{L}{2\epsilon}\right)^{1/2}\int_{-L}^{L} d\xi\frac{\sigma(\xi)-\sigma_0}{(L^2-\xi^2)^{1/2}} \quad [27]$$

which exhibits the expected scaling $K \propto \Delta\sigma\sqrt{L}$. Physically plausible crack models require that stresses are finite everywhere. From eqns [26] and [27], the requirement of the absence of a stress singularity is met if (and only if)

$$\int_{-L}^{L} d\xi\frac{\sigma(\xi)-\sigma_0}{(L^2-\xi^2)^{1/2}} = 0 \quad [28]$$

Any realistic distribution of the driving stress resolved on the crack surface must satisfy the integral constraint [28]. This implies that the driving stress $(\sigma(x) - \sigma_0)$ must change sign along the crack ($|x| < L$). For example, the stress drop $(\sigma(x) < \sigma_0)$ in the central part of the crack needs to be balanced by the material strength or the high transient friction $(\sigma_s \geq \sigma(x) > \sigma_0)$ within the process zone.

Displacements of the crack walls $u_x(x)$ corresponding to the instantaneous shear stress $\sigma(x)$ can be found by differentiating eqn [17] for $y = 0$,

$$\frac{du_x}{dx} = \frac{2(1-\nu)}{\mu}\Im\phi_1' \quad [29]$$

and making use of expression [26] to integrate the resulting differential eqn [29] with the initial condition $u_x(-L) = 0$. The respective solution is

$$D(x) = \frac{2(1-\nu)}{\pi\mu}\int_{x}^{L}(L^2-\xi^2)^{1/2}\,d\xi\int_{-L}^{L}\frac{\sigma(t)dt}{(L^2-t^2)^{1/2}(t-\xi)} \quad [30]$$

where $D(x) = 2u_x(x)$ is slip between the crack walls. The problem is closed by specifying a constitutive relationship between slip D and kinetic friction σ. In general, such a relationship may include dependence of friction of slip rate, as well as on the amount of slip, temperature, and other state variables (Dieterich, 1979; Ruina, 1983; Blanpied *et al.*, 1995). To gain a further analytic insight, here we consider a simple slip-weakening relationship

$$\sigma(x) = \sigma_s \text{ for } D(x) < D_c$$
$$\sigma(x) = \sigma_d \text{ for } D(x) > D_c \quad [31]$$

where σ_s is the yield strength, or static friction in the process zone, σ_d is the residual kinetic friction on the developed part of the crack, and D_c is the critical slip-weakening displacement corresponding to a transition from σ_s to σ_d (see **Figure 3**). The size of the process zone, R, is defined by a requirement that the fault slip is subcritical within the process zone, $D(|F|) = D_c$, where $F = L - R$ is the length of the developed part of a crack on which the friction has dropped to a residual value σ_d. **Figure 4** illustrates the geometry of the problem.

In case of a piecewise constant distribution of shear stress along the crack $\sigma(x)$ given by eqn [32], as shown in **Figure 4**, expressions [28] and [30] can

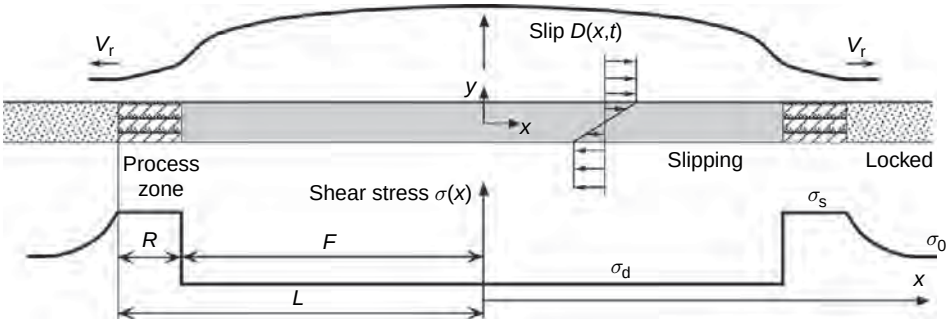


Figure 4 A two-dimensional plain strain shear crack in an infinite elastic medium. The imposed shear stress at the infinity is σ_0. The developed part of the crack is assumed to have a constant residual shear stress σ_d. At the crack tips, there are process zones having length R and shear stress σ_s.

be readily integrated to provide closed-form analytic solutions. In particular, integration of expression [28] allows one to determine a relative size of the process zone with respect to the crack half-length:

$$\frac{R}{L} = 2\sin^2\left(\frac{\pi}{4}\frac{\sigma_0-\sigma_d}{\sigma_s-\sigma_d}\right) \qquad [32]$$

Evaluation of integral [30] gives rise to

$$D(x) = \frac{2(1-\nu)L}{\pi\mu}(\sigma_0-\sigma_d)I(\chi,\psi) \qquad [33]$$

where $\chi = x/L$ and $\psi = F/L$ are the nondimensional along-crack coordinate and the half-length of the developed part of the crack, respectively, and function I is given by the following equation:

$$I(U,V) = (V+U)\log\left|\frac{1+UV+\sqrt{(1-U^2)(1-V^2)}}{V+U}\right| + (V-U)\log\left|\frac{1-UV+\sqrt{(1-U^2)(1-V^2)}}{V-U}\right| \qquad [34]$$

At the base of the process zone, $\chi = \psi$, eqn [33] reduces to

$$D(F) = \frac{4(1-\nu)}{\pi\mu}F(\sigma_s-\sigma_d)\log\frac{1}{\psi} \qquad [35]$$

For a crack that is on the verge of propagating, slip at the base of the process zone ($x = F$) equals the critical weakening displacement D_c. A simple dimensional analysis of eqn [35] then reveals a characteristic length scale L_c:

$$L_c = \frac{\pi D_c}{4(1-\nu)}\frac{\mu}{\sigma_s-\sigma_d} \qquad [36]$$

Using eqn [36], eqn [35] may be written as

$$\frac{L_c}{F} = \log\frac{L}{F} \qquad [37]$$

A further rearrangement of eqn [37] yields an explicit expression for the length of the process zone R:

$$R = F\left(\exp\left(\frac{L_c}{F}\right)-1\right) \qquad [38]$$

It is instructive to consider two end-member cases: the long crack limit, $F \gg L_c$, and the microcrack limit, $F \ll L_c$. In case of sufficiently long cracks ($F \gg L_c$), an asymptotic expansion of eqn [38] (i.e., $\exp(x) \approx 1 + x$ for $x \ll 1$) indicates that the process zone length is independent of the crack length, and equals L_c. This is the realm of small-scale yielding, $R = \text{const.} = L_c \ll L$, $F \approx L$, in which cracks propagate by preserving the structure of the near-tip stress field. By comparing eqns [37] and [3], one can see that the general scaling relationship suggested by dimensional arguments holds; for the case of a constant breakdown stress σ_s within the process zone, the nondimensional coefficient γ in estimate [3] equals $\pi/4(1-\nu)$. Exact analytic solutions assuming small-scale yielding give rise to $\gamma = \pi/2(1-\nu)$ for a linear slip-weakening relationship (see **Figure 3**; e.g., Chen and Knopoff, 1986), and $\gamma = 9\pi/16(1-\nu)$ for a breakdown stress that linearly decreases with distance away from the crack tip (Palmer and Rice, 1973). In the microcrack limit, the length of the equilibrium process zone is not constant even if the critical weakening displacement D_c and the strength drop $(\sigma_s - \sigma_d)$ are intrinsic material properties independent of the ambient stress and loading conditions. In particular, R is predicted to exponentially increase as the length of the stress drop region F decreases. Equations [38] and [32] suggest that the presence of sufficiently short cracks (such that $F < L_c$) has no effect on the macroscopic 'strength' of rocks. In particular, the prestress required for the crack extension must approach the static yield limit, $\sigma_0 \to \sigma_s$, and the size of the yield zone increases without bound, $L \approx R \to \infty$, for $F \to 0$ (see eqn [38]).

Quantitative estimates of the critical length scale L_c are not straightforward because it is not clear whether the slip-weakening distance D_c is indeed a scale-independent material constant (e.g., Barton, 1971; Rudnicki, 1980; Ohnaka, 2003). Laboratory measurements of the evolution of friction on smooth slip interfaces indicate that D_c may be of the order of 10^{-5} m (Li, 1987; Dieterich, 1979; Marone, 1998). For $\mu \sim 10^{10}$ Pa, and $(\sigma_s - \sigma_d) \sim 10^7 - 10^8$ Pa (likely spanning the range of strength drops for both 'strong' and 'weak' faults), from eqn [36] one obtains $L_c \sim 10^{-3} - 10^{-1}$ m, negligible compared to the characteristic dimension of the smallest recorded earthquakes, but comparable to the typical sample size used in the laboratory experiments. An upper bound on L_c may be obtained from estimates of the effective fracture energies of earthquake ruptures. For large (moment magnitude greater than 6) earthquakes, the seismically inferred fracture energies $G_c = (\sigma_s - \sigma_d)\, D_c$ are of the order of 10^6–10^7 J m^{-2} (Ida, 1972; Husseini, 1977; Beroza and Spudich, 1988; Abercrombie and Rice, 2005), rendering the effective $D_c \sim 0.01$–1 m. Assuming that the seismically inferred values of D_c are applicable to quasi-static cracks, eqn [36] suggests that a transition from 'micro' to 'macro' rupture regimes occurs at length scales of the order of 1–10^3 m.

The magnitude of a prestress required to initiate the crack propagation can be found by combining eqns [37] and [32]:

$$\frac{\sigma_s - \sigma_0}{\sigma_s - \sigma_d} = \frac{2}{\pi} \arcsin \exp\left(-\frac{L_c}{F}\right) \qquad [39]$$

In the microcrack, or large-scale yielding limit, $F \ll L_c$, eqn [39] predicts $\sigma_0 \approx \sigma_s$, that is, the crack propagation requires ambient stress comparable to the peak static strength of crustal rocks, as discussed above. In the small-scale yielding limit, $F \gg L_c$, eqn [39] gives rise to a well-known inverse proportionality between the stress drop $(\sigma_0 - \sigma_d)$, and the square root of the crack length F (e.g., Rice, 1968a; Kostrov, 1970; Cowie and Scholz, 1992):

$$\sigma_0 = \sigma_d + \frac{2}{\pi}(\sigma_s - \sigma_d)\left(\frac{2L_c}{F}\right)^{1/2} \qquad [40]$$

It follows from eqn [40] that for sufficiently large ruptures, the background tectonic stress required for the rupture propagation does not need to appreciably exceed the residual friction on the slipped surface σ_d. That is, the stress drop associated with the rupture propagation, $(\sigma_0 - \sigma_d)$, may be much smaller than the strength drop, $(\sigma_s - \sigma_d)$, provided that the rupture size significantly exceeds the critical nucleation size L_c. This statement forms the basis of the 'statically strong, but dynamically weak' fault theory (Lapusta and Rice, 2004). According to this theory, major crustal faults may operate at relatively low driving stresses (e.g., sufficient to explain the so-called heat flow paradox of the San Andreas fault; Brune *et al.*, 1969; Lachenbruch, 1980), even if the peak failure stress required for the onset of dynamic weakening is consistent with the Byerlee's law and hydrostatic pore pressures (Byerlee, 1978; Marone, 1998; Scholz, 2002), provided that $\sigma_d \ll \sigma_s$. If so, earthquake ruptures must nucleate in areas where σ_0 approaches σ_s (either due to locally increased ambient stress, or decreased static strength, for example, due to high pore fluid pressures), and propagate into areas of relatively low ambient stress. Under this scenario, the overall fault operation must be such that the average stress drop $\Delta\sigma$ remains relatively small (of the order of 0.1–10 MPa), and essentially independent of the rupture size (Kanamori and Anderson, 1975; Scholz, 2002; Abercrombie, 1995). Because the overall seismic moment release is dominated by the largest events, the implication from Lapusta and Rice (2004) model is that the Earth crust is not able to support high deviatoric stresses in the vicinity of large active faults. Phenomenologically, this is consistent with the 'weak fault' theory maintaining that the average shear stress σ_0 resolved on mature faults is of the order of the earthquake stress drops (i.e., up to a few tens of megapascals), and is considerably less than predictions based on the Byerlee's law (a few hundreds of megapascals) (e.g., Kanamori and Heaton, 2000). The 'statically strong but dynamically weak' fault theory seeks to reconcile laboratory results from rock friction experiments with seismic observations. Neither the peak shear stress σ_s nor the residual dynamic friction σ_d can be estimated from seismic data. Both of these parameters are likely scale dependent; for example, the peak shear stress σ_s may vary from gigapascals on the scale of microasperities and gouge particles (10^{-6} m) to the Mohr–Coulomb stress $f_s(\sigma_n - p)$, where f_s is the static coefficient of friction, σ_n is the fault-normal stress, and p is the pore fluid pressure, on the scale of centimeters to meters (i.e., consistent with laboratory data), to values that may be lower still on scales of hundreds of meters to kilometers. Similarly, the residual dynamic friction may depend on the amount of slip (and the rupture size) (e.g., Kanamori and Heaton, 2000; Abercrombie and Rice, 2005; Rice, 2006); implications from such behavior are further discussed is Section 4.03.6.

Some parallels may be drawn between the results rendered by the fracture mechanics analysis above, and empirical inferences from the rate-and-state friction formulation (Dieterich, 1992). In particular, for a given change in slip velocity, parameters a and b of the rate-and-state friction are analogous (upon scaling by the normal stress) to the 'direct effect' ($\sigma_s - \sigma_0$) and the strength drop ($\sigma_s - \sigma_d$), respectively (**Figure 4**). The critical length scale L_c (eqn [36]) is then essentially coincident with the critical nucleation size established by Dieterich (1992) based on numerical simulations coupling the rate-and-state friction and elasticity. The correspondence between the two length scales may be interpreted as indicating that no elastodynamic instability is possible in the microcrack regime ($F < L_c$). Recently, Rubin and Ampuero (2005) showed that L_c is in fact the minimum nucleation length, and that for a weakly weakening ($a/b \to 1$, i.e., nearly velocity neutral) material contact there may be a different nucleation size of the order of $b^2/(b-a)^2\ L_c$. Condition $a/b \to 1$ implies $\sigma_0 \approx \sigma_d$, that is, small-scale yielding (see eqn [40]). Indeed, conditions of small-scale yielding were satisfied prior to the onset of elastodynamic instability in all numerical experiments of Rubin and Ampuero (2005) performed under the assumption of a weak velocity weakening. Using eqns [32] and [38], one can see that in the limit $R/L \ll 1$, $b^2/(b-a)^2\ L_c \sim L/R\ L_c = L$, that is, the 'velocity-neutral' nucleation length of Rubin and Ampuero (2005) is essentially the length of a quasi-static crack on the verge of propagation under the condition of a constant remote stress σ_0. This is the maximum length of a slip zone satisfying conditions of a quasi-static equilibrium. Full numerical solutions in the framework of rate-and-state friction take into account continuous variations in slip velocity, implying a time-dependent evolution of σ_s and σ_d. The above analogy may be therefore considered a high-velocity limit of the coupled elastic-rate-and-state solution, given a logarithmic dependence of the effective shear stresses on relative variations in slip velocity.

4.03.5 Dynamic Effects

Quasistatic solutions considered in the previous section are valid only for rupture speeds that are well below the shear wave velocity V_s. As the rupture velocity increases, the inertial term in the equilibrium eqn [10] eventually becomes non-negligible, and the near-tip stress field is significantly altered when V_r becomes a sizeable fraction of V_s (Andrews, 1976a; Broberg, 1978; Freund, 1979; Rice, 1980). The most pronounced effects of a high rupture speed are the relativistic shrinking of the in-plane process zone R, and simultaneous increase of stress perturbations off the crack plane (Kame and Yamashita, 1999; Poliakov *et al.*, 2002; Rice *et al.*, 2005). The net result is an increased tendency for branching, bifurcation, and nonsteady propagation, all of which significantly complicate the analytic and numerical treatment of the elastodynamic rupture problem. In particular, the intermittent propagation of the rupture front invalidates the equivalence between the LEFM and slip-weakening formulations (e.g., Freund, 1979), implying a greater dependence of the model results on a specific choice of fracture criteria.

Analysis of the full elastodynamic equilibrium eqn [10] reveals that solutions exist for rupture velocities below a limiting speed V_l, which equals to the Rayleigh wave velocity V_R for Mode II cracks, and shear wave velocity for Mode III cracks (e.g., Freund, 1979; Kostrov and Das, 1988). Solutions become singular as $V_r \to V_l$; in particular, the dynamic stress intensity factor and the energy release rate at the crack tip asymptotically vanish. Main fracture mechanics parameters of a steady-state elastodynamic rupture (such as the process zone size and the stress intensity factors) may be readily obtained by multiplying or dividing the respective results from quasi-static solutions by dimensionless coefficients that depend on rupture velocity only. The corresponding coefficients are

$$f_{II} = (1-\nu)\frac{4\sqrt{1-\frac{V_r^2}{V_p^2}}\sqrt{1-\frac{V_r^2}{V_s^2}}-\left(2-\frac{V_r^2}{V_s^2}\right)^2}{\frac{V_r}{V_s}\sqrt{1-\frac{V_r^2}{V_s^2}}} \qquad [41]$$

$$f_{III} = \sqrt{1-\frac{V_r^2}{V_s^2}} \qquad [42]$$

for the mode II and III loading, respectively (e.g., Rice, 1980; Freund, 1998). In eqn [41], V_p is the P wave velocity. Coefficients f_{II} and f_{III} monotonically decrease from unity at $V_r = 0$ to zero at $V_r = V_l$. For instance, the length of the dynamic process zone R^* is given by $R^*(V_r) = f_{II,\,III}(V_r)R$, and the dynamic stress intensity factor for a self-similar expanding crack is $K_{II,\,III}(V_r) = f_{II,\,III}(V_r)K_{II,\,III}(0)$. Note that for a steady-state self-healing pulse with a fixed stress distribution in the reference frame of a moving pulse, the process zone shortens at high rupture velocity, similar to the case of a self-similar crack, but the dynamic stress

intensity factor is independent of V_r. Rice *et al.* (2005) confirmed these results with full analytic solutions for a self-healing pulse with a linearly weakening process zone, propagating at a constant rupture speed. They found that the ratio of the dynamic process zone R^* to the pulse length L^* is independent of the rupture speed, but is dependent on the ratio of stress drop to strength drop:

$$\frac{\sigma_0-\sigma_d}{\sigma_s-\sigma_d}=\frac{\xi}{\pi}-\frac{\xi-\sin\xi}{2\pi\sin^2(\xi/2)} \qquad [43]$$

where $\xi=2\arcsin\sqrt{R^*/L^*}$ (cf. eqn [32]). The velocity invariance of the relative size of the process zone gives rise to a somewhat unintuitive result that the pulse length L^* vanishes as the rupture accelerates to a limiting speed. At the same time, the amount of slip produced by a self-healing pulse is also invariant with respect to the rupture velocity. Thus, the dynamically shrinking rupture size gives rise to a dramatic increase in the near-field coseismic strain. **Figure 5** shows the dependence of the near-tip stress concentration on the rupture speed inferred from the Rice *et al.* (2005) model. As one can see from **Figure 5**, the extent of the off-plane yielding and damage substantially increases at high rupture speeds, so that the assumption of a thin in-plane process zone ceases to be valid. Similar results were reported for a semi-infinite crack by Poliakov *et al.* (2002). Note that the deduced areas of yielding likely underestimate the extent of off-fault damage, as stresses inside the shaded areas in **Figure 5** are beyond the failure envelope. By explicitly allowing inelastic deformation, the excess stresses will be relaxed, and the yielding zone will further expand. The enhanced off-fault damage may be one of the mechanisms preventing extreme contractions of the rupture length at high propagation velocities. It may also appreciably modify the expenditure part of the earthquake energy balance, as discussed in the following section.

Elastodynamic slip instabilities with rupture velocity below the limiting speed V_l are referred to as subsonic or subshear rupture. The majority of earthquakes for which high-quality measurements of the rupture speed are available appear to be subsonic. Theoretical models indicate that immediately above the limiting rupture speed V_l, the flux of mechanical energy into the crack tip region becomes negative for tensile and mode II shear cracks, effectively prohibiting self-sustained fracture (Freund, 1998; Broberg, 1999). However, physically admissible solutions do exist for shear cracks with rupture speeds spanning the interval between the S wave and P wave velocities;

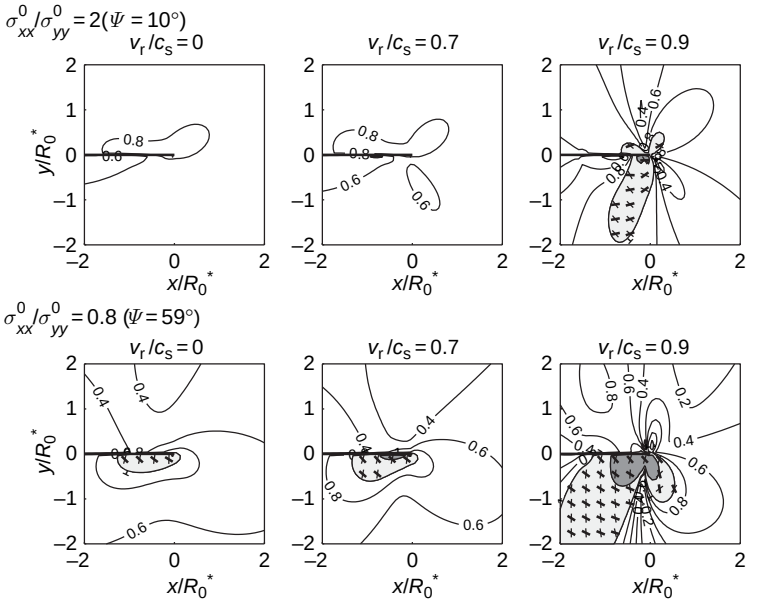


Figure 5 Stress perturbation around the tip of a propagating slip pulse with a linear slip-weakening process zone (figure 6, Rice *et al.*, 2005). Light shading denote areas where the Mohr–Coulomb failure envelope is exceeded (likely to fail in shear), and dark shading denotes areas of absolute tension (likely to fail by tensile cracking). Thick tick marks indicate the orientation of optimal planes for right-lateral slip, and thin tick marks indicate the orientation of optimal planes for left-lateral slip. ψ is an angle between the maximum compression axis and the fault plane. Process zone comprises 10% of the rupture length. Axes are normalized by the quasi-static process zone size.

such mode of propagation is referred to as intersonic, transonic, or supershear rupture (Burridge, 1973; Andrews, 1976b; Simonov, 1983; Broberg, 1989). Although the mechanics of transition from the subsonic ($V_r < V_l$) to transonic ($V_r > V_s$) propagation is not fully understood, there is experimental (Samudrala *et al.*, 2002; Xia *et al.*, 2004) and seismologic (Archuleta, 1984; Wald and Heaton, 1994; Bouchon and Vallee, 2003; Dunham and Archuleta, 2004) evidence that transonic rupture speeds may be achieved under certain conditions. The spatial structure of the near-tip stress field, and the radiation pattern in the transonic regime, are markedly different from those due to subsonic ruptures (*see* Chapter 4.06).

4.03.6 Fracture Energy

The concept of fracture energy was originally introduced for tensile cracks by Griffith (1920) to quantify the irreversible work associated with breaking of the intermolecular bonds and creation of a stress-free crack surface. Griffith's definition based on a global energy balance was subsequently shown to be equivalent to local definitions based on the LEFM and small-scale yielding models (Willis, 1967; Rice, 1968a). For example, for a Barenblatt-type process zone model, the fracture energy is the work spent against the cohesive stress σ_s in the process zone on separating the crack walls by the critical opening distance D_c. An elegant demonstration of the equivalence of global and local definitions of fracture energy for tensile cracks was provided by Rice (1968b) in a form of the path-independent $\mathcal{J}$-integral (also, see Eshelby, 1956; Cherepanov, 1968). Palmer and Rice (1973) extended this technique to the case of shear cracks, and defined the shear fracture energy as work required to evolve shear stress on the slip interface from the yield stress (or static friction) σ_s to the residual dynamic friction σ_d:

$$G_c = \int_0^{D_c} (\sigma(D) - \sigma_d)\mathrm{d}D \qquad [44]$$

where $\sigma(D)$ varies between σ_s and σ_d for $0 < D < D_c$, respectively. A similar formulation was introduced by Ida (1972). Equation [44] allows a simple insight into the fracture process, and has been widely used for interpretations of seismic data. However, several factors may limit its application to the analysis of earthquake ruptures. First, the displacement-weakening model assumes that all inelastic deformation is limited to the slip plane. Both theoretical models (Rudnicki, 1980; Andrews, 2005; Rice *et al.*, 2005; **Figure 5**) and field observations (Li *et al.*, 1998; Fialko *et al.*, 2002; Fialko, 2004b; Chester *et al.*, 2005) suggest that the earthquake-induced damage likely extends well off the fault plane, and the energy dissipated in the fault damage zone may be quite significant (e.g., Wilson *et al.*, 2004; Andrews, 2005; Ben-Zion and Shi, 2005). Second, the fracture energy given by eqn [44] has a clear physical interpretation if the residual dynamic stress σ_d is constant (or at least if the along-fault variations in σ_d are small compared to the strength drop, $\sigma_s - \sigma_d$). The second point can be illustrated by considering a traditional representation of the earthquake energy budget:

$$\Delta U_p = U_r + U_f + U_G \qquad [45]$$

where ΔU_p is the change in the total potential energy (which includes changes in the elastic strain energy ΔU_e, gravitational potential energy, etc.), U_r is the energy radiated in seismic waves, U_f is the energy dissipated on the well-slipped portion of the fault due to friction, communition, phase transitions, and other irreversible losses, and U_G is the fracture energy spent on overcoming high resisting stresses near the crack tip (Kostrov, 1974; Dahlen, 1977; Rudnicki and Freund, 1981; Rivera and Kanamori, 2005). A significant part of U_f is believed to be ultimately converted into heat (Sibson, 1980; Fialko, 2004a). Under the approximation of the displacement-weakening model, $U_G \propto \sigma_s D_c$, and $U_f \propto \sigma_d D_m$. Assuming that the residual friction is of the order of the peak strength, $\sigma_d \sim O(\sigma_s)$, and the critical slip-weakening distance is much smaller than the coseismic offset, $D_c \ll D_m$, the fracture energy is negligible compared to frictional losses in the earthquake energy balance [45]. However, if the fault friction progressively decreases with slip, as suggested by the experimental observations and theoretical inferences of the dynamic weakening (Tsutsumi and Shimamoto, 1997; Goldsby and Tullis, 2002; Di Toro *et al.*, 2004; Abercrombie and Rice, 2005; Fialko and Khazan, 2005), the effective slip-weakening distance D_c is expected to scale with the slip magnitude, and the fracture energy U_G may not be small compared to U_f. Because neither the slip-weakening distance D_c nor the residual friction σ_d in this case are material properties (in particular, they may depend on the details of slip history, thickness and permeability of the slip zone, etc.), a distinction between U_G and U_f terms in the earthquake energy balance eqn [45] becomes somewhat arbitrary. Note that for a rupture on a pre-existing fault there is little 'physical' difference between U_G and U_f, as both terms represent spatially and temporally variable frictional losses associated

with fault slip; both U_G and U_f ultimately contribute to wear and heating on the slip interface. The situation is further complicated if the dynamic friction is a non-monotonic function of slip (e.g., Brune and Thatcher, 2003; Hirose and Shimamoto, 2005; Rivera and Kanamori, 2005; Tinti *et al.*, 2005). While a formal distinction between the frictional and fracture losses associated with shear ruptures may be problematic, the entire amount of work spent on inelastic deformation of the host rocks during the crack propagation is unambiguous and can be readily quantified. For simplicity, we consider the case of a quasi-static crack growth here. Formulation presented below can also be generalized to the case of dynamic cracks.

Consider an equilibrium mode II crack in a medium subject to initial stress σ_{ij}^0. The medium is elastic everywhere except inside the crack, and within a finite process zone near the crack tips (**Figure 6(a)**). The inelastic zone is demarcated by a surface S_f. Let external forces do some work δW on a medium, as a result of which the crack acquires a new equilibrium configuration. In the new configuration, some area ahead of the crack front undergoes inelastic yielding and joins the process zone (see an area bounded by surfaces ΔS_1 and S_f in **Figure 6(a)**. At the trailing end of the process zone, slip exceeds D_c, and some fraction of the process zone (bounded by surface ΔS_2 in **Figure 6(a)** joins the developed part of the crack. The external work δW is spent on changes in the elastic strain energy δU_e, and irreversible inelastic deformation δU_G (which includes friction, breakdown, communition, etc.):

$$\delta W = \delta U_e + \delta U_G \qquad [46]$$

Changes in the elastic strain energy are given by (e.g., Timoshenko and Goodier, 1970; Landau and Lifshitz, 1986)

$$\delta U_e = \frac{1}{2}\left(\sigma_{ij}^1 \varepsilon_{ij}^1 - \sigma_{ij}^0 \varepsilon_{ij}^0\right) \qquad [47]$$

where σ_{ij}^1 and ε_{ij}^1 are stresses and strains, respectively, in the elastic part of a medium after the crack extension. The assumption of linear elasticity [9] implies that

$$\sigma_{ij}^0 \varepsilon_{ij}^1 \equiv \sigma_{ij}^1 \varepsilon_{ij}^0 \qquad [48]$$

for any σ_{ij}^k and ε_{ij}^k. The identity [48] allows one to write eqn [47] as follows:

$$\begin{aligned}\delta U_e &= \frac{1}{2}\left(\sigma_{ij}^1 + \sigma_{ij}^0\right)\varepsilon_{ij}^1 - \frac{1}{2}\left(\sigma_{ij}^1 + \sigma_{ij}^0\right)\varepsilon_{ij}^0 \\ &= \frac{1}{2}\left[\left(\sigma_{ij}^1 + \sigma_{ij}^0\right)\delta u_i\right]_{,j}\end{aligned} \qquad [49]$$

where $\delta u_i = u_i^1 - u_i^0$ is the displacement field produced by crack propagation. Expressions [47] and [49] assume that the strains are infinitesimal, so that the relationship between strain and displacement gradients is given by eqn [8]. Also, expression [49] makes use of the fact that under quasi-static conditions, the divergence of stress is zero, $\sigma_{ij,j} = 0$ (see the equilibrium eqn [10]; note that the body forces may be excluded from consideration by incorporating the effects of gravity in prestress). Using the Gauss theorem along with a condition that the crack-induced deformation must vanish at infinity, from the energy balance eqn [46] one obtains the following expression for the work done on inelastic deformation:

$$\begin{aligned}\delta U_G &= \delta W - \delta U_e \\ &= -\int_{S_f} \sigma_{ij}^0 n_j \delta u_i \mathrm{d}S - \frac{1}{2}\int_{\Delta S_1 + \Delta S_2} \left(\sigma_{ij}^0 + \sigma_{ij}^1\right) n_j \delta u_i \mathrm{d}S\end{aligned} \qquad [50]$$

In the limit of an ideally brittle fracture, the area of inelastic yielding has a negligible volume (i.e., $S_f \to 0$), so that the first integral on the right-hand side of eqn [50] vanishes. In the second integral, the surface ΔS_2 also vanishes, while the surface ΔS_1 becomes the crack length increment δL (**Figure 6(b)**). Within δL, the stresses are weakly singular, $\sigma_{ij} \propto 1/\sqrt{r}$, where r is distance to the crack tip, and the crack wall displacements scale as $\delta u_i \propto \sqrt{r}$, so that the product of

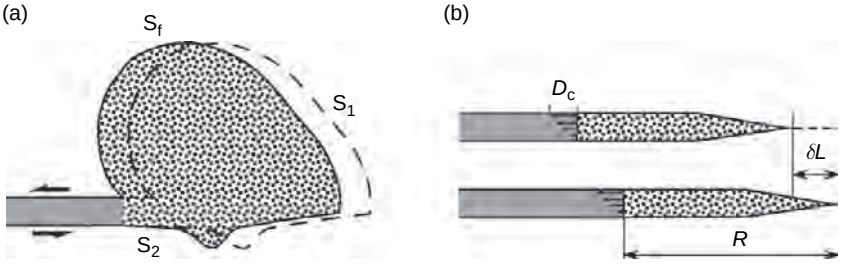


Figure 6 Schematic view of inelastic deformation associated with crack propagation. (a) Finite damage zone extending off the fault plane. (b) Thin in-plane damage zone.

stresses and displacements is of the order of unity, and the corresponding integral in eqn [50] is of the order of $\Delta S_1 = \delta L$. This is the well-known LEFM limit, for which the fracture energy is given by eqn [2].

For more realistic models that explicitly consider failure at the crack tip, the stresses are finite everywhere, so that the second integral on the right-hand side of eqn [50] is of the order of $\sigma^0_{ij}\delta u_i(\Delta S_1 + \Delta S_2)$, that is, negligible compared to the integral over the finite inelastic zone $\sim O\left(\sigma^0_{ij}\delta u_i S_f\right)$. Provided that the displacement field associated with the crack extension can be represented as $\delta u_i = \delta L \partial u_i/\partial L$, eqn [50] allows one to introduce the fracture energy G_c as the total inelastic work per increment of the crack length, or the energy release rate,

$$G_c = \frac{\partial U_G}{\partial L} = -\frac{1}{2}\int_{S_f} \frac{\partial u_i}{\partial L}\sigma^0_{ij} n_j \mathrm{d}S \qquad [51]$$

Factor of $1/2$ in eqn [51] stems from the assumption of a bilateral crack propagation. Equation [51] in general cannot be readily evaluated analytically because the size and geometry of the inelastic zone S_f are not known in advance, and have to be found as part of a solution. Further insights are possible for special cases. For example, assuming that all yielding is confined to a crack plane (**Figure 6(b)**), eqn [51] reduces to

$$G_c = \int_0^L \sigma(x)\frac{\partial D(x)}{\partial L}\mathrm{d}x \qquad [52]$$

where we took into account that the total offset between the crack walls is $D(x) = 2u_x$, and the sense of slip is opposite to that of the resisting shear tractions acting on the crack walls. The integral [52] is still intractable for an arbitrary loading, as the derivative $\partial D(x)/\partial L$ must be calculated along the equilibrium curve (e.g., see eqn [33]). Closed-form analytic solutions can be obtained for limiting cases of small-scale ($L \gg R$) and large-scale ($L \approx R$) yielding. First, consider a crack that is much longer than the critical size L_c [36], and has a complete stress drop, $\sigma(x) = 0$ for $D > D_c$. As shown in Section 4.03.2, for such a crack, the size of the process zone is independent of the crack length, $R = L_c$, and the crack propagation does not modify the slip distribution within the process zone in the reference frame of a propagating crack tip. In this case,

$$\partial D(x)/\partial L = -\partial D(x)/\partial x \qquad [53]$$

so that eqn [52] gives rise to

$$G_c = \int_{L-R}^{L} \sigma(x)\frac{\partial D(x)}{\partial L}\mathrm{d}x = \int_0^{D_c} \sigma(D)\mathrm{d}D \qquad [54]$$

Expression [54] is analogous to the result obtained using the $\mathcal{J}$-integral technique (Rice, 1968b). An expression for the $\mathcal{J}$-integral contains a derivative of the crack wall displacement with respect to the integration variable, rather than the crack length. As noted by Khazan and Fialko (2001), the two derivatives coincide (up to a sign) in a limiting case of a very long crack (for which the $\mathcal{J}$-integral was derived), but the difference may be significant if the small-scale yielding approximation does not hold (also, see Rice, 1979).

For a case of a constant, but nonvanishing residual friction, $\sigma(x) = \sigma_d$ for $D_c < D < D_m$, evaluation of integral [52] gives rise to

$$G_c = \sigma_d(D_m - D_c) + \int_0^{D_c} \sigma(D)\mathrm{d}D \qquad [55]$$

thanks to self-similarity of the along-crack displacement $D(x)$ over the interval $0 < x < L - R$ for long cracks, such that the relationship [53] still holds. For a constant yield stress within the process zone, $\sigma(x) = \text{const.} = \sigma_s$ (e.g., **Figure 4**), eqn [55] gives rise to a simple expression $G_c = (\sigma_s - \sigma_d)\,D_c + \sigma_d D_m$, in which the first and second terms may be recognized as the traditionally defined fracture energy U_G, and the frictional work U_f, respectively (cf. eqn [45]).

In the case of a large-scale yielding, relationship [53] is generally not applicable. Assuming $\sigma_s = \text{const.}$, expression [52] can be integrated by parts to yield

$$G_c = \sigma_s \frac{\partial}{\partial L}\int_0^L D(x)\mathrm{d}x - \sigma_s \int_0^F \frac{\partial D}{\partial L}\mathrm{d}x \qquad [56]$$

In the developed part of the crack with the full stress drop ($x < F$), slip is essentially constant and is equal to D_c, so that the second integral on the right-hand side of eqn [56] can be neglected. Taking advantage of expressions [33] and [35] to evaluate the first term in eqn [56], one obtains (Khazan and Fialko, 2001)

$$G_c = \frac{\pi F}{2L_c}\sigma_s D_c \qquad [57]$$

Equation [57] indicates that the fracture energy for the case of small cracks (or large-scale yielding, $F \ll L_c$) is substantially different from the fracture energy for large cracks (or small-scale yielding, $F \gg L_c$). In the case of large-scale yielding, the fracture energy is not constant even if both the yield strength σ_s and the critical slip-weakening displacement D_c are material constants independent of loading conditions. In particular, G_c is predicted to linearly increase with the size of the developed part of the crack F. A linear scaling of fracture energy implies increases in the apparent fracture toughness

K_Q proportional to a square root of the developed crack length, $K_Q \propto \sqrt{F}$ (see eqn [2]), in the large-scale yielding regime. Increases in the apparent fracture toughness for small cracks are well known from laboratory studies of tensile fracture (e.g., Ingraffea and Schmidt, 1978; Bazant and Planas, 1998).

The apparent scaling of the earthquake fracture energy with the earthquake size has been inferred from seismic data (e.g., Husseini, 1977; Kanamori and Heaton, 2000; Abercrombie and Rice, 2005). Arguments presented above indicate that several mechanisms may be responsible for the observed increases in the seismically inferred fracture energies with the rupture length, in particular, (1) off-fault damage that scales with the rupture length (larger ruptures are expected to produce broader zones of high stress near the rupture fronts, presumably advancing the extent of off-fault damage; see **Figure 5** and eqn [51]); (2) a continuous degradation of dynamic friction on a fault plane, for example, due to thermal pressurization or any other slip-weakening mechanism; and (3) rupture propagation under conditions of large-scale yielding (eqn [57]), although it remains to be seen whether elastodynamic instability can occur when the process zone comprises a substantial fraction of the crack length (see Section 4.03.4). These mechanisms are not mutually exclusive, and may jointly contribute to the observed scaling $G_c \propto F$. For example, the third mechanism might be relevant for small earthquakes, while the first and second ones perhaps dominate for large events. Note that the second mechanism is ultimately limited by a complete stress drop, beyond which no further increase in fracture energy is possible. The same limit may also apply to the first mechanism, as the size of the dynamic damage zone scales with the quasi-static one for a given displacement-weakening relationship (Rice *et al.*, 2005). Establishing the relative importance of contributions of various mechanisms to the effective fracture energy is an important but challenging task. In particular, if contributions from the off-fault yielding are substantial, interpretations of seismic data that neglect such yielding may systematically overestimate the magnitude of the effective slip-weakening distance D_c. Unfortunately, distinguishing between different contributions to the overall value of G_c is unlikely to be accomplished based on the seismic data alone.

In summary, the fracture energy defined by eqn [51] is analogous to the Griffith's concept for tensile cracks, provided that the stress drop is complete ($\sigma_d = 0$), and the small-scale yielding condition is met. For a nonvanishing friction on the crack surface, eqn [51] combines inelastic work spent against residual friction, as well as work spent on evolving the shear stress on a fault to a residual level (i.e., the traditionally defined fracture energy). Separation between these two contributions is justified if small-scale yielding condition applies, but may be ill-defined otherwise. For models with a continuous strength degradation, there is a continuous repartitioning of the energy budget, such that the effective fracture energy increases at the expense of a diminishing frictional dissipation.

4.03.7 Coupling between Elastodynamics and Shear Heating

One of the factors that can strongly affect the dynamic friction on the slipping interface, and thereby the seismic radiation, efficiency, and stress drop, is the coseismic frictional heating. Rapid slip during seismic instabilities may substantially raise temperature on a fault surface. The dependence of dynamic friction on temperature may stem from several mechanisms, including thermal pressurization by pore fluids (Sibson, 1973; Lachenbruch, 1980; Mase and Smith, 1987), frictional melting (Jeffreys, 1942; McKenzie and Brune, 1972; Sibson, 1975; Maddock, 1986), and flash heating of contact asperities (Rice, 2006). Recent experimental measurements confirm appreciable variations in the dynamic friction at slip velocities approaching the seismic range of order of a meter per second (Goldsby and Tullis, 2002; Di Toro *et al.*, 2004; Hirose and Shimamoto, 2003; Spray, 2005). The documented variations are significantly larger than predictions of the rate-and-state friction (Dieterich, 1979; Ruina, 1983) extrapolated to seismic slip rates. The likely importance of the thermally induced variations in friction warrants a quantitative insight into the dynamics of fault heating during seismic slip. Major questions include: What are the dominant mechanisms of fault friction at high slip rates? How do increases in temperature affect the dynamic shear stress on a slipping interface? How robust is the thermally activated weakening? Is dynamic friction a monotonically decaying function of temperature? If not, what are the mechanisms, conditions, and significance of the thermally activated strengthening? When and where is the onset of the thermally induced weakening or strengthening likely to occur on a slipping interface, and what are the implications for the dynamics of earthquake ruptures?

In a simple case of the LEFM crack with a constant residual shear stress, the frictional heating

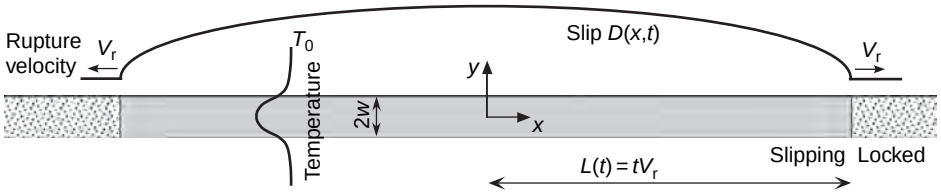


Figure 7 A schematic view of a dynamically propagating mode II crack. The crack has a thickness 2ω, and is rupturing bilaterally at a constant velocity $dL/dt = V_r$.

problem admits closed analytic solutions (Richards, 1976; Fialko, 2004a). Consider a mode II (plain strain) crack rupturing bilaterally at a constant speed V_r (**Figure 7**), so that $L(t) = tV_r$. The thickness of the gouge layer that undergoes shear and the shear strain rate across the gouge layer are assumed to be constant (Cardwell *et al.*, 1978; Mair and Marone, 2000). Because the thickness of the gouge layer $2w$ is negligible compared to any other characteristic length scale in the problem (e.g., the rupture size $2L$ and the amount of slip D), the temperature evolution in the gouge layer and in the ambient rock is well described by the one-dimensional diffusion equation with a heat source,

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial y^2} + \frac{Q}{c\rho} \qquad [58]$$

where y is the crack-perpendicular coordinate, and Q is the rate of frictional heat generation within the slipping zone:

$$Q(x, y, t) = \begin{cases} \dfrac{\sigma_d(x)}{2w(x)} \dfrac{\partial D(x, t)}{\partial t}, & t > 0, \ |y| < w \\ 0, & |y| > w \end{cases} \qquad [59]$$

$\partial D/\partial t$ being the local slip velocity. A solution to eqn [58] subject to the initial condition $T(x, y, 0) = T_0$, where T_0 is the temperature of the host rocks prior to faulting, is (Fialko, 2004a)

$$T - T_0 = \frac{1}{4c\rho w} \int_{x/V_r}^{t} \left(\operatorname{erf}\left[\frac{y + w}{2\sqrt{\kappa(t-\tau)}} \right] - \operatorname{erf}\left[\frac{y - w}{2\sqrt{\kappa(t-\tau)}} \right] \right) \frac{\partial D(x, \tau)}{\partial \tau} \sigma_d(x) d\tau \qquad [60]$$

Dimensional arguments suggest the following similarity variables:

$$\text{Nondimensional along-fault coordinate:} \quad \chi = \frac{x}{tV_r} \qquad [61]$$

$$\text{Nondimensional fault thickness:} \quad \bar{w} = \sqrt{\frac{2}{\kappa t}} w \qquad [62]$$

For the LEFM crack with a constant stress drop, the along-crack displacement profile $D(x, t)$ is self-similar in that it may be expressed in terms of a single similarity variable $\chi = \chi(x, t)$:

$$D(x, t) = L(t)\varepsilon\sqrt{1 - \chi^2}, \quad t > 0, \ |\chi| < 1 \qquad [63]$$

where ε is the characteristic shear strain due to the crack, $\varepsilon = D(0, t)/L(t) = 2(1 - \nu)(\sigma_0 - \sigma_d)/\mu$. Here, ε is taken to be independent of L, as the earthquake stress drops $(\sigma_0 - \sigma_d)$ do not exhibit any scale dependence across a wide range of earthquake magnitudes (Kanamori and Anderson, 1975; Scholz, 2002; Abercrombie, 1995). The local slip rate in terms of new variables is

$$\frac{\partial D}{\partial t} = \frac{V_r \varepsilon}{\sqrt{1 - \chi^2}} \qquad [64]$$

Equations [60] and [64] suggest the following similarity variable for temperature:

$$\text{Nondimensional temperature:} \quad \theta = \frac{T - T_0}{\hat{T}} \qquad [65]$$

where

$$\hat{T} = \frac{\sigma_d V_r \varepsilon}{c\rho} \sqrt{\frac{t}{\pi\kappa}} \qquad [66]$$

is a characteristic temperature scale for frictional heating assuming a perfectly sharp fault contact.

Substituting eqn [64] into [60], and making use of the similarity variables [62] and [65], one obtains the following expression for the along-crack temperature distribution in the middle of the slip zone ($y = 0$):

$$\theta(\chi) = \frac{\sqrt{\pi}}{\bar{w}\sqrt{2}} \int_{\chi}^{1} \operatorname{erf}\left[\frac{\bar{w}}{2\sqrt{2(1 - \xi)}} \right] \frac{\xi d\xi}{\sqrt{\xi^2 - \chi^2}} \qquad [67]$$

Solutions to eqn [67] are shown in **Figure 8**. A family of curves in **Figure 8** illustrates a spatiotemporal

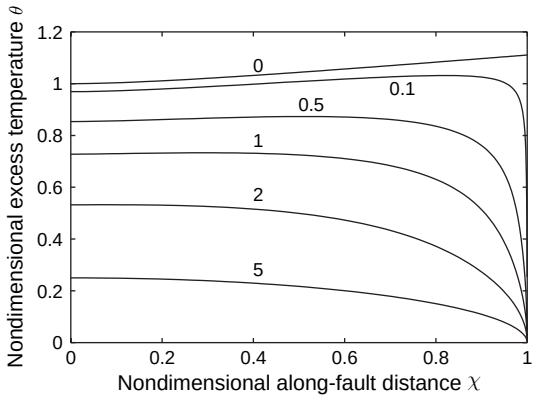


Figure 8 Variations of the nondimensional excess temperature $\theta(\chi)$ along a mode II crack propagating at a constant rupture speed under constant frictional stress. Labels denote the nondimensional thickness of a slipping zone $\bar{w}$ (see eqn [62]).

evolution of temperature on the slipping fault surface. For faults that are thicker than the thermal diffusion length scale, or at early stages of rupture (i.e., $\bar{w} > 1$), the temperature increase along the fault is proportional to the amount of slip. For thin faults, or later during the rupture ($\bar{w} \ll 1$), the temperature is maximum near the crack tip, and decreases from the tip toward the crack center. However, the instantaneous temperature maximum near the crack tip does not imply cooling of the crack surface behind the tip; at any given point the temperature on the crack surface $T(x, 0, t)$ steadily increases with time (Fialko, 2004a). For the nondimensional fault thickness $\bar{w}$ of the order of unity, the maximum temperature is reached somewhere between the crack center and the rupture front (**Figure 8**). The inferred anticorrelation between the temperature and the amount of slip stems from a competition between the rates at which the frictional heat is generated at the crack surface, and removed to the ambient rocks by conduction. Generation of frictional heat at the tip of the LEFM crack is singular as the thickness of the conductive boundary layer is zero, while the slip velocity is infinite (see eqn [64]). Nonetheless, the excess temperature at the tip is zero for cracks having finite thickness ($\bar{w} > 0$). For cracks that are much thinner than the conductive boundary layer ($\bar{w} \ll 1$), the temperature field develops a shock-like structure, with the tip temperature exceeding the temperature at the crack center by about 10% (**Figure 8**). Assuming that the thickness of the slip zone is constant during an earthquake, eqn [67] predicts that the maximum temperatures are initially attained at the center of a crack-like shear instability. As the earthquake rupture expands, the temperature maximum may migrate toward the rupture fronts. For the thermal diffusivity of the ambient rocks $k = 10^{-6}$ m^2 s^{-1}, and rupture durations of $t = 1$–10 s (corresponding to the rupture sizes of ~5–50 km), this transition will occur for faults that have thickness of the order of $\sqrt{2\kappa t} \sim 2-5$mm or less. The critical fault thickness may be larger still if the heat removal from the fault involves some advective transport by the pressurized pore fluids, and the *in situ* hydraulic diffusivity exceeds the thermal diffusivity k.

For sufficiently large ruptures, a model of a self-healing slip pulse may be a better approximation than the crack-like models (e.g., Kanamori and Anderson, 1975; Heaton, 1990; Beroza and Mikumo, 1996; Olsen *et al.*, 1997). A self-healing mode II pulse having a constant length L (Freund, 1979) generates a temperature field that is steady state in the reference frame of a moving rupture front (Fialko, 2004a). The appropriate similarity variables are

$$\text{Along-fault coordinate:} \quad \chi = \frac{x - tV_r}{L} + 1 \qquad [68]$$

$$\text{Nondimensional fault thickness:} \quad \bar{w} = \sqrt{\frac{2V_r}{L\kappa}} w \qquad [69]$$

$$\text{Nondimensional temperature:} \quad \theta = \frac{T - T_0}{\hat{T}} \qquad [70]$$

$$\hat{T} = \frac{\sigma_d \varepsilon}{c\rho} \sqrt{\frac{LV_r}{\pi\kappa}} \qquad [71]$$

The coseismic displacements and the rate of slip are assumed to have the LEFM-like characteristics at the rupture front ($\chi = 1$), and a nonsingular healing at the trailing edge ($\chi = 0$):

$$D(\chi) = L\varepsilon\sqrt{1-\chi^2} \qquad [72]$$

$$\frac{\partial D}{\partial \chi} = \frac{V_r \varepsilon \chi}{\sqrt{1-\chi^2}} \qquad [73]$$

As before, we assume a constant dynamic friction on the slipping interface. Upon nondimensionalization using variables [71], eqn [60] gives rise to the following expression for the along-fault temperature variations in the middle of the slip zone ($y = 0$):

$$\theta(\chi) = \frac{\sqrt{\pi}}{\bar{w}\sqrt{2}} \int_{\chi}^{1} \operatorname{erf}\left[\frac{\bar{w}}{2\sqrt{2(\chi - \xi)}}\right] \frac{\xi \, d\xi}{\sqrt{1-\xi^2}} \qquad [74]$$

Solutions to eqn [74] are shown in **Figure 9**.

The near-tip structure of the temperature field due to a steady-state pulse is similar to that due to a

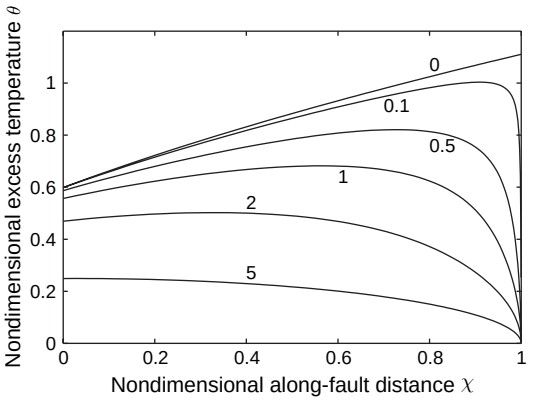


Figure 9 Variations of the nondimensional excess temperature $\theta(\chi)$ along a self-healing pulse. Labels denote the nondimensional thickness of a slipping zone $\bar{w}$ (see eqn [71]).

self-similar expanding crack (cf. **Figures 8** and **9**). At the leading edge of an infinitesimally thin shear pulse, there is a thermal shock of amplitude $\hat{T}$ (eqn [71]). The fault temperature monotonically decreases toward the healing front, where the temperature falls to about one-half of the maximum value (**Figure 9**). For 'thick' pulses ($\bar{w} \gg 1$), the fault temperature increases toward the healing front proportionally to the amount of slip. For intermediate nondimensional fault thicknesses of the order of unity, the fault surface initially heats up to a maximum temperature, and then cools down before the arrival of the healing front. This behavior is qualitatively different from that on a surface of an expanding crack, which indicates a progressive heating at every point along the crack as long as the rupture continues. The inferred cooling toward the healing front of the steady state pulse for $\bar{w} < 5$ is caused by a decreasing heat generation due to a vanishing slip velocity, and efficient removal of heat by thermal diffusion. For the characteristic rise times L/V_r of the order of seconds, the steady-state LEFM pulses need to be thinner than $5\sqrt{2\kappa L/V_r} \approx 1$ cm to experience maximum temperatures at the rupture front.

Several factors may accentuate the tendency for the temperature peaks near the rupture front. First, higher stresses in the process zone near the rupture tip imply enhanced heating. Numerical simulations indicate that the thermal effect of the process zone can be significant even under conditions of small-scale yielding; in particular, for thin faults, the instantaneous temperature increase within the process zone is predicted to be a factor of σ_s/σ_d greater than the temperature increase on the rest of the slipping interface (Fialko, 2004a). Second, in the presence of a continued dynamic weakening (e.g., Abercrombie and Rice, 2005), the frictional heating is expected to progressively decay behind the rupture front, further suppressing the excess temperature. In the context of thermal weakening, such coupling between the coseismic heating and dynamic friction may be conducive to self-sustained slip pulses (Perrin *et al.*, 1995; Zheng and Rice, 1998), as well as to a transition to a pulse-like behavior for ruptures that initially propagate in a crack-like mode.

If thermal weakening mechanisms that operate at relatively small increases in the average fault temperature, such as thermal pressurization and flash melting (e.g., Lachenbruch, 1980; Lee and Delaney, 1987; Andrews, 2002; Rice, 2006), do not give rise to substantial reductions in the dynamic friction, a continued dissipation and heating due to a localized slip will result in macroscopic melting on the slip interface, and a transition from the asperity-contact friction to viscous rheology. High transient stresses associated with shearing of thin viscous films of melt have been considered as one of possible mechanisms of thermally induced strengthening (Tsutsumi and Shimamoto, 1997; Fialko, 2004a; Koizumi *et al.*, 2004). However, recent laboratory, field, and theoretical studies suggest that transient viscous braking may not be an efficient arresting mechanism, especially in the lower part of the brittle layer (Fialko and Khazan, 2005; Di Toro *et al.*, 2006). If so, earthquakes that produced macroscopic melting and pseudotachylites (e.g., Sibson, 1975; Swanson, 1992; Wenk *et al.*, 2000; Davidson *et al.*, 2003) must have been accompanied by nearly complete stress drops. More generally, one may argue that highly localized slip zones are an indicator of nearly complete stress drops for sizeable earthquakes, regardless of whether macroscopic melting took place. This stems from the fact that melting of a narrow slip interface could have been prevented only if the dynamic friction were already low (Fialko, 2004a). Theoretical and observational results of Fialko and Khazan (2005) and Di Toro *et al.* (2006) also raise a question about the mechanism of rupture arrest below the brittle–ductile transition. A currently prevailing view is that the bottom of the seismogenic layer represents a rheologic transition between the velocity-weakening and velocity-strengthening friction (e.g., Scholz, 1998; Marone, 1998). While this transition may well explain the stick-slip behavior in the brittle layer, and steady interseismic creep in the ductile substrate, it is

unclear whether the velocity strengthening is relevant for arresting of seismic ruptures that propagate into the ductile substrate from the brittle upper crust. For highly localized ruptures, the onset of melting or some other thermally activated mechanism may occur immediately behind the rupture front (aided by high ambient temperature and stress concentration), potentially overwhelming the effect of rate and state friction. In this case, the rupture arrest may require either delocalization of seismic slip near the rupture front (thereby limiting the temperature rise and increasing the effective fracture energy), or low deviatoric stress below the brittle–ductile transition. Discriminating between these possibilities may provide important insights into long-standing questions about the effective mechanical thickness and strength of tectonically active crust and lithosphere (England and Molnar, 1997; Jackson, 2002; Lamb, 2002).

4.03.8 Conclusions

Shear tractions on a slipping fault may be controlled by a multitude of physical processes that include yielding and breakdown around the propagating rupture front, rate- and slip-dependent friction, acoustic fluidization, dynamic reduction in the fault-normal stress due to dissimilar rigidities of the fault walls, thermally activated increases in the pore fluid pressure, flash heating of the asperities, and macroscopic melting. A traditional distinction between the fracture energy consumed near the rupture front and frictional losses on a well-slipped portion of a fault may be ill-defined if the slipping interface undergoes a substantial yet gradual weakening behind the rupture front. Both theoretical models and experimental data indicate that a continuous degradation of the dynamic fault strength is an expected consequence of a highly localized slip. In this case, earthquake ruptures violate a basic assumption of the LEFM, namely that the zone of the material breakdown is small compared to the rupture size. The assumption of small-scale yielding is also violated if the extent of the off-fault damage is not negligible compared to the characteristic length of a slipping region. Field observations of kilometer-wide damage zones around major crustal faults (Ambraseys, 1970; Fialko *et al.*, 2002; Chen and Freymueller, 2002; Fialko, 2004b) are evidence that the off-fault yielding may constitute a significant fraction of the earthquake energy balance (Wilson *et al.*, 2004).

Both the dynamically induced variations in shear stress along the slip interface and distributed inelastic deformation off the primary slip surface affect the distribution of slip and slip rate, dynamic stress drop, and radiated seismic energy. Inferences from seismic data about constitutive properties of the fault zone material, such as the peak breakdown stress and critical slip-weakening distance, may be biased if the above-mentioned phenomena are not taken into account. In particular, simplified rupture models that neglect off-fault damage likely overestimate the critical slip-weakening distance D_c. Similarly, models that neglect the physics of high-speed slip may attribute the dynamic (e.g., thermally induced) weakening to large '*in situ*' values of D_c interpreted as an intrinsic rock fracture property. Constitutive relationships governing evolution of the dynamic strength at seismic slip velocities are difficult to formalize, as temperature, pore fluid pressure, and shear tractions on the fault interface are sensitive to a number of parameters such as the effective thickness of the slip zone, fluid saturation, dynamic damage and changes in the fault wall permeability, and ambient stress. Many of these parameters are also poorly constrained by the available data. Nonetheless, further insights into the dynamics of earthquakes may require fully coupled elastodynamic–thermodynamic models of seismic rupture. Development of such models is a significant challenge for future work.

Acknowledgments

The author wishes to thank Jim Rice for kindly providing results of his calculations of off-fault damage (**Figure 5**). This work was supported by NSF through grants EAR-0338061 and EAR-0450035.

References

Abercrombie R (1995) Earthquake source scaling relationships from −1 to 5M_L using seismograms recorded at 2.5-km depth. *Journal of Geophysical Research* 100: 24015–24036.

Abercrombie R and Rice JR (2005) Can observations of earthquake scaling constrain slip weakening? *Geophysical Journal International* 162: 406–424.

Aki K (1967) Scaling law of seismic spectrum. *Journal of Geophysical Research* 72: 1217–1231.

Aki K and Richards PG (1980) *Quantitative Seismology: Theory and Methods, I and II*, 577 pp. San Francisco: W. H. Freeman.

Ambraseys N (1970) Some characteristic features of the North Anatolian fault zone. *Tectonophysics* 9: 143–165.

Andrews DJ (1976a) Rupture propagation with finite stress in antiplane strain. *Journal of Geophysical Research* 81: 3575–3582.

Andrews DJ (1976b) Rupture velocity of plane strain shear cracks. *Journal of Geophysical Research* 81: 5679–5687.

Andrews DJ (2002) A fault constitutive relation accounting for thermal pressurization of pore fluid. *Journal of Geophysical Research* 107: 10.1029/2002JB001,942.

Andrews DJ (2005) Rupture dynamics with energy loss outside the slip zone. *Journal of Geophysical Research* 110: B01 307, doi:10.1029/2004JB003,191.

Aochi H and Madariaga R (2003) The 1999 Izmit, Turkey, earthquake: Nonplanar fault structure, dynamic rupture process, and strong ground motion. *Bulletin of the Seismological Society of America* 93: 1249–1266.

Archuleta RJ (1984) A faulting model for the 1979 Imperial Valley earthquake. *Journal of Geophysical Research* 89: 4559–4585.

Atkinson BK (1987) Introduction to fracture mechanics. *In: Fracture Mechanics of Rock*, chap. 1, pp. 1–23. London: Academic Press.

Barenblatt GI (1959) The formation of equilibrium cracks during brittle fracture, general ideas and hypotheses, axially-symmetric cracks. *Applied Mathematics and Mechanics* 23: 622–636 (in Russian).

Barton N (1971) A relationship between joint roughness and joint shear strength. *Proc. Int. Symp. Rock Mech., Nancy,* Pap. I–8, 1971.

Bazant ZP and Planas J (1998) *Fracture and Size Effect in Concrete and Other Quasibrittle Materials*. Boca Raton, FL: CRC Press.

Ben-Zion Y (2003) Appendix 2, key formulas in earthquake seismology. In: Lee W, Kanamori H, Jennings P, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology, 81A*, pp. 1857–1875. London: Academic Press.

Ben-Zion Y and Rice JR (1997) Dynamic simulations of slip on a smooth fault in an elastic solid. *Journal of Geophysical Research* 102: 17,771–17,784.

Ben-Zion Y and Shi Z (2005) Dynamic rupture on a material interface with spontaneous generation of plastic strain in the bulk. *Earth and Planetary Science Letters* 236: 486–496.

Beroza GC and Mikumo T (1996) Short slip duration in dynamic rupture in the presence of heterogeneous fault properties. *Journal of Geophysical Research* 101: 22,449–22,460.

Beroza GC and Spudich P (1988) Linearized inversion for fault rupture. *Journal of Geophysical Research* 93: 6275–6296.

Bilby BA and Eshelby JD (1968) Dislocations and the theory of fracture. In: Liebowitz H (ed.) *Fracture, An Advanced Treatise*, vol. 1, pp. 99–182. New York: Academic Press.

Blanpied M, Tullis T, and Weeks J (1995) Frictional slip of granite at hydrothermal conditions. *Journal of Geophysical Research* 100: 13,045–13,064.

Bouchon M and Vallee M (2003) Observation of long supershear rupture during the magnitude 8.1 Kunlunshan earthquake. *Science* 301: 824–826.

Broberg K (1978) Transient sliding motion. *Geophysical Journal of the Royal Astronomical Society* 52: 397–432.

Broberg K (1987) On crack paths. *Engineering Fracture Mechanics* 28: 663–679.

Broberg K (1989) The near-tip field at high crack velocities. *International Journal of Fracture* 39: 1–13.

Broberg K (1999) *Cracks and Fracture*. London: Academic Press.

Brodsky EE and Kanamori H (2001) Elastohydrodynamic lubrication of faults. *Journal of Geophysical Research* 106: 16,357–16,374.

Brune J and Thatcher W (2003) Strength and energetics of active fault zones. In: Lee W, Kanamori H, Jennings P, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology*, 81A, pp. 569–588. London: Academic Press.

Brune JN, Henyey T, and Roy R (1969) Heat flow, stress, and rate of slip along San Andreas fault, California. *Journal of Geophysical Research* 74: 3821–4009.

Burridge R (1973) Admissible speeds for plane-strain self-similar shear cracks with friction but lacking cohesion. *Geophysical Journal of the Royal Astronomical Society* 35: 439–455.

Burridge R and Halliday G (1971) Dynamic shear cracks with friction as models for shallow focus earthquakes. *Geophysical Journal of the Royal Astronomical Society* 25: 261–283.

Byerlee J (1978) Friction of rock. *Pure and Applied Geophysics* 116: 615–626.

Cardwell R, Chinn D, Moore G, and Turcotte D (1978) Frictional heating on a fault zone with finite thickness. *Geophysical Journal of the Royal Astronomical Society* 52: 525–530.

Chen Q and Freymueller J (2002) Geodetic evidence for a near-fault compliant zone along the San Andreas fault in the San Francisco Bay area. *Bulletin of the Seismological Society of America* 92: 656–671.

Chen YT and Knopoff L (1986) Static shear crackwith a zone of slip-weakening. *Geophysical Journal of the Royal Astronomical Society* 87: 1005–1024.

Cherepanov GP (1968) Cracks in solids. *International Journal of Solids and Structures* 4: 811–831.

Cherepanov GP (1979) *Mechanics of Brittle Fracture*. New York: McGraw-Hill.

Chester JS, Chester FM, and Kronenberg AK (2005) Fracture surface energy of the Punchbowl fault, San Andreas system. *Nature* 437: 133–136.

Cowie PA and Scholz CH (1992) Physical explanation for the displacement–length relationship of faults using a post-yield fracture mechanics model. *Jounral of Structural Geology* 14: 1133–1148.

Dahlen FA (1977) The balance of energy in earthquake faulting. *Geophysical Journal of the Royal Astronomical Society* 48: 239–261.

Dahlen FA and Tromp J (1998) *Theoretical Global Seismology*, 944pp. Princeton, NJ: Princeton University Press.

Davidson C, Davis K, Bailey C, Tape C, Singleton J, and Singer B (2003) Age, origin, and significance of brittle faulting and pseudotachylyte along the coast shear zone, prince rupert, british columbia. *Geology* 31: 43–46.

Day SM (1982) Three-dimensional simulation of spontaneous rupture: The effect of nonuniform prestress. *Bulletin of the Seismological Society of America* 72: 1881–1902.

Delouis B, Giardini D, Lundgren P, and Salichon J (2002) Joint inversion of InSAR, GPS, teleseismic, and strong-motion data for the spatial and temporal distribution of earthquake slip: Application to the 1999 Izmit mainshock. *Bulletin of the Seismological Society of America* 92: 278–299.

Di Toro G, Goldsby DL, and Tullis TE (2004) Friction falls towards zero in quartz rock as slip velocity approaches seismic rates. *Nature* 427: 436–439.

Di Toro G, Hirose T, Nielsen S, Pennacchioni G, and Shimamoto T (2006) Natural and experimental evidence of melt lubrication of faults during earthquakes. *Science* 311: 647–649.

Dieterich JH (1979) Modeling of rock friction 1. Experimental results and constitutive equations. *Journal of Geophysical Research* 84: 2161–2168.

Dieterich JH (1992) Earthquake nucleation on faults with rate- and state-dependent strength. *Tectonophysics* 211: 115–134.

Dugdale DSJ (1960) Yielding of steel sheets containing slits. *Journal of Mechanics and Physics of Solids* 8: 100–115.
Dunham EM and Archuleta RJ (2004) Evidence for a supershear transient during the 2002 Denali fault earthquake. *Bulletin of the Seismological Society of America* 94: S256–S268.
England PC and Molnar P (1997) Active deformation of Asia: From kinematics to dynamics. *Science* 278: 647–650.
Eshelby JD (1956) The continuum theory of lattice defects. In: Seitz F and Turnbull D (eds.) *Progress in Solid State Physics*, pp. 79–144. New York: Academic Press.
Fialko Y (2004) Temperature fields generated by the elastodynamic propagation of shear cracks in the Earth. *Journal of Geophysical Research* 109(B01): 303 (10.1029/2003JB002,497).
Fialko Y (2004b) Probing the mechanical properties of seismically active crust with space geodesy: Study of the coseismic deformation due to the 1992 M_w7.3 Landers (southern California) earthquake. *Journal of Geophysical Research* 109(B03): 307 (10.1029/2003JB002,756).
Fialko Y and Khazan Y (2005) Fusion by earthquake fault friction: Stick or slip? *Journal of Geophysical Research* 110(B12): 407 (doi:10.1029/2005JB003,869).
Fialko Y, Sandwell D, Agnew D, Simons M, Shearer P, and Minster B (2002) Deformation on nearby faults induced by the 1999 Hector Mine earthquake. *Science* 297: 1858–1862.
Fialko Y, Rivera L, and Kanamori H (2005) Estimate of differential stress in the upper crust from variations in topography and strike along the San Andreas fault. *Geophysical Journal International* 160: 527–532.
Fialko YA and Rubin AM (1997) Numerical simulation of high pressure rock tensile fracture experiments: Evidence of an increase in fracture energy with pressure? *Journal of Geophysical Research* 102: 5231–5242.
Freund LB (1979) The mechanics of dynamic shear-crack propagation. *Journal of Geophysical Research* 84: 2199–2209.
Freund LB (1998) *Dynamic Fracture Mechanics*, 563 pp. New York: Cambridge University Press.
Gakhov FD (1966) *Boundary Value Problems*. London: Pergamon Press.
Goldsby D and Tullis T (2002) Low frictional strength of quartz rocks at subseismic slip rates. *Geophysical Research Letters* 29: 1844 (10.1029/2002GL015,240).
Griffith AA (1920) The phenomena of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London, A* 221: 163–197.
Gutenberg B and Richter C (1949) *Seismicity of the Earth and Associated Phenomena*. Princeton: Princeton University Press.
Hartzell SH and Heaton T (1983) Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America* 73: 1553–1583.
Hashida T, Oghikubo H, Takahashi H, and Shoji T (1993) Numerical-simulation with experimental-verification of the fracture-behavior in granite under confining pressures based on the tension-softening model. *International Journal of Fracture* 59: 227–244.
Heaton T (1990) Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors* 64: 1–20.
Hirose T and Shimamoto T (2003) Fractal dimension of molten surfaces as a possible parameter to infer the slip-weakening distance of faults from natural pseudotachylytes. *Journal of Structural Geology* 25: 1569–1574.
Hirose T and Shimamoto T (2005) Growth of molten zone as a mechanism of slip weakening of simulated faults in gabbro during frictional melting. *Journal of Geophysical Research* 110(B05): 202 (doi:10.1029/2004JB003,207).
Husseini MI (1977) Energy balance for formation along a fault. *Geophysical Journal of the Royal Astronomical Society* 49: 699–714.
Ida Y (1972) Cohesive force across the tip of a longitudinal-shear crack and Griffith's specific surface energy. *Journal of Geophysical Research* 77: 3796–3805.
Inglis CE (1913) Stresses in a plate due to the presence of cracks and sharp corners. *Transactions of the Institute of Naval Architect* 55: 219–241.
Ingraffea AR and Schmidt RA (1978) Experimental verification of a fracture mechanics model for tensile strength prediction of Indiana limestone. In: Kim Y-S (ed.) *Proceedings of the 19th US Symposium on Rock Mechanics,* pp. 247–253. University of Nevada (Reno).
Irwin GR (1957) Relation of stresses near a crack to the crack extension force. *Proceedings of the 9th International Congress and Applied* VIII: 245–251.
Jackson J (2002) Strength of the continental lithosphere: Time to abandon the jelly sandwich? *GSA Today* 12: 4–9.
Jeffreys H (1942) On the mechanics of faulting. *Geological Magazine* 79: 291–295.
Kame N and Yamashita T (1999) Simulation of the spontaneous growth of a dynamic crack without constraints on the crack tip path. *Geophysical Journal International* 139: 345–358.
Kanamori H (2004) The diversity of the physics of earthquakes. *Proceedings of the Japan Academy Series B* 80: 297–316.
Kanamori H and Anderson DL (1975) Theoretical basis of some empirical relations in seismology. *Bulletin of the Seismological Society of America* 65: 1073–1095.
Kanamori H and Brodsky EE (2004) The physics of earthquakes. *Reports on Progress in Physics* 67: 1429–1496.
Kanamori H and Heaton TH (2000) Microscopic and macroscopic physics of earthquakes. In: Rundle J, Turcotte DL, and Klein W (eds.) *Physics of Earthquakes, Geophysical Monograph*, 106, pp. 117–136. Washington, DC: AGU.
Khazan YM and Fialko Y (2001) Tensile and shear cracks in the Dugdale-Barenblatt model. *Geofizika Zhurnal* 23: 13–30 (in Russian).
Khazan YM and Fialko YA (1995) Fracture criteria at the tip of fluid driven cracks in the Earth. *Geophysical Research Letters* 22: 2541–2544.
Koizumi YK, Otsuki TA, and Nagahama H (2004) Frictional melting can terminate seismic slips: Experimental results of stick-slips. *Geophysical Research Letters* 31: Art. No. L21,605 (doi:10.1029/2004GL020,642).
Kolosoff GV (1909) *On an Application of Complex Function Theory to a Plane Problem of the Mathematical Theory of Elasticity.* Tartu: Yuriev University.
Kostrov BV (1970) Tectonic earthquake focal theory. *Izvestiya Akademii Nauk SSR, Earth Physics* 4: 84–101.
Kostrov BV (1974) Seismic moment and energy of earthquakes and seismic flow of rock. *Izvestiya Akademii Science, USSR, Physics of the Solid Earth (Engl. Translation)* 1: 23–40.
Kostrov BV and Das S (1988) *Principles of Earthquake Source Dynamics*, 286 pp. Cambridge: Cambridge University Press.
Lachenbruch AH (1980) Frictional heating, fluid pressure, and the resistance to fault motion. *Journal of Geophysical Research* 85: 6097–6112.
Lamb S (2002) Is it all in the crust? *Nature* 420: 130–131.
Landau LD and Lifshitz EM (1986) *Theory of Elasticity*, 187 pp. Oxford: Pergamon Press.
Lapusta N and Rice JR (2004) Earthquake sequences on rate and state faults with strong dynamic weakening. *Eos Transactions of the American Geological Union, Supplement* 84: S51B-02.
Lawn B (1993) *Fracture of Brittle Solids – Second Edition*, 378 pp. Cambridge: Cambridge University Press.
Lee TC and Delaney P (1987) Frictional heating and pore pressure rise due to fault slip. *Geophysical Journal of the Royal Astronomical Society* 88: 569–591.

Leonov MY and Panasyuk VV (1959) The development of very small cracks in a solid. *Priklodnoi Mekhanika (Applied Mechanics)* 5: 391–401 (in Ukrainian).

Li VC (1987) Mechanics of shear rupture applied to earthquake zones. In: Atkinson B (ed.) *Fracture Mechanics of Rock*, pp. 351–428. London: Academic Press.

Li Y, Aki K, Vidale J, and Alvarez M (1998) A delineation of the Nojima fault ruptured in the M7.2 Kobe, Japan, earthquake of 1995 using fault zone trapped waves. *Journal of Geophysical Research* 103(B4): 7247–7263.

Lockner DA, Byerlee J, Kuksenko V, Ponomarev A, and Sidorin A (1992) Observations of quasistatic fault growth from acoustic emissions. In: Evans B and Wong T (eds.) *Fault Mechanics and Transport Properties of Rocks*, pp. 3–31. San Diego, CA: Academic Press.

Madariaga R (1976) Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America* 66: 636–666.

Maddock R (1986) Frictional melting in landslide-generated frictionites (hyalomylonites) and fault-generated pseudotachylytes – discussion. *Tectonophysics* 128: 151–153.

Mair K and Marone C (2000) Shear heating in granular layers. *Pure and Applied Geophysics* 157: 1847–1866.

Malvern LE (1969) *Introduction to the Mechanics of a Continuum Medium*, 713 pp. Englewood Cliffs, NJ: Prentice-Hall.

Manighetti I, King G, Gaudemer Y, Scholz C, and Doubre C (2001) Slip accumulation and lateral propagation of active normal faults in Afar. *Journal of Geophysical Research* 106: 13667–13696.

Marone C (1998) Laboratory-derived friction laws and their application to seismic faulting. *Annual Reviews of the Earth and Planetary Science* 26: 643–696.

Mase CW and Smith L (1987) Effects of frictional heating on the thermal, hydrologic, and mechanical response of a fault. *Journal of Geophysical Research* 92: 6249–6272.

McKenzie D and Brune JN (1972) Melting on fault planes during large earthquakes. *Geophysical Journal of the Royal Astronomical Society* 29: 65–78.

Melin S (1986) When does a crack grow under mode II conditions. *International Journal of Fracture* 30: 103–114.

Molinari A, Estrin Y, and Mercier S (1999) Dependence of the coefficient of friction on the sliding conditions in the high velocity range. *Journal of Tribology* 121: 35–41.

Muskhelishvili NI (1953) *Some Basic Problems of Mathematical Theory of Elasticity*. Groningen, Holland: Noordhoff.

Oglesby DD and Archuleta RJ (2001) Fault geometry and the dynamics of the 1999 Chi-Chi (Taiwan) earthquake. *Bulletin of the Seismological Society of America* 91: 1099–1111.

Ohnaka M (2003) A constitutive scaling law and a unified comprehension for frictional slip failure, shear fracture of intact rock, and earthquake rupture. *Journal of Geophysical Research* 108: B2080 (doi:10.1029/2000JB000,123).

Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America* 75: 1135–1154.

Olsen K, Madariaga R, and Archuleta R (1997) Three-dimensional dynamic simulation of the 1992 Landers earthquake. *Science* 278: 834–838.

Palmer AC and Rice JR (1973) The growth of slip surfaces in the progressive failure of over-consolidated clay. *Proceedings of the Royal Society of London, Series A* 332: 527–548.

Perrin G, Rice JR, and Zheng G (1995) Self-healing slip pulse on a frictional surface. *Journal of Mechanics and Physics of Solids* 43: 1461–1495.

Peyrat S, Olsen K, and Madariaga R (2001) Dynamic modeling of the 1992 Landers earthquake. *Journal of Geophysical Research* 106: 26467–26482.

Poliakov AB, Dmowska R, and Rice JR (2002) Dynamic shear rupture interactions with fault bends and off-fault secondary faulting. *Journal of Geophysical Research* 107: 2295 (doi:10.1029/2001JB000,572).

Power WL and Tullis TE (1995) A review of the fractal character of natural fault surfaces with implications for friction and the evolution of fault zones. In: Barton CC and LaPointe PR (eds.) *Fractals in the Earth Sciences*, pp. 89–105. New York: Plenum.

Reid H (1910) The mechanism of the earthquake. In: Report of the State Earth Quake Investigation Commission (ed.) *The California Earthquake of April 19, 1906*, vol. 2. Washington, DC: Carnegie Institution.

Rice JR (1968a) Mathematical analysis in the mechanics of fracture. In: Liebowitz H (ed.) *Fracture, An Advanced Treatise*, vol. 2, pp. 191–311. New York: Academic Press.

Rice JR (1968b) A path-independent integral and the approximate analysis of strain concentration by notches and cracks. *Journal of Applied Mechanics* 35: 379–386.

Rice JR (1979) The mechanisms of quasistatic crack growth. In: Kelly RE (ed.) *Proceedings of the US National Congress and Applied Mechanics 8th, UCLA, June 1978*, pp. 191–216. North Hollywood, CA: Western Periodicals.

Rice JR (1980) The mechanics of earthquake rupture. In: Dziewonski AM and Boschi E (eds.) *Physics of the Earth's Interior*, pp. 555–649. North-Holland: Amsterdam.

Rice JR (2006) Heating and weakening of faults during earthquake slip. *Journal of Geophysical Research* B05311, doi:10.1029/2005JB00406.

Rice JR, Sammis CG, and Parsons R (2005) Off-fault secondary failure induced by a dynamic slip pulse. *Bulletin of the Seismological Society of America* 95: 109–134.

Richards P (1976) Dynamic motions near an earthquake fault – 3-dimensional solution. *Bulletin of the Seismological Society of America* 66: 1–32.

Rivera LA and Kanamori H (2005) Representations of the radiated energy in earthquakes. *Geophysical Journal International* 162: 148–155.

Rubin AM (1993) Tensile fracture of rock at high confining pressure: Implications for dike propagation. *Journal of Geophysical Research* 98: 15,919–15,935.

Rubin AM and Ampuero JP (2005) Earthquake nucleation on (aging) rate and state faults. *Journal of Geophysical Research* 110(B11): 312 (doi:10.1029/2005JB003,686).

Rudnicki JW (1980) Fracture mechanics applied to the Earth's crust. *Annual Reviews of the Earth and Planetary Science* 8: 489–525.

Rudnicki JW and Freund LB (1981) On energy radiation from seismic sources. *Bulletin of the Seismological Society of America* 71: 583–595.

Ruina A (1983) Slip instability and state variable friction laws. *Journal of Geophysical Research* 88: 10,359–10,370.

Samudrala O, Huang Y, and Rosakis AJ (2002) Subsonic and intersonic shear rupture of weak planes with a velocity weakening cohesive zone. *Journal of Geophysical Research* 107: B2170 (10.1029/2001JB000,460).

Sandwell DT and Fialko Y (2004) Warping and cracking of the Pacific Plate by thermal contraction. *Journal of Geophysical Research* 109: B10411, doi:10.1029/2004JB0030917.

Savage J (1998) Displacement field for an edge dislocation in a layered half-space. *Journal of Geophysical Research* 103: 2439–2446.

Scholz CH (1998) Earthquakes and friction laws. *Nature* 391: 37–42.

Scholz CH (2002) *The Mechanics of Earthquakes and Faulting* 2nd Ed., 496 pp. New York: Cambidge University Press.

Segall P (1991) Fault mechanics. *Reviews of Geophysics* 29: 864–876.

Segall P and Rice JR (1995) Dilatancy, compaction, and slip instability of a fluid infiltrated fault. *Journal of Geophysical Research* 100: 22,155–22,171.

Sibson R (1980) Power dissipation and stress levels on faults in the upper crust. *Journal of Geophysical Research* 85: 6239–6247.

Sibson RH (1973) Interaction between temperature and pore-fluid pressure during earthquake faulting – A mechanism for partial or total stress relief. *Nature* 243: 66–68.

Sibson RH (1975) Generation of pseudotachylite by ancient seismic faulting. *Geophysical Journal of the Royal Astronomical Society* 43: 775–794.

Simonov IV (1983) Behavior of solutions of dynamic problems in the neighborhood of the edge of a cut moving at transonic speed in the elastic medium. *Mechanics of Solids (Mekhanika Tverdogo Tela)* 18: 100–106.

Spray JG (2005) Evidence for melt lubrication during large earthquakes. *Geophysical Research Letters* 32: (doi:10.1029/2004GL022,293, Art. No L07,301).

Steketee JA (1958) Some geophysical applications of the elasticity theory of dislocations. *Canadian Journal of Physics* 36: 1168–1198.

Swanson MP (1992) Fault structure, wear mechanisms and rupture processes in pseudotachylyte generation. *Tectonophysics* 204: 223–242.

Timoshenko S and Goodier JN (1970) *Theory of Elasticity* 3rd edn., 398 pp. New York: McGraw-Hill.

Tinti E, Spudich P, and Cocco M (2005) Earthquake fracture energy inferred from kinematic rupture models on extended faults. *Journal of Geophysical Research* 110(B12): 303 (doi:10.1029/2005JB003,644).

Tsutsumi A and Shimamoto T (1997) High-velocity frictional properties of gabbro. *Geophysical Research Letters* 24: 699–702.

Vvedenskaya AV (1959) Determination of displacement fields for earthquakes by means of the dislocation theory. *Izvestiya Akademii Nauk SSR, Seriya Geophysika* 3: 277–284 (in Russian),.

Wald DJ and Heaton TH (1994) Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake. *Bulletin of the Seismological Society of America* 84: 668–691.

Wenk H-R, Johnson L, and Ratschbacher L (2000) Pseudotachylites in the Eastern Peninsular Ranges of California. *Tectonophysics* 321: 253–277.

Westergaard H (1939) Bearing pressures and cracks. *Journal of Applied Mechanics and Transactions of the ASME* 6: A49–A53.

Willis JR (1967) A comparison of the fracture criteria of Griffith and Barenblatt. *Journal of Mechanics and Physics of Solids* 15: 151–162.

Wilson B, Dewers T, Reches Z, and Brune J (2004) Particle size and energetics of gouge from earthquake rupture zones. *Nature* 434: 749–752.

Xia K, Rosakis AJ, and Kanamori H (2004) Laboratory earthquakes: The sub-Rayleigh-to-supershear rupture transition. *Science* 303: 1859–1861.

Zheng G and Rice J (1998) Conditions under which velocity-weakening friction allows a self-healing versus a cracklike mode of rupture. *Bulletin of the Seismological Society of America* 88: 1466–1483.

4.04 Applications of Rate- and State-Dependent Friction to Models of Fault Slip and Earthquake Occurrence

J. H. Dieterich, University of California, Riverside, CA, USA

4.04.1 Fault Slip Phenomena

Slip of active faults takes a variety of forms including continuous creep, episodic creep events or slow earthquakes, earthquake slip, decelerating postseismic creep, and by inference localized accelerating slip during earthquake nucleation. Any comprehensive theory of fault slip and earthquake occurrence must be based on a physical understanding of these diverse sliding phenomena. Sliding behaviors analogous to those of active faults are also seen in laboratory fault-slip experiments, where sliding processes and constitutive properties may be studied directly. The rate- and state-dependent constitutive formulation describes characteristic dependencies of friction on slip, slip rate, slip history, and normal stress history observed in laboratory studies, and it provides a unified framework for predictive modeling of the various sliding phenomena observed for faults in nature and in the laboratory. A formulation for earthquake rates, which is derived using rate- and state-dependent friction, forms the basis of an integrated approach to quantitatively model the effects of stress changes on earthquake rates. Foreshocks, aftershocks, and triggering phenomena can be modeled; and the earthquake rate formulation has been used in an inverse mode to infer stress changes from changes of seismicity rates.

4.04.2 Rate- and State-Dependent Friction

The sliding resistance of rocks approximately conforms to the Coulomb friction model, wherein the coefficient of friction μ is a constant. The nominal coefficient of friction of rocks, like other unlubricated materials, is generally about 0.6–0.7 (Byerlee, 1978; Stesky, 1978). However, laboratory studies have established that in detail the coefficient friction is not constant – there are subtle, but characteristic dependencies on slip, sliding rate, contact time,

and normal stress history (Dieterich, 1979, 1981; Ruina, 1983; Blanpied and Tullis, 1986; Tullis and Weeks, 1986; Tullis, 1988; Linker and Dieterich, 1992; Beeler *et al.*, 1994; Marone, 1998). The rate- and state-dependent constitutive formulation quantifies these frictional dependencies, providing a framework to unify observations of dynamic/static friction, displacement weakening at the onset of macroscopic slip, time-dependent healing, slip history dependence, and slip speed dependence (Ruina, 1983; Dieterich and Kilgore, 1996a). Although developed expressly for fault slip, rate-state constitutive properties appear to be characteristic of most unlubricated materials – this has led to its use in diverse applications including the analysis of friction in microelectromechanical devices (MEMS), servo-control design for precise positioning systems, failure processes in composite materials, and even the design of clamps in window frames.

Several closely related rate- and state-dependent formulations have been used to study sliding phenomena and earthquake processes (Dieterich, 1979, 1981; Ruina, 1983; Rice, 1983). The Ruina (1983) simplification of the Dieterich (1981) formulation for sliding resistance is widely used and may be written as

$$\tau = \sigma\left[\mu_0 + a\ln\left(\frac{\dot{\delta}}{\dot{\delta}^*}\right) + b\ln\left(\frac{\theta}{\theta^*}\right)\right] \qquad [1]$$

where τ and σ are shear and effective normal stress (compression positive), respectively; μ_0, a, and b are experimentally determined constants; $\dot{\delta}$ is sliding speed; θ is a state variable (discussed below) that evolves with slip- and normal stress-history; and $\dot{\delta}^*$ and θ^* are normalizing constants. The constant θ^* in eqn [1] is often replaced by $\theta^* = D_c/\dot{\delta}^*$ where D_c is a characteristic slip parameter described below. The nominal coefficient of friction, μ_0, is defined at a reference slip rate and state $\left(\dot{\delta} = \dot{\delta}^*, \theta = \theta^*\right.$ and generally has values of 0.6–0.7. For silicates at room temperature, a and b have roughly similar values in the range 0.005–0.015 (Dieterich, 1981; Ruina, 1983; Tullis and Weeks, 1986; Linker and Dieterich, 1992; Kilgore *et al.*, 1993; Marone and Kilgore, 1993; Marone, 1998).

As $\dot{\delta}$ or θ approach zero, eqn [1] yields unacceptably small (or negative) values of sliding resistance. To limit the minimum value of μ, some studies (Dieterich, 1986, 1987; Okubo and Dieterich, 1986; Okubo, 1989; Shibazaki and Iio, 2003) use a modified form of eqn [1] that adds constants to the logarithmic terms

$$\tau = \sigma\left[\mu_0 + a\ln\left(\frac{\dot{\delta}}{\dot{\delta}^*} + 1\right) + b\ln\left(\frac{\theta}{\theta^*} + 1\right)\right] \qquad [2]$$

Here μ_0 takes the value of the minimum friction. However, at present very few experimental data exist that let us assign limiting values to $\dot{\delta}^*$ or θ^*. In granitic rocks, Okubo and Dieterich (1986) analyzed recordings of slip speed and frictional strength at the tips of propagating frictional instabilities in granite and detected a lower limit for the state effect with an average value $\theta^* = 0.03$ s. However, controlled sliding experiments with granite to speeds as low as $10^{-4}\,\mu\mathrm{m\,s^{-1}}$ failed to detect a lower limit to the rate effect (Kilgore *et al.*, 1993).

The state variable θ has the dimensions of time and is interpreted to be a measure of the age of the population load-supporting contacts, which evolve with contact time, slip, and changes of normal stress (Dieterich, 1979, 1981; Dieterich and Kilgore, 1994). The connection between age of contacts and frictional strength appears to originate from localized creep, which occurs under the extreme stress conditions that exist at microcontacts. Microscopic imaging of contacts in transparent materials (including minerals, glass, and plastics) reveals that the actual area of contact increases with the logarithm of contact time (Dieterich and Kilgore 1994, 1996b). The contact creep observations are closely related to the well-known microindentation creep effect (Walker and Demer, 1964; Westbrook and Jorgensen, 1965; Scholz and Engelder, 1976; Evans and Goetz, 1979; Dieterich and Conrad, 1984; Dieterich and Kilgore, 1994, 1996b; Goldsby *et al.*, 2004).

Different empirical evolution laws for the state variable θ have been proposed (Ruina, 1983; Blanpied and Tullis, 1986; Linker and Dieterich, 1992; Perrin *et al.*, 1995). Common features of the laws are evolution of state over a characteristic slip distance D_c, and steady-state dependence of state on slip rate whereby contact population age is inversely proportional to slip speed

$$\theta_{ss} = D_c/\dot{\delta} \qquad [3]$$

The evolution law that is used in the following discussion incorporates time-dependent evolution of stationary contacts, which accounts for observed strengthening with the logarithm of time during stationary contact (Dieterich, 1972), displacement

evolution proposed by Ruina (1983), and normal stress dependence of Linker and Dieterich (1992). It is

$$d\theta = dt - \frac{\theta}{D_c}d\delta - \frac{\alpha\theta}{b\sigma}d\sigma \quad [4]$$

where, α is a constant $0 \leq \alpha \leq \mu$. Measurements using bare surfaces of Westerly granite give α in the range 0.25–0.50 (Linker and Dieterich, 1992). Characteristic slip distance D_c increases with surface roughness, gouge particle size, and gouge thickness, and generally varies from 2 to 100 μm (Dieterich, 1981; Tullis and Weeks, 1986; Dieterich and Kilgore, 1994; Marone, 1998). Under conditions of constant normal stress the evolution laws have the property that θ decreases with time at stresses above the steady-state friction, which can, under circumstances described below, lead to frictional weakening and unstable (earthquake) slip. Conversely, θ increases at stresses below the steady-state friction, which leads to strengthening of the frictional resistance. This strengthening at lower stresses provides a mechanism for recovery of fault strength following unstable slip events.

At constant normal stress the steady-state dependence of eqn [3] combined with eqn [1] gives the following relation for the coefficient of friction at steady state

$$\mu_{ss} = \mu_0 + (a-b)\ln\left(\frac{\dot{\delta}}{\dot{\delta}^*}\right) \quad [5]$$

Equation [5] employs the definition $\theta^* = D_c/\dot{\delta}^*$. With this substitution μ_0 now represents the steady-state friction at slip, the reference speed $\dot{\delta}^*$. Note that if $a > b$ then friction increases with increasing slip speed (rate strengthening), and if $a < b$ then friction decreases with increasing slip speed (rate weakening).

Interplay between the rate-dependent and state-dependent effects and the evolution of state over the distance D_c are revealed in the characteristic response of the coefficient of friction to rapid step-like transitions from one sliding speed to another (**Figure 1**). In **Figure 1** the top curve gives the idealized response from the rate- and state-dependent constitutive formulation and the lower curves give the observed response for a variety of materials. Step changes of sliding speed, under conditions of constant normal stress, result in immediate jumps in frictional resistance followed by displacement-dependent decay and stabilization at a new steady-state sliding friction. In rocks, this characteristic rate- and state-dependent frictional behavior has been documented for bare surfaces and simulated fault surfaces separated by a layer of gouge under a wide range of conditions accessible in laboratory experiments. These include wet and dry conditions, the range of temperatures and pressures characteristic of crustal earthquakes, and sliding rates from mm yr^{-1} to mm s^{-1} (Dieterich, 1979, 1981; Ruina, 1983; Weeks and Tullis, 1985; Blanpied and Tullis, 1986; Tullis and Weeks, 1986; Tullis, 1988; Blanpied *et al.*, 1991; Linker and Dieterich, 1992; Marone and Kilgore, 1993; Kilgore, *et al.*, 1993; Marone, 1998).

Rate- and state-dependent effects described here are largely based on laboratory studies conducted at slow slip speeds. Hence, they are clearly appropriate for modeling fault creep processes and earthquake nucleation. Experimental data for friction during stick-slip events with small displacements (Okubo and Dieterich, 1986) and plate impacts with very high slips speeds (tens of meters per second) (Prakash, 1998) indicate rate-state behaviors also operate at seismic slip speeds. Models of the earthquake cycle that employ rate-state effects during earthquake slip (Tse and Rice, 1986; Stuart, 1988; Okubo, 1989; Kato and Hirasawa, 1997; Lapusta *et al.*, 2000; Lapusta and Rice, 2003) yield results that are consistent with a broad scope of observations. However, the conditions and processes that control sliding resistance during the large displacements and high slip speeds of large earthquakes remain uncertain. Processes that may significantly alter, or replace, the rate- and state-dependent effects include thermal interactions due to shear heating and pore-fluid effects (Sleep, 1995a, 1995b, 1997; Sleep and Blanpied, 1992; Segall and Rice, 1995; Rice, 2006). These and other processes, together with fault constitutive properties for slip in large earthquakes, are the subject of Chapter 4.05.

4.04.3 Models of Sliding Phenomena

4.04.3.1 Stability of Slip

Sliding instabilities are a common, but generally unwelcome, phenomenon in a variety of mechanical systems that vary in scale from microns in MEMS devices, to a thousand kilometers or more with earthquake faults. The conditions required for unstable slip with rate- and state-dependent friction have been extensively analyzed for a single degree of freedom system consisting of a slider connected to a

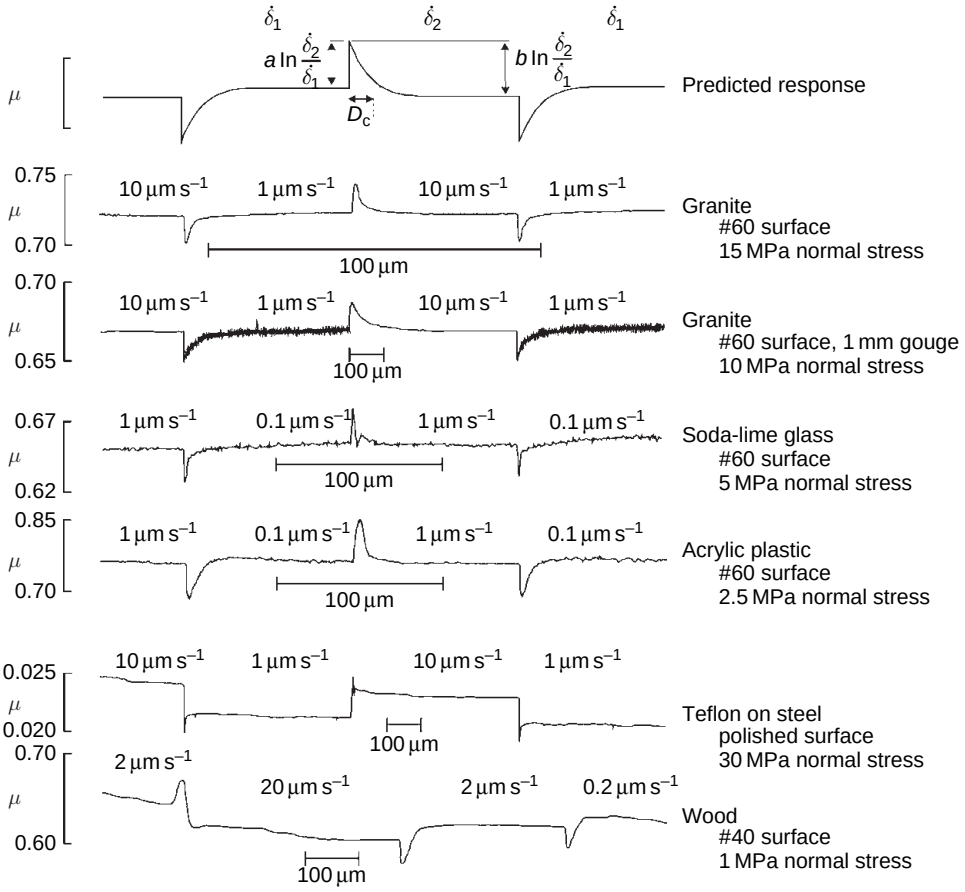


Figure 1 Characteristic rate-state frictional response to step changes of slip speed. Remarkably similar behavior is seen for a wide variety of materials. Top curve indicates the predicted rate-state response. D_c is observed to depend on surface roughness and characteristics of gouge separating the surfaces (Dieterich and Kilgore, 1996b).

loading point through a spring of stiffness K (Rice and Ruina, 1983; Ruina, 1983; Gu *et al.*, 1984; Weeks and Tullis, 1985; Blanpied and Tullis, 1986; Dieterich and Linker, 1992). Here stiffness is defined as the change of stress at the slider divided by change in spring elongation. For sliders at constant normal stress, those studies establish that unstable slip occurs only when the system stiffness is less than the critical value

$$K_c = \frac{\xi\sigma}{D_c} \qquad [6]$$

where ξ is a parameter that depends on loading conditions and constitutive parameters. For perturbations from slow steady-state slip at constant normal stress, $\xi = b - a$ (Rice and Ruina, 1983, Ruina, 1983). This result is independent of the evolution law used. The terms K_c, σ, and D_c, are positive – hence, an added requirement for unstable slip is that ξ must also be positive (i.e., $b > a$). This is readily understood from eqn [5], which shows that steady-state frictional resistance decreases with increasing slip rate if $b > a$. Conversely the strengthening that occurs with increasing slip speed if $b < a$ will act to prevent acceleration of slip leading to instability. Laboratory studies confirm the association of stick-slip with rate weakening, and additionally verify predictions of the critical stiffness for instability, including the dependence on σ and D_c. Stable slip always occurs if $b > a$, or if $K > K_c$ independent of the relative values of a and b.

In many circumstances, including slip of nonplanar faults, and nonvertical faults that interact with the free surface, both normal and shear components of stress couple to slip. The Ruina stability analysis has been generalized to include cases where both shear and normal stresses are coupled to slip through a spring inclined at an angle to the sliding surface

(Dieterich and Linker, 1992). This leads to two stability criteria. The first is a generalized form of the Ruina and Rice results, in which the term ξ in the critical stiffness equation is

$$\xi = \frac{b-a}{1+(\mu_{ss}-\alpha)\tan\phi}, \quad \left(\text{for } \phi > -\cot^{-1}\mu_{ss}\right) \quad [7]$$

where α is the normal stress parameter in the evolution eqn [4], and ϕ is the angle between the spring and the sliding surface (taken as positive if slip increases the compressive stress. For the case $\phi = 0$, normal stress remains constant during slip. It is seen from this result that conditions in which normal stress decreases with increasing slip (negative spring angles) are less stable than situations in which normal stress increases with slip; that is, negative spring angles result in larger ξ than positive spring angles and therefore require a stiffer spring to stabilize slip. However, the requirement of $b > a$ for instability remains valid (for $\phi > -\cot^{-1}(\mu_{ss})$). The second stability condition arises for large negative spring angles. If $\phi < -\cot^{-1}(\mu_{ss})$ then either slider lock-up, or instability occur. The second condition is independent of the system stiffness or specific constitutive properties (i.e., b may be less than a).

4.04.3.2 Depth Dependence of Constitutive Parameters

Laboratory data on steady-state friction at different temperatures (Stesky, 1978, Blanpied *et al.*, 1991, 1995) indicate that rocks with rate-weakening friction at low temperatures undergo a transition to rate-strengthening friction above temperatures of about 300–350°C (**Figure 2(a)**). Tse and Rice (1986) note that the condition for slip instability, $b > a$ given in eqn [6], together with the temperature dependence of steady-state friction, are consistent with the observed depth ranges of earthquakes along plate boundaries; that is, the predicted depth of the isotherm for the transition from unstable slip to fault creep, based on thermal gradients at plate boundaries, corresponds to the observed maximum depth of earthquakes. These observations indicate a model for the depth dependence of a and b (**Figure 2(b)**) that is consistent with the depth range of seismicity, fault stability criteria, and predictions of constitutive properties based on crustal temperatures. The temperature-dependent model for constitutive parameters has been widely adopted for modeling the earthquake cycle and creep processes on strike-slip faults and in subduction zones

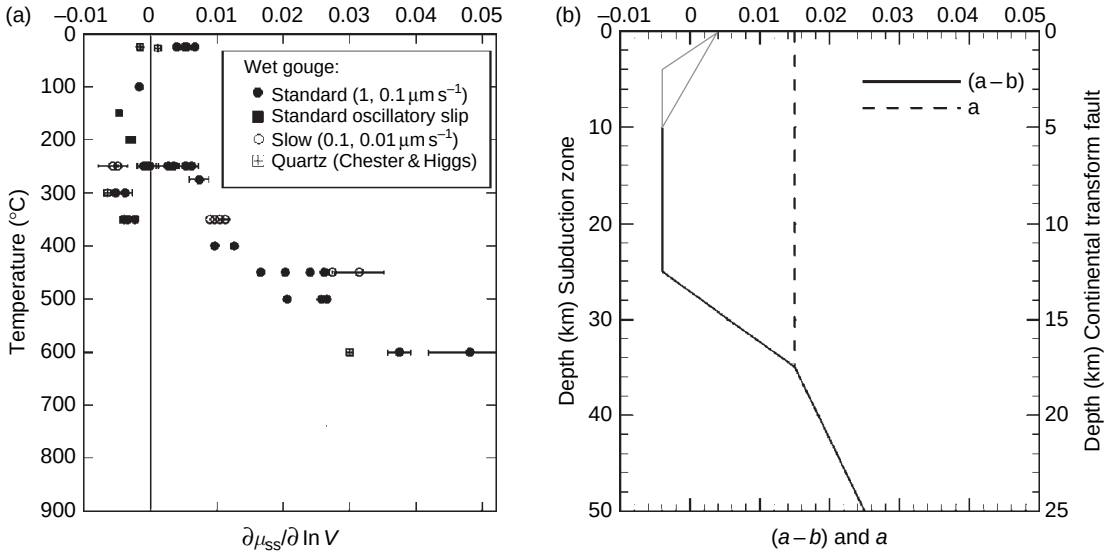


Figure 2 Depth distribution of friction parameters. (a) Experimental data on steady-state friction under hydrothermal conditions from Blanpied *et al*. (1995). (b) Depth distribution of $(a - b)$ and a based on thermal structure of continental transform faults and subduction zones. The shallow rate-strengthening zone may be of variable thickness depending on the degree of consolidation of sediments. Subduction of cold sediments may produce zones of rate-weakening at greater depths than generally found along strike-slip faults. Adapted from Liu and Rice (2005) and Lapusta N and Rice JR (2003) Nucleation and early seismic propagation of small and large events in a crustal earthquake model. *Journal of Geophysical Research* 108(B4): 2205 (doi:10.1029/2001JB000793).

(Tse and Rice, 1986; Stuart, 1988; Kato and Hirasawa, 1997; Lapusta *et al.*, 2000; Lapusta and Rice, 2003). Shallow portions of faults with rate strengthening are sometimes invoked to represent poorly consolidated sediments in the shallow portions of subducting slabs and crustal strike-slip faults. This association is based on observations of rate strengthening at low normal stresses in poorly consolidated layers of gouge (Marone *et al.*, 1990, Marone, 1998). Such a shallow zone is consistent with observations of shallow fault creep, which tends to correlate with poorly consolidated sediments.

4.04.3.3 Minimum Fault Dimension for Unstable Slip

An approximate, but useful representation of effective stiffness for slip of a fault patch, obtained from the elasticity solutions for slip of a crack embedded in an elastic media, takes the form

$$K = \frac{\Delta\tau}{\delta} = \frac{G\eta}{L} \qquad [8]$$

where L is the length (or diameter) of the fault patch, G is the shear modulus (Poisson's ratio taken to be 0.25), and η is a factor with value near 1 that depends upon the geometry of the slip patch and slip or stress conditions on the patch. **Table 1** gives values of η for some simple patch geometries. Combining eqn [8] with the critical stiffness relation for instability (eqn [6]) gives

$$L_c = \frac{G\eta D_c}{\xi\sigma} \qquad [9]$$

where L_c is the patch length corresponding to K_c (Dieterich, 1986, 1992). Slip of fault patches with dimensions less than L_c will have an effective stiffness in excess of K_c and sliding will be stable.

Assuming the dimensions of an earthquake rupture must be equal to, or greater than, the size of the patch that nucleates the instability, L_c defines a lower bound for the dimensions of the earthquake source. Equation [9] has been approximately verified by experiments with a large-scale testing apparatus that accommodates samples with a 2 m fault – a size that is sufficient to permit confined sliding instabilities that do not reach to the ends of the fault (Dieterich and Kilgore, 1996a).

A primary uncertainty in applying eqn [9] to faults in nature relates to poor constraints on the parameter D_c, which continues to be an open question. In the laboratory, D_c increases with surface roughness, gouge particle size, and possibly gouge thickness. Compared to natural faults, laboratory experiments employ exceptionally smooth faults with thin layers of fine gouge. It seems plausible, therefore, that natural faults may have much larger values of D_c than laboratory faults, which could substantially increase estimates of the minimum dimensions of earthquakes that originate as frictional instabilities. However, the existence of very small earthquakes with source dimensions of 5 m or less (Abercrombie, 1995) require small nucleation sources and values of D_c, which may be similar to those measured in laboratory experiments. For example using an estimate of D_c at the lower end of laboratory values (2 μm) gives $L_c \sim 5\text{cm}$ ($G = 18\,000$ MPa, $\xi = 0.004$, $\eta = 1$, with effective $\sigma = 100$ MPa). Increasing D_c to the upper range of laboratory values (100 μm) gives $L_c \sim 2$ m.

4.04.3.4 Creep Events and Silent Earthquakes

Aseismic creep of active faults is observed in a variety of tectonic settings where it takes the form of either continuous creep or creep events. Continuous creep at more or less steady rates is readily understood in

Table 1 Value of parameter η for different patch geometries[a]

η	L	Patch	Reference
$7\pi/12$	Diameter	Circular patch	Eshelby (1957)
4/3	Length	Plane strain, constant shear stress	Star (1928)
$8/3\pi$	Length	Plane strain, constant slip	Chinnery (1969)
1	Length (vertical)	Anti-plane strain, constant shear stress	Knopoff (1958)
$2/\pi$	Length (vertical)	Anti-plane strain, constant slip	Chinnerry (1969)

[a]Assumes Poisson's ration $\nu = 0.25$, slip δ is measured at center of patch.

terms of the stability relations discussed in Section 4.04.3.3, which show that steady creep will occur when $a > b$. However, the origins of creep events are more problematic.

Creep events were first observed with surface creepmeters along portions of the creeping section of the San Andreas Fault in central California (Nason and Weertman, 1973; Wesson, 1988). The events there are periodic, have durations that vary from an hour to a few days, and have displacements of up to a few tens of millimeters per slip event. Analyses of strain signals associated with these events indicate that the sources are shallow, extending to maximum depths of a few kilometers (McHugh and Johnston, 1976; Langbein, 1981; Gladwin *et al.*, 1994).

Worldwide improvements in monitoring capabilities have revealed widespread occurrence of deep creep events. The events span a very large range of sizes (equivalent to moment magnitudes up to $M \sim 7$) and have durations varying from rapid short-term slip that is transitional with earthquake slip, to very slow events that continue more than a year. The terms slow earthquake and silent earthquake are sometimes used to refer to short-term and long-term events, respectively. Creep events of various durations have been reported in widespread locations and geologic settings including Japan (Sacks *et al.*, 1981; Fujii, 1993; Hirose *et al.*, 1999; Kawasaki, 2004; Ozawa *et al.*, 2002; Linde and Sacks, 2002; Sagiya, 2004; Kawasaki, 2004), California (Linde and Sacks, 2002; Kanamori and Hauksson, 1992), Alaska (Freymueller, *et al.*, 2002), Cascadia (Dragert *et al.*, 2001; Miller *et al.*, 2002; Rogers and Dragert, 2003), Italy (Crescentini *et al.*, 1999), Mexico (Lowry *et al.*, 2001; Kostoglodov *et al.*, 2003), Hawaii at Kilauea Volcano (Cervelli *et al.*, 2002; Brooks *et al.*, 2006), and oceanic transform faults (Beroza and Jordan (1990). Repeating slip events have been reported for Cascadia and Hawaii. The Hawaiian events appear to be highly periodic (Miller *et al.*, 2002; Brooks *et al.*, 2006).

The continuous spectrum of event durations from minutes to years, suggests a common origin for creep events. In very general terms, the characteristics of creep events resemble those of an earthquake cycle in that both have long intervals in which the fault is essentially locked (at least to the resolution of deformation observations), followed by an interval of rapid slip. From a mechanical perspective, creep events require a weakening mechanism to enable rapidly accelerating slip, and some type of damping mechanism to quench the acceleration of slip before it reaches seismic slip speeds. In the case of earthquakes, acceleration of slip is made possible by slip-rate weakening ($a < b$), and slip speed is limited by inertial damping from the generation of seismic waves.

Oscillatory creep is observed in laboratory experiments and in spring-slider simulations near the stable–unstable slip transition of eqn [6]. Near this boundary, small decreases of effective normal stress can move a system from conditionally stable to unstable state. Kodaira *et al.* (2004) have identified a region of high pore-fluid pressure, a subduction zone where silent slip occurs in the Tokai (Japan) region. They propose that high pore-fluid pressures extend the region of conditionally stable slip to greater depths, which promotes creep events over large regions. However, the necessary conditions for oscillations and the amplitudes of the fluctuations of slip speeds are very restricted in single-degree-of-freedom sliders.

Recent 3-D simulations of earthquakes and creep in subduction zones were found to spontaneously generate larger slip pulses in the region of transition from seismic to aseismic slip when mild heterogeneities are introduced (Liu and Rice, 2005). The simulated creep events have some characteristics of observed creep events. However, in their current form these simulations do not produce the highly periodic events with relatively short periods seen in some areas, nor do they address the small shallow periodic surface creep events seen in the creeping section of the San Andreas Fault system. It is not yet clear if this approach can reproduce the spectrum of slip rates and slip amplitudes evident in natural systems.

Shibazaki and Iio (2003) propose a model in which creep events arise spontaneously from the modified form of the rate-state formulation given by eqn [2]. In particular, the use of a lower cutoff for θ together with $b > a$ produces a transition from steady-state rate weakening to rate strengthening at $\theta^* = \theta_{ss} = D_c/\dot{\delta}$. At slow speeds rate weakening prevails, which provides the necessary mechanism for acceleration of slip, and at higher speeds the fault experiences rate strengthening, which provides the necessary mechanism to halt the acceleration before seismic slip speeds are reached. Characteristics of creep events with this mechanism are readily illustrated with a spring-slider model using a rate-state law with a single cutoff in the state term

$$\tau = \sigma\left[\mu_0 + a\ln\left(\frac{\dot{\delta}}{\dot{\delta}^*}\right) + b\ln\left(\frac{\theta}{\theta^*} + 1\right)\right] \qquad [10]$$

Figure 3 illustrates steady-state friction from eqn [10], and the oscillatory response of a slider with these properties. This simple version of the Shibazaki and Lio (2003) model produces highly periodic creep events. By changing θ^* the duration and amplitude of the creep events is continuously adjustable from slow low-amplitude events, through rapid, events, to fully seismic slip. However, at present there is no direct experimental data to either justify or discount the use cutoffs at the values required to produce oscillations resembling those in nature. Most experiments to evaluate rate-state friction effects have used unaltered quartzo-feldspathic materials that give relatively strong rate weakening at slow slip speeds and room temperature. There are few data for either the highly altered gouges and sediments that are likely to occur in zones of shallow creep, or at high temperatures that characterize deep silent earthquakes.

4.04.3.5 Earthquake Afterslip

Earthquake afterslip is an episode of stable slip that characteristically increases with the logarithm

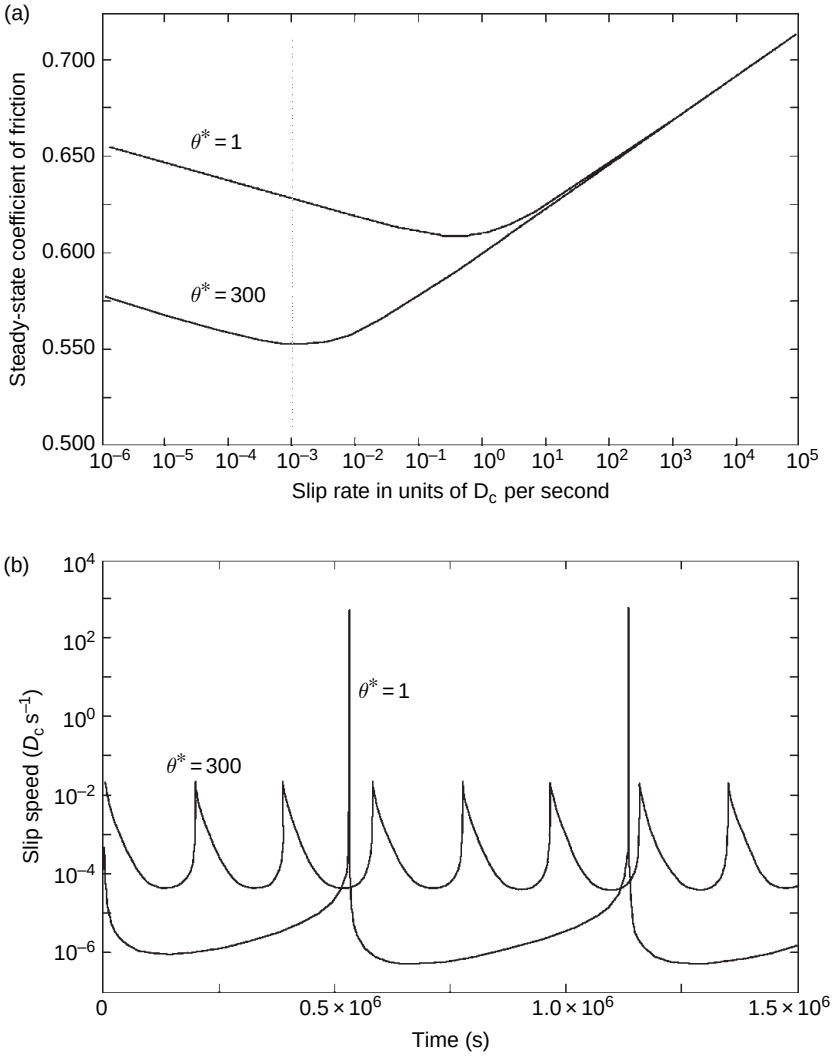


Figure 3 Spring-slider simulation of creep events. (a) Steady-state friction as given by eqn [10], which produces a minimum in friction around $\dot{\delta} = D_c/\theta^*$. Dotted line is load point displacement rate. Units of θ^* are in seconds. (b) The minimum in friction produces oscillatory slip, which increases in amplitude and period as the minimum friction moves to increasing higher speeds.

of time following an earthquake. It was originally recognized through direct measurements of fault offset at the Earth's surface, and it generally occurs in regions where thick poorly consolidated sediments lie above seismogenic faults (Smith and Wyss, 1968; Burford, 1976; Bucknam *et al.*, 1978). In addition, deep afterslip is inferred from postseismic strain observations (Pollitz *et al.*, 1998; Bürgmann *et al.*, 2002; Hsu *et al.*, 2002; Hutton *et al.*, 2002; Miyazaki *et al.*, 2004). Both shallow and deep afterslip are well represented by models that incorporate rate- and state-dependent friction. However, surface deformation observations may be insufficient to differentiate deep afterslip from bulk viscoelastic flow in some circumstances.

In general terms significant afterslip occurs whenever earthquake slip increases the stresses acting on an adjacent fault segment that has rate-strengthening characteristics ($a > b$). Models of the seismic cycle on strike-slip faults and subduction zones that employ a deep zone with rate-strengthening friction (**Figure 2**) spontaneously generate afterslip following earthquake slip events (Tse and Rice, 1986; Stuart, 1988; Kato and Hirasawa, 1997; Lapusta *et al.*, 2000; Lapusta and Rice, 2003). Those studies show that earthquake slip is strongly impeded when it encounters regions where $b > a$, which results in large stress increases and accelerated post-earthquake creep rates. Creep rates decay as the afterslip relaxes the stresses.

Marone *et al.* (1991) and Wennerberg and Sharp (1997) model surface afterslip as a single degree of freedom slider. In the case of shallow afterslip, Marone *et al.* (1991) point out that poorly consolidated fault gouges have rate-strengthening characteristics. Their models fit afterslip observations with considerable fidelity and incorporate normal stress and stiffness parameters scaled to the depth of afterslip in natural fault systems. Because the seismogenic portion of the fault is essentially locked following an earthquake, the slider represents the afterslip zone, and it is assigned rate-strengthening friction. The spring stiffness scales by eqn [6], where L is the thickness of the afterslip zone. Earthquake slip distorts the spring and loads the slider, which responds by slipping at a faster rate. Both studies assume a quasi steady-state condition, which assumes slip is always at steady-state. This appears to be a reasonable approximation because slip speed changes slowly and slip is large (centimeters to meters) compared to laboratory values of D_c (2–100 μm). Earthquake slip (spring load point displacement) has a finite duration and causes some coseismic slip of the slider, which is readily calculated. This fixes the stress and initial afterslip speed. Marone *et al.* (1991) give the following solution for afterslip with a spring and slider:

$$\delta = \frac{\sigma(a-b)}{K}\ln\left[\left(\frac{K\dot{\delta}_{CS}}{\sigma(a-b)}\right)t+1\right]+\dot{\delta}_0 t \qquad [11]$$

where t is time from the end of the earthquake, $\dot{\delta}_{CS}$ is the thickness-averaged slip speed in the rate-strengthening area at $t = 0$, and $\dot{\delta}_0$ is the long-term creep rate in the rate-strengthening area.

4.04.4 Earthquake Nucleation

4.04.4.1 Numerical Simulations of Spontaneous Earthquake Nucleation

The events and interactions that lead to the initiation of unstable fault slip are collectively defined as earthquake nucleation (Dieterich, 1987, 1992). Characterization of the earthquake nucleation process is of considerable interest because nucleation establishes the time and place of occurrence of earthquakes, which in turn determine their space-time patterns.

Dieterich (1992) and Rubin and Ampuero (2005) examined models with heterogeneous initial conditions, uniform loading, and uniform constitutive properties, and Lapusta and Rice (2003) modeled nucleation at the base of a seismogenic zone where there are strong stressing gradients and spatial variations of constitutive properties. Each of the studies obtains qualitatively similar results, though details vary as a consequence of differences in model configurations and initial conditions. **Figure 4** illustrates some characteristic features of the nucleation process. Nucleation arises spontaneously and consists of an extended interval of stable accelerating slip, during which slip speed increases by $1/t$, where t is the time remaining to instabililty. As nucleation proceeds, it rapidly localizes to a portion of the fault, and becomes increasingly insensitive to external stressing. Dieterich (1992) and Rubin and Ampuero (2005) find a region of model parameter space where accelerating slip remains localized to a zone of fixed length L_c, which scales by eqn [9] with $\xi \approx 0.4b$. In simulations where slip is forced to begin over a region $L > L_c$, the region of most active slip contracts to L_c as the time of instability approaches. In addition, there is a region of parameter space where the initial

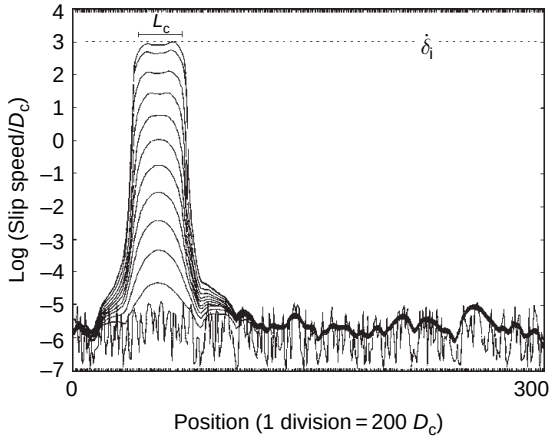


Figure 4 Model of spontaneous earthquake nucleation on a fault with rate- and state-dependent friction (Dieterich, 1992). The model was initialized with random normal stress and uniform shear stress. Growth of slip speeds favors wavelengths around L_c.

localization of accelerating slip is followed by expansion of the nucleation zone prior to unstable slip. This type of nucleation is characteristic of the Lapusta and Rice (2003) models. Dieterich (1992) and Rubin and Ampuero (2005) obtain different results on the extent of the parameter space where final expansion occurs, most likely as a consequence of differences in the way the computations are initiated.

4.04.4.2 Nucleation Solutions for Single Degree of Freedom Slider

Analytic solutions for a simplified model, which represents the nucleation source as a fault patch of fixed length are given by Dieterich (1992). The localization process observed in the numerical models allows use of eqn [8] to represent slip of the patch as a single-degree-of-freedom slider with spring stiffness K. The model pertains to conditions that operate once nucleation is in progress ($\theta >> D_c/\dot{\delta}$). Although several approximations are employed, the solutions are in good agreement with both detailed 2-D simulations, and laboratory observations of the nucleation process on a 2 m fault (**Figure 5**). These solutions form the basis of the earthquake rate formulation discussed in Section 4.04.5 and are useful for understanding earthquake-triggering processes.

In the following, the Dieterich (1992) solutions have been generalized using the evolution eqn [4] for conditions in which normal stress is not constant. A general expression for acceleration of slip during nucleation is given by

$$d\omega = \frac{-1}{a\sigma}\left[b\,dt + \omega\,d\tau + \left(\frac{\tau}{\sigma} - \alpha\right)\omega\,d\sigma\right] \qquad [12]$$

where $\omega = 1/\dot{\delta}$ and b is a term containing model and constitutive parameters

$$b = -K + \frac{b\sigma}{D_c} \qquad [13]$$

Assuming the terms $a\sigma$ and $b\sigma$ are constant (small $\Delta\sigma/\sigma$) and taking $\mu' = \tau/\sigma - \alpha$ to be an effective constant coefficient of friction during nucleation, eqn [12] reduces to

$$d\omega = \frac{-1}{a\sigma}[b\,dt + \omega\,dS] \qquad [14]$$

where S is a modified Coulomb stress function

$$S = \tau - \mu'\sigma \qquad [15]$$

Some solutions of eqn [14] are listed in **Table 2**. Solutions in terms of shear stress, with constant σ are identical to those in **Table 2** with the replacement ($\tau = S$). The solutions in **Table 2** require specification of slip speed as an initial condition. Hence, the response of a nucleation source may be followed for a complex stressing history made up of any number of stressing ramps and stress steps, by using the slip-rate solutions to track the initial conditions (slip speed) through each time/stress step.

4.04.5 Seismicity Rate Models

4.04.5.1 Earthquake Rate Formulation

From eqn [1] it is evident that changes of stress result in changes of slip speed. The nonlinearity of this effect, together with the dependence of nucleation time on slip speed (eqns [18] and [21]), indicate that earthquake rates will be sensitive to stress changes. Effects of stress changes on earthquake rates have been modeled by treating seismicity as a sequence of nucleation events in which the stressing history controls the timing of earthquakes (Dieterich, 1979, 1994). This conceptual approach has been implemented with numerical models that follow the evolution of single degree of freedom sliders to the point of

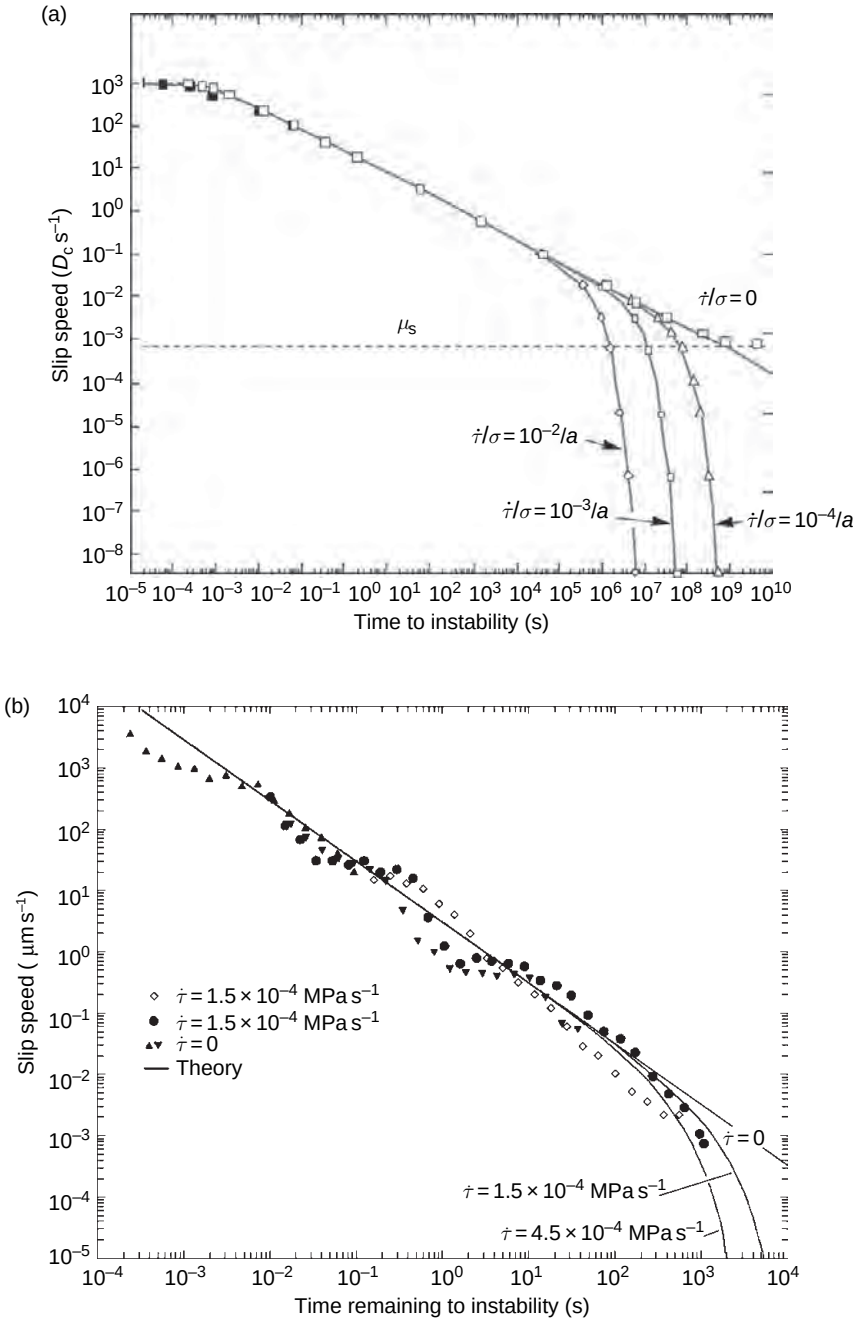


Figure 5 Time to nucleate vs slip speed. (a) Comparison of numerical calculations with 2-D boundary element model illustrated in **Figure 4**, (data points) with analytic solutions for spring and slider (solid lines) as given by eqns [17] and (20) (Dieterich, 1992). (b) Comparison of experimental results for nucleation on 2 m laboratory fault (data points) with spring-slider solutions (Dieterich and Kilgore, 1996a). The triangle represents experiments in which external stressing was stopped when premonitory displacements were detected, demonstrating the self-driven nature of the nucleation process.

instability, using perturbed and unperturbed stressing, to explicitly determine the effects of stress steps, earth tides, and seismic transients on the discrete nucleation sources (Dieterich, 1979, 1987; Gomberg *et al.*, 1997, 1998, 2000; Belardinellie *et al.*, 2003).

Table 2 Nucleation solutions[a] (eqn [14])

	$\dot{S} = \text{const} \neq 0$		$\dot{S} = 0$	
Slip	$\delta = \frac{-a}{h}\ln\left\{\frac{\dot{\delta}_0 h\sigma}{\dot{S}}\left[1-\exp\left(\frac{\dot{S}t}{a\sigma}\right)\right]+1\right\}$	[16]	$\delta = \frac{-a}{h}\ln\left\{1-\frac{\dot{\delta}_0 ht}{a}\right\}$	[19]
Slip speed	$\dot{\delta} = \left\{\left[\frac{1}{\dot{\delta}_0}+\frac{h\sigma}{\dot{S}}\right]\left[\exp\left(\frac{-\dot{S}t}{a\sigma}\right)\right]-\frac{h\sigma}{\dot{S}}\right\}^{-1}$	[17]	$\dot{\delta} = \left\{\frac{1}{\dot{\delta}_0}-\frac{-ht}{a}\right\}^{-1}$	[20]
Time to instability[b]	$t_{EQ} = \frac{a\sigma}{\dot{S}}\ln\left[\frac{\dot{S}}{h\sigma\dot{\delta}_0}+1\right]$	[18]	$t_{EQ} = \frac{a}{h}\left[\frac{1}{\dot{\delta}_0}\right]$	[21]

[a] These solutions are subject to the requirement that evolution of state does not reach the steady condition, ($\theta_{ss} = D_c/\dot{\delta}$) prior to the onset of seismic slip (see Dieterich (1992) for discussion). Initial condition is slip speed $\dot{\delta}_0$ at $t = 0$.
[b] Because intertial effects are ignored in this simplified model, instability is defined as the time were $\dot{\delta}_0 \to \infty$. Time to instability t_{EQ} is the time remaining to instability given δ_0.

Following a similar conceptual approach, a general analytic formulation for earthquake rates has been derived using the solutions for earthquake nucleation in **Table 2** (Dieterich, 1994). This formulation avoids the necessity of following individual nucleation sources, and it is valid for stressing histories of arbitrary complexity. Earthquake rate R in some magnitude interval is given by

$$R = \frac{r}{\gamma\dot{\tau}_r} \quad [22]$$

$$d\gamma = \frac{1}{a\sigma}\left[dt - \gamma\, d\tau - \gamma\left(\frac{\tau}{\sigma}-\alpha\right)d\sigma\right] \quad [23]$$

where r is a constant steady-state background rate at the reference shear stressing rate $\dot{\tau}_r$, and γ is a variable that evolves with time and stress. Note that constitutive parameters b and D_c that appear in the nucleation solutions cancel out of the rate formulation. Assuming small changes of stress ($a\sigma$ and τ/σ constant), eqns [22] and [23] can be written using the modified Coulomb stress function of eqn [15] (Dieterich *et al.*, 2000, 2003).

$$R = \frac{r}{\gamma\dot{S}_r} \quad [24]$$

$$d\gamma = \frac{1}{a\sigma}[dt - \gamma\, dS] \quad [25]$$

For positive shear stressing rates with $d\sigma = 0$, eqn [23] has the property that γ seeks the steady-state value, $\gamma_{ss} = 1/\dot{\tau}$ with the characteristic relaxation time of $t_a = a\sigma/\dot{\tau}$. Similarly in terms of Coulomb stress $\gamma_{ss} = 1/\dot{S}$ and $t_a = a\sigma/\dot{S}$.

In forward models of earthquake rates, one must find γ as a function of time for a specified stressing history. Earthquake rates as a function of time are then directly obtained from eqns [22] or [24]. Solutions for $\gamma(t)$ as a function of time may be found through numerical evaluation of the evolution equations. Alternatively, any complex stressing history may be reduced to a series of small time increments with constant stressing rates and stress steps. This permits use of analytic solutions of the evolution equations (**Table 3**) to follow the evolution of γ through the sequence of stress/time steps.

4.04.5.2 Aftershocks

4.04.5.2.1 Aftershock-rate equations

Dieterich (1986, 1992) propose that aftershocks are caused by the step-like changes of stress that occur at the time of a mainshock, which perturbs earthquake nucleation times. Two general observations support this model. First, aftershocks cluster in those areas near a mainshock where stress changes favor fault slip in aftershocks (Reasenberg and Simpson, 1992; King *et al.*, 1994; Harrts *et al.*, 1995; Toda *et al.*, 1998; Stein, 1999). Second, rapid increases of stress unrelated to an earthquake mainshock also stimulate bursts of seismicity that obey the $1/t$ Omori aftershock decay law. Examples of the latter are given by the aftershock-like earthquake sequences at Kilauea Volcano that begin at the time of magmatic intrusion events (Dvorak *et al.*, 1986).

Solutions of the earthquake-rate equations as a function of time following a step of Coulomb stress

Table 3 Solutions of evolution equations for γ[a]

	Equation [23]	Equation [25][b]
Evolution with time, fixed stress	$\gamma = \gamma_0 + \frac{t}{a\sigma}$	$\gamma = \gamma_0 + \frac{t}{a\sigma}$ [26]
$\dot{\tau} = \text{const} \neq 0 \quad \dot{S} = \text{const} \neq 0$	$\gamma = \left(\gamma_0 - \frac{1}{\dot{\tau}}\right)\exp\left(\frac{-t}{t_a}\right) + \frac{1}{\dot{\tau}}$ [27]	$\gamma = \left(\gamma_0 - \frac{1}{\dot{S}}\right)\exp\left(\frac{-t}{t_a}\right) + \frac{1}{\dot{S}}$ [29]
Stress step	$\gamma = \gamma_0\left(\frac{\sigma}{\sigma_0}\right)\exp\left(\frac{\tau_0}{a\sigma_0} - \frac{\tau}{a\sigma}\right)$ [28]	$\gamma = \gamma_0 \exp\left(\frac{S_0 - S}{a\sigma}\right)$ [30]
log t stressing[c] $S = S_0 + u\ln(wt+1)$	$\gamma = \gamma_0(wt+1)^{-m} + \frac{(wt+1) - (wt+1)^{-m}}{a\sigma w(m+1)}$, where $m = \frac{u}{a\sigma}$ [31]	

[a]Parameter values at the start of time/stress increment are indicated by the subscript values 0. At he start of a time increment $t = 0$.
[b]Solutions in terms of τ are at constant normal stress σ.
[c]Identical solution for γ is obtained for change of τ by log of time with constant normal stress.

Table 4 Earthquake rates following a sress step[a]

	Solution in terms of S
$\dot{\tau}_1 \neq 0,\ \dot{S}_1 \neq 0$	$R = \frac{r\dot{S}_1/\dot{S}_0}{\left[\frac{\dot{S}_1}{\dot{S}_0}\exp(-\Delta S/(a\sigma)) - 1\right]\exp(-t/t_a) + 1}$, where $t_a = a\sigma/\dot{S}_1$ [32]
$\dot{\tau}_1 = 0,\ \dot{S}_1 = 0$	$R = \frac{r}{\exp(-\Delta S/a\sigma) + \frac{t\dot{S}_0}{a\sigma}}$ [33]
Log stressing[b]	$R = \frac{r}{(wt+1)^{-m}\exp\left(\frac{-\Delta S}{a\sigma}\right) + \frac{(wt+1) - (wt+1)^{-m}}{\dot{S}_0 a\sigma w(m+1)}}$, where $m = \frac{u}{a\sigma}$ [34]

[a]Parameter values at the start and end of time/stress increment are indicated by the subscript values 0 and 1, respectively. Time $t = 0$ at the time of the stress step ΔS.
[b]Stressing by log of time following stress step as given in Table 3, eqn [31].

are given in Table 4 and Figure 6. These results are obtained through the stepwise application of the evolution equations in Table 3. Solutions for shear stressing at constant normal stress are identical to those in Table 4, with the substitution $\tau = S$.

Equation [33] can be written in a form that is identical the Omori's aftershock decay law

$$R = \frac{c_1}{c_2 + t} \qquad [35]$$

where the empirical parameters in the Omori equation are expressed in terms of constitutive parameters and stressing conditions in this model:

$$c_1 = \frac{ra\sigma}{\dot{S}_r} \quad \text{and} \quad c_2 = \frac{ra\sigma}{\dot{S}_r}\exp\left(\frac{-\Delta S}{a\sigma}\right) \qquad [36]$$

Equation [31] (Table 4, Figure 6), which incorporates continued stressing after the stress step, also has the form of Omori's aftershock decay law at

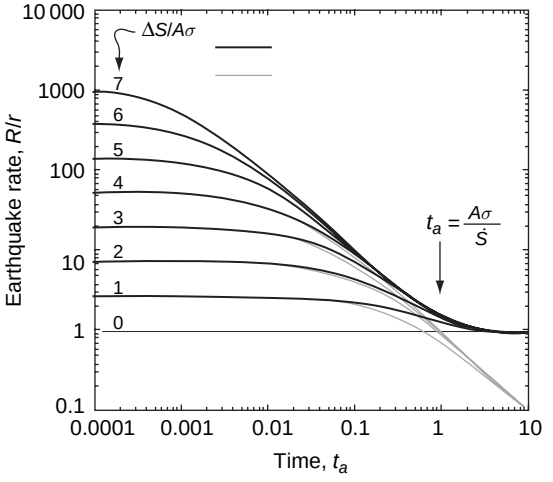


Figure 6 Plot of aftershock rate from eqn [32] (black lines), which has continuing stressing at $\dot{S}$ following a stress step, and from eqn [33] (gray lines), which has $\dot{S} = 0$ following the step.

times $t < t_a$, and earthquake rates asymptotically return to a constant steady-state background at times $t > t_a$. Hence, t_a is the aftershock duration, which is inversely proportional to stressing rate. This inverse dependence on stressing rate is supported by observations of aftershock sequences (**Figure 7**), which have durations that span a range time from a few days to 78 years (Dieterich, 1994; Toda *et al.*, 2002).

4.04.5.2.2 Variability of aftershock decay rates

Aftershocks decay as $1/t^p$, generally with $p \sim 1$. However, deviations from $p = 1$ appear to be reliable features of aftershock data. With the exception of eqn [34], which is discussed below, the stress step equations of **Table 4** predict aftershock decay exactly at $p = 1$. However, predicted aftershock rates integrated over a region with heterogeneous stress changes generally give decay with $p \neq 1$. For example, $p \approx 0.8$ is obtained by integrating aftershock rates over a finite region that has stress variations characteristic of those adjacent to a shear crack (Dieterich, 1994). Also, slip of a nonplanar fault with random fractal roughness results in highly heterogeneous stress changes adjacent to the fault (**Figure 8(a)**). Those stresses generate clustering of aftershocks near the fault and a systematic variation of the decay exponent p by distance from the earthquake rupture (**Figure 8(b)**). Finally, time-dependent relaxation of an earthquake stress step, perhaps fault creep, can result in aftershock decay with $p > 1$. For example, eqn [34] with values set to relax the stress step with the logarithm of time following an earthquake produces aftershock decay with $p > 1$(Dieterich, 1994).

4.04.5.3 Foreshocks

Faults with rate- and state-dependent properties appear capable of generating foreshocks by two, somewhat different, mechanisms (Dieterich and Kilgore, 1996a). The first mechanism assumes that mainshocks preceded by foreshocks are in essence aftershocks that happen to be larger than the prior earthquake. With this model, aftershocks are presumed to obey the Gutenberg–Richter frequency distribution of earthquake magnitudes. Hence, there is a finite probability that any aftershock will exceed the magnitude of the prior mainshock. The probability of this is determined by the rates of aftershocks, which are computed with eqn [32] using the stress changes around the region of the prior earthquake, together with the frequency distribution of earthquake magnitudes. The second mechanism assumes foreshocks are driven by mainshock nucleation. In this case local stressing is proportional to slip speed during mainshock nucleation and accelerates with time (eqns [17] and [20]). The accelerating rate of stressing causes locally accelerating rates of seismicity. For the second mechanism to operate, the dimensions of the mainshock nucleation zone must be significantly greater than the foreshock nucleation sources so that (1) mainshock nucleation is able to drive the foreshock sequences and (2) the region of stress perturbation is of sufficient size to yield an appropriate probability of triggering foreshock. Simple models of each process can be fit to data for foreshock–mainshock pairs (Dieterich and Kilgore, 1996a). However, there is presently no strong evidence to support, or reject, the requirement of the second mechanism that nucleation zones of mainshocks are larger than the nucleation zones of foreshocks.

Numerical simulations of seismicity that implicitly incorporate the ingredients of the first mechanism (rate- and state-dependent nucleation, stress transfer among simulated events, and development of magnitude/frequency distributions for the simulated earthquakes) spontaneously generate

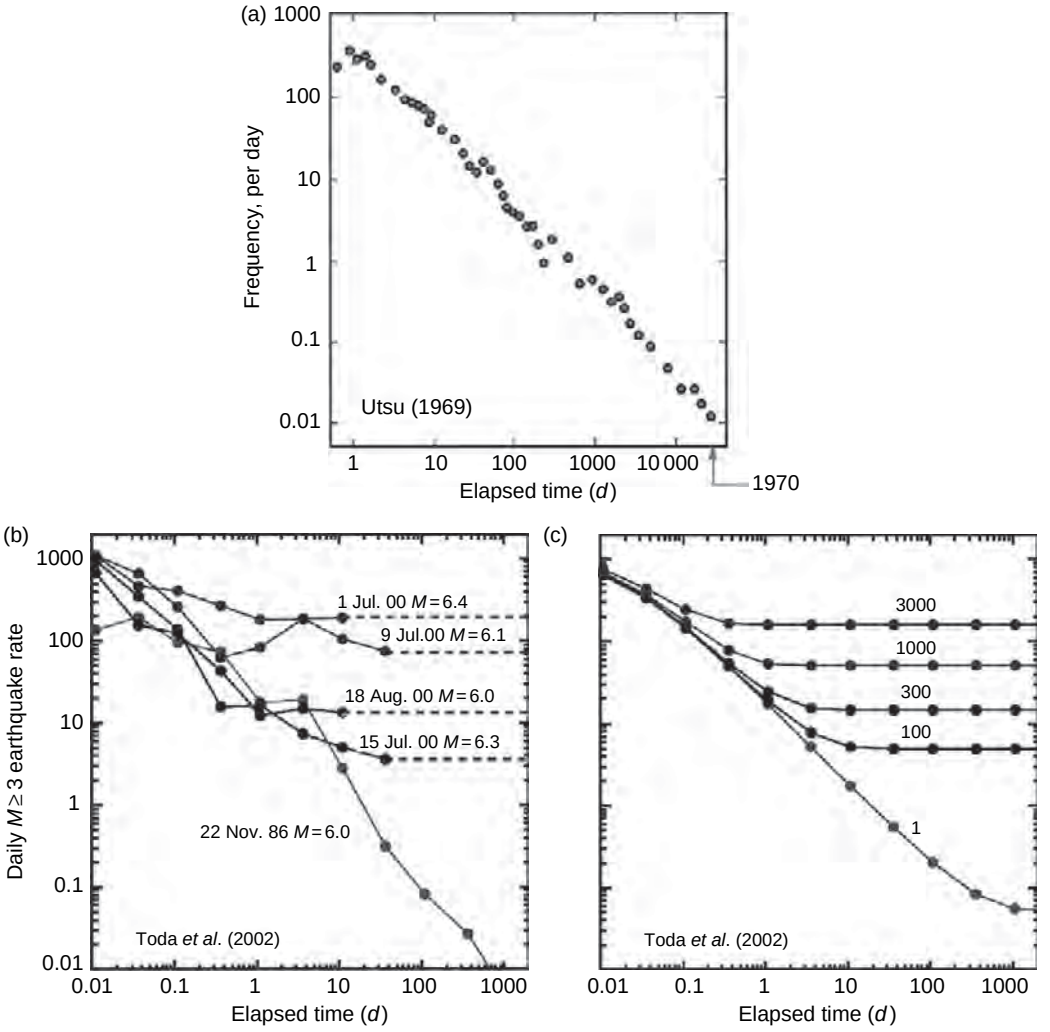


Figure 7 Duration of aftershocks. (a) Aftershocks of the $M \sim 8$, Nobi, Japan earthquake of 1891 continued at least to 1969 (Utsu, 1969) with no suggestion of returning to a background level after 79 years. Estimated recurrence time for this earthquake is 5000–20 000 years (Tsukuda, 1983), which indicates very low stressing rates. (b) Ultra short-duration aftershocks of 1–10 days following $M \sim 6$ earthquake that occurred during the intense Izu islands earthquake swarm of 2000 (Toda *et al*., 2002). (c) Predicted aftershock decay as a function of the stressing rate for the Izu Islands swarm from eqn [32]. Estimated stressing rates during the large intrusion that caused the swarm were as much as 3000 times larger than the background tectonic stressing rate (indicated on curves).

foreshocks (Dieterich, 1995; Ziv, 2003). This suggests that at least some foreshocks originate by this mechanism. In a study of foreshocks in the Western United States, Abercrombie and Mori (1995) find decreasing frequency of foreshocks with mainshock depth, which indicates that increasing normal stress inhibits foreshocks. This result is consistent with the normal stress dependence of the solutions in **Table 4** and suggests a triggered origin for foreshock processes.

4.04.5.4 Earthquake Triggering

Triggering of earthquakes by stress transients (principally tides and seismic waves) is a large topic with many unresolved questions. It seems quite reasonable that tidal stresses should have a significant imprint on earthquake occurrence because the stress changes that occur during a tidal cycle of a few hours are generally comparable in magnitude to months of tectonic stressing in seismically active

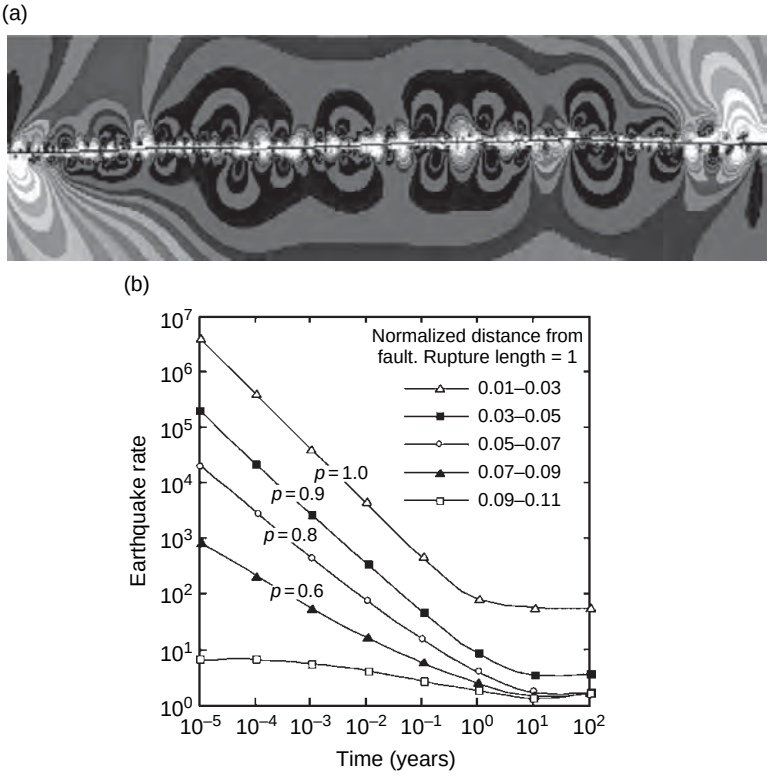


Figure 8 Coulomb stresses and aftershocks near a nonplanar fault (Dieterich, 2005). (a) Boundary element calculation of Coulomb stresses near a fault with random fractal roughness. Light shades indicate areas of stress increase and dark shades area of stress decrease. (b) Example of calculation of aftershocks as a function of distance from a fractal fault showing spatial variation of aftershock decay exponent p.

regions. Similarly, large earthquakes produce stress transients that are comparable to decades of tectonic stressing. Although tidal loading (Cochran *et al.*, 2004) and seismic waves (Prejean *et al.*, 2004; West *et al.*, 2005) are known to directly trigger earthquakes, observations show that triggering effects are surprisingly subtle. This can be ascribed to nucleation processes, which are time-dependent, and become increasingly insensitive to stress perturbations as the time to instability decreases (Section 4.04.4). The following discussion reviews some characteristics of, and constraints on, earthquake triggering, within the framework of rate- and state-dependent friction.

Constitutive formulations for earthquake rates (eqns [24] and [25]) indicate that earthquake rates will fluctuate in response to tectonic stressing with a superimposed periodic component. Equations [24] and [25] may be evaluated numerically for oscillating stresses using the methods outlined in Section 4.04.5.1. However, useful analytic solutions are readily obtained for conditions of steady-state seismicity ($\gamma_{ss} = 1/\dot{S}$), which is perturbed by periodic stresses. For these conditions, time-dependent evolution of γ becomes negligible during stress oscillations with periods less than $\sim 0.1 t_c$ where

$$t_c = \frac{a\sigma}{\dot{S}} \exp\left(\frac{-\Delta S}{a\sigma}\right) \qquad [37]$$

Tectonic stressing with perturbation by earth tides, ocean tidal loading, and seismic waves fall well below this limit. Hence, evolution with time is negligible and only the direct stress effect as given in eqns [28] and [30] operates on γ. For periodic stressing of the form $S(\phi) = C \cos \phi$, where ϕ is the phase angle, and C is the amplitude of the cyclic stress component

$$\gamma(\phi) = \frac{1}{\dot{S}} \exp\left(\frac{-C \cos \phi}{a\sigma}\right) \qquad [38]$$

where $\dot{S}$ is the tectonic stressing rate. This assumes steady-state seismicity ($\gamma_{ss} = 1/\dot{S}$) at $\phi = \pi/2,\ \pi/2$. Setting $\dot{S}_r$ and r equal to the tectonic stressing rate

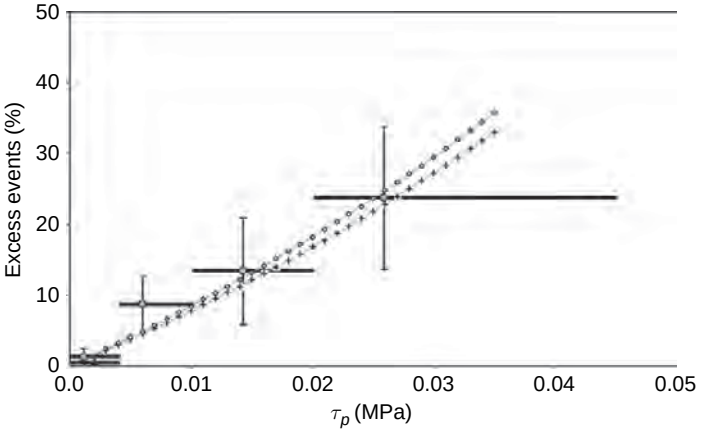


Figure 9 Extra earthquakes N_{ex} occurring at times of positive tidal stressing plotted against the tidal amplitudes τ_p(Cochran *et al.*, 2004). The quantity N_{ex} is the number of earthquakes in excess of the background rate that occur during intervals when stress exceeds the average value, expressed as a percentage of the background earthquakes. The solid triangle near the origin is for California strike-slip earthquakes. The gray circles are for shallow global thrust earthquakes. The stress amplitudes are mainly driven by ocean tidal loading. The slightly lower curve is a fit to the data using the rate-state equation [39], The upper curve is a fit based on a stress corrosion model.

and background steady-state earthquake rate, respectively, gives

$$R(\phi) = r \exp\left(\frac{C \cos \phi}{a\sigma}\right) \qquad [39]$$

Figure 9 shows a fit of this equation to the data of Cochran *et al.* (2004) for earth and ocean tidal loading.

Distant triggering of earthquakes was noted at the time of the M 7.3 Landers earthquake of 1992 (Hill *et al.*, 1993). Direct triggering of local earthquakes during the passage of the wave train has been widely reported for a number earthquake including the 3 November 2002 M 7.9 Denali, Alaska, earthquake where triggering was observed at Mt. Rainier, Washington, Geysers geothermal field, California, Mammoth Mountain/Long Valley Caldera, California, and Coso Geothermal Field, California (Prejean *et al.*, 2004). Some other examples of direct trigging include the 22 December 2003 M 6.5 San Simeon, California, earthquake (Hardebeck *et al.*, 2004), the 6 December 1999 M 7.0 Karluk Lake, Alaska, earthquake (Power *et al.*, 2001), and the 26 December 2004 M 9.3 Sumatra earthquake (West *et al.*, 2005).

The triggering of earthquakes during the passage of stress waves may be evaluated in a similar manner to that used for analysis of tidal stressing because the stressing conditions for seismic waves fall well below the limit where evolution of γ with time is important (eqn [37]); that is, eqns [28] and [30] may be directly applied to follow the change of γ with stress. **Figure 10** illustrates a calculation for triggering by an idealized surface wave. The dependence of triggering on amplitude of stress waves is quite nonlinear, and stress amplitudes of at least $(\Delta S/a\sigma) > 3$ appear necessary to achieve significant levels of triggering. Generic values of $a\sigma$ from aftershock studies (Dieterich, 1994) are in the range 0.1–0.2 MPa, which indicates that somewhat high stress amplitudes, in excess of 0.3 MPa, are required for triggering by short-duration transients. Remote triggering appears to occur preferentially in geothermal and volcanic areas (Hill *et al.*, 1993; Prejean *et al.*, 2004; Stark and Davis, 1996; West *et al.*, 2005) where fluid pressures may be anomalously high, and where earthquakes are generally quite shallow – these would reduce the effective normal stress in $(\Delta S/a\sigma)$ permitting a comparable decrease in the required stress triggering stress ΔS. Additionally, the passage of seismic waves may lead to significant amplification of pore-pressure transients (Brodsky *et al.*, 2003), which would amplify fluctuations in the effective normal component in the stress transient.

Remotely triggered earthquakes often continue for days or weeks after the passage of seismic waves through a region, with the elevated rates decaying in an aftershock-like manner (Hill *et al.*, 1993; Brodsky, 2006). Diverse mechanisms have been proposed to explain continuing elevated rates of seismicity. In the context of rate-state friction, no viable mechanisms have been identified that would

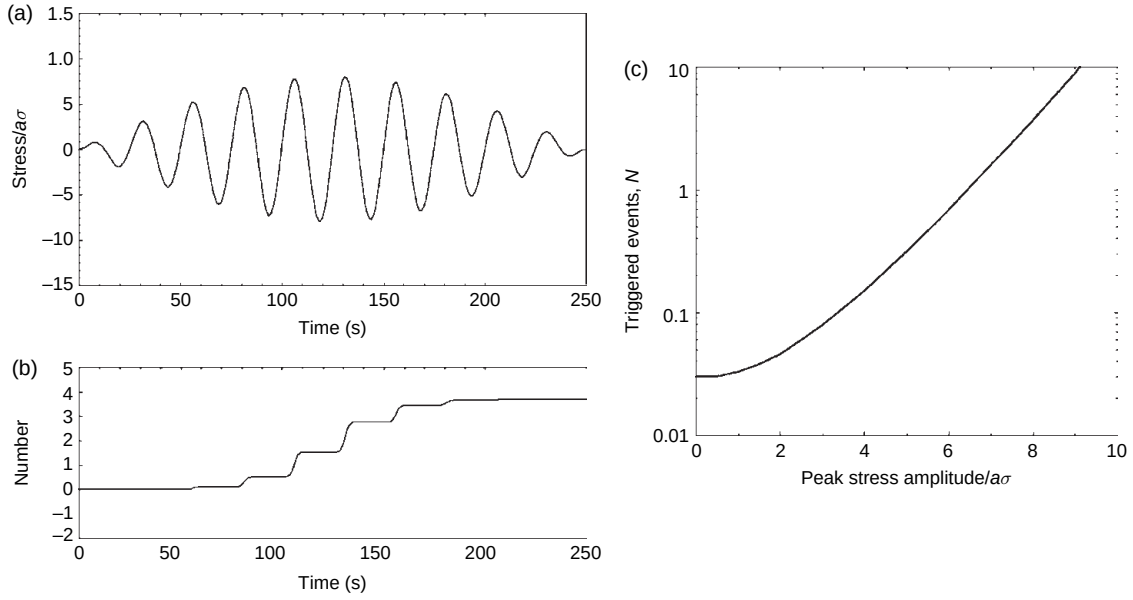


Figure 10 Example of a simulation of dynamic stress triggering. (a) Stresses from an idealized surface wave with peak amplitude of 8$a\sigma$. (b) Triggered events calculated with eqns [24] and [30] using (a) as an input. The background rate is set at 10/day. (c) Dependence of number of triggered earthquake on peak amplitude. All parameters are held fixed except for the amplitude of the stress oscillations. Black circle denotes conditions for (a) and (b).

explain sustained seismicity as arising from seismic shaking alone. However, aftershock-like decay of the elevated seismicity suggests that sustained triggering may simply represent an aftershock-like response to a local stress event, triggered by the seismic waves, as described by stress step equations of **Table 4**. Brodsky (2006) presents evidence in support of a proposal by Hough and Kanamori (2002) that continuing earthquakes are normal aftershocks to earthquakes directly triggered by the seismic waves. Other possible sources for triggered local stress events include, aseismic fault creep (Anderson *et al.*, 1994; Bodin and Gomberg, 1994), fluid pressure steps caused by destruction of flow barriers (Hill *et al.*, 1993; Brodsky *et al.*, 2003), and various adjustments in magmatic systems (Hill *et al.*, 1993).

4.04.6 Stress Changes Estimated from Earthquake Rates

Stress changes in the Earth's crust are generally estimated from model calculations that use near-surface deformation as an observational constraint. The widespread correlation of changes of earthquake activity with stress (Reasenberg and Simpson, 1992; King *et al.*, 1994; Harrts *et al.*, 1995; Stein, 1999) suggests that stress changes might be calculated from changes of earthquake rates. This possibility has considerable appeal, because seismicity data are routinely collected and have good spatial and temporal resolution. The time- and stress-dependencies contained in the formulation for earthquake rate, eqns [24] and [25], appear to be consistent with a number of earthquake phenomena over range of timescales and stressing conditions. The effectiveness of this formulation in forward modeling of earthquake phenomena suggests that it can be used to estimate stress changes from earthquake-rate data.

Recently, a method to do this has been implemented and applied to stress changes at Kilauea Volcano, Hawaii (Dieterich *et al.*, 2000, 2003). It employs observed seismicity rates as a function of time $R(t)$ to directly calculate γ as a function of time. From eqn [24], it follows that

$$\gamma(t) = \frac{r}{R(t)\dot{S}_r} \qquad [40]$$

The differential eqn [25] for γ is then solved for Coulomb stress change as a function of time. The method yields stresses that drive the earthquake process. As such, it is distinct from other seismological methods that yield measures of stress changes resulting from earthquakes.

Figure 11 illustrates some stress-change computations from Kilauea Volcano. Initial results for the stress calculations appear promising – the magnitudes and patterns agree with other estimates of stress change, including those obtained from boundary element modeling of surface deformation observations. The sequence of stress changes shown in Figure 11 indicate a pattern of progressive stress transfer following a dike intrusion event at the onset of the Puu Oo eruption in 1983. Figure 12 shows a recurring pattern of stress accumulation and release associated with repeating $M \sim 5$ earthquakes on the south flank of Kilauea Volcano.

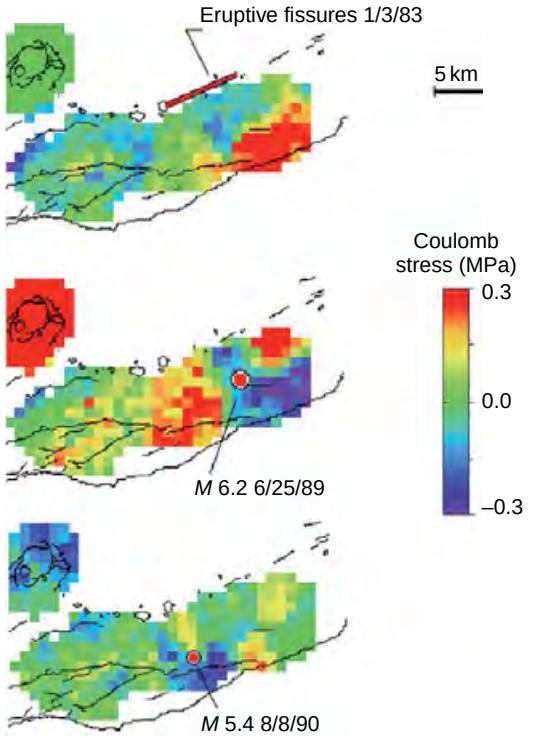


Figure 11 Coulomb stress interactions at Kiluaea volcano, Hawaii, calculated from changes of earthquake rates. Kilauea caldera is at upper left of each image. A pattern of progressive stress transfer began with the dike intrusion and eruption at the beginning of 1983 (marked by the eruptive fissures along southeast rift). The intrusion opened the upper portion of the rift zone by about 2 m, which stressed a portion of the south flank of Kilauea (red). The region stressed by the intrusion experienced a large earthquake in 1989, which appeared to largely relax the intrusion stress, but resulted in stress increase in the region to the southwest. Fourteen months later in 1990 an M 5.4 earthquake occurred in this area. Data from Dieterich JH, Cayol V, and Okubo P (2003) Stress changes before and during the Puu Oo–Kupianaha eruption. *In*: Heliker C, Donald AS, and Taeko J (eds.) *US Geological Survey Professional Paper 1676*: *The Puu Oo–Kupaianaha Eruption of Kilauea Volcano*, *Hawaii*: *The First 20 Years*, pp. 187–202. Reston, VA: US Geological Survey.

4.04.7 Conclusions and Future Directions

The rate- and state-dependent representation of fault friction has proven to be effective both in characterizing frictional properties in laboratory experiments, and in modeling diverse fault slip and earthquake phenomena. To date, most modeling efforts with rate-state friction have been based on simple spring-slider models and 2-D analyses with planar faults. Faults and fault systems are intrinsically 3-D structures with exquisitely complex geometries. One may speculate that new processes and deeper insights will emerge from fully 3-D analyses – particularly so when those models incorporate the geometric characteristics of faults in nature.

From an experimental standpoint there is considerable need for characterizing fault constitutive properties with a variety of lithologies appropriate for subducting sediments, and deep fault zone, and over a wider range of pressure, temperature, and slip rates. For example, there are very few experiments that use the materials and pressure–temperature conditions appropriate for deep fault creep, including the evolving lithologies of subducting plates. The detailed characteristics of frictional properties under these conditions are likely to be of critical importance for understanding and modeling the deep aseismic strain events, which increasingly appear to be characteristic of deep fault slip. Also, slip of nonplanar faults and fault systems necessarily involves coupled interactions among slip, normal stress, and state. Current evolution relations for evolution of state with changes of normal stress are based on a small data set.

Finally, it is noted again that the rate-state evolution equations are entirely empirical. To some extent the micromechanical processes that give rise to rate-state friction have been characterized. However, we currently lack a comprehensive theoretical framework for deriving the functional dependencies of state evolution from more fundamental parameters – this limits our ability to extrapolate fault properties to conditions that are difficult to access in laboratory experiments.

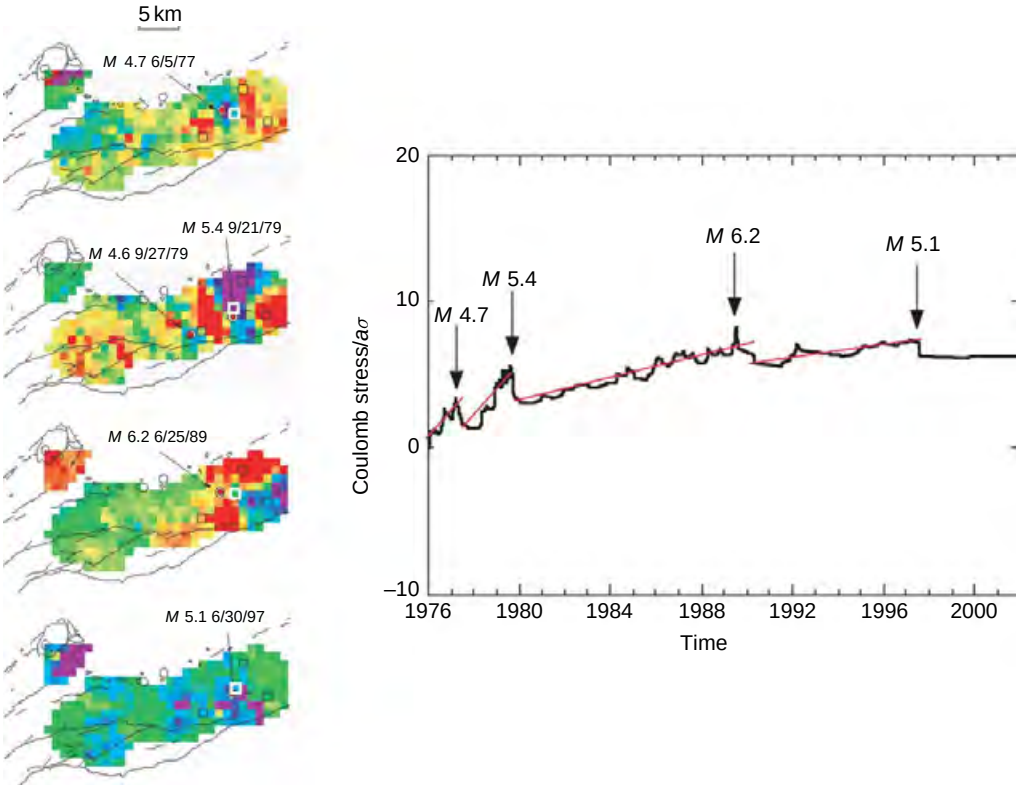


Figure 12 Coulomb stress changes around an area of repeating earthquakes, calculated from changes of earthquake rates. This is the same region shown in **Figure 11**. Left panels show Coulomb stress changes at the time of four earthquakes that affected the same area. In every case there is a region of decreased stress (stress shadows) near the hypocenters. Right panel shows stress versus time computed from earthquake rate change at the point marked by the white square in the bottom panel. It shows repeating pattern of progressive stress increase followed by stress drop at the time of the earthquakes. Data from Dieterich JH, Cayol V, and Okubo P (2003) Stress changes before and during the Puu Oo–Kupianaha eruption. *In*: Heliker C, Donald AS, and Taeko J (eds.) *US Geological Survey Professional Paper 1676*: *The Puu Oo–Kupaianaha Eruption of Kilauea Volcano*, *Hawaii*: *The First 20 Years*, pp. 187–202. Reston, VA: US Geological Survey.

References

Abercrombie RE (1995) Earthquake source scaling relationships from −1 to 5 *ML* using seismograms recorded at 2.5 km depth. *Journal of Geophysical Research* 100: 24015–24036.

Abercrombie RE and Mori J (1995) Occurrence patterns of foreshocks to large earthquakes in the Western United States. *Nature* 381: 303–307.

Anderson JG, Brune JN, Louie J, *et al.* (1994) Seismicity in the western Great Basin apparently triggered by the Landers, California, earthquake, June 28, 1992. *Bulletin of the Seismological Society of America* 84: 863–891.

Beeler NM, Tullis TE, and Weeks JD (1994) The roles of time and displacement in the evolution effect in rock friction. *Geophysical Research Letters* 21: 1987–1990.

Bodin P and Gomber J (1994) Triggered seismicity and deformation between the Landers, California, and Little Skull Mountain, Nevada earthquakes. *Bulletin of the Seismological Society of America* 84: 835–843.

Brodsky EE (2006) Long-range triggered earthquakes that continue after the wave train passes. *Geophysical Research Letters* 33: L15313 (doi:10.1029/2006GL026605).

Brodsky EE, Roeloffs E, Woodcock D, Gall I, and Manga M (2003) A mechanism for sustained groundwater pressure changes induced by distant earthquakes. *Journal of Geophysical Research* 108: 2390 (doi:10.1029/2002JB002321).

Brooks BA, Foster JH, Bevis M, Fazer LN, Wolfe CJ, and Behn M (2006) Periodic slow earthquakes on the flank of Kilauea Volcano Hawaii. *Earth and Planatery Science Letters* 246: 207–216.

Belardinelli ME, Bizzari A, and Cocco M (2003) Earthquake triggering by static and dynamic stress changes. *Journal of Geophysical Research* 108(B3): 2135 (doi:10.1029/2002JB001779).

Beroza GC and Jordan TH (1990) Searching for slow and silent earthquakes using free oscillations. *Journal of Geophysical Research* 95: 2485–2510.

Blanpied ML and Tullis TE (1986) The stability and behavior of a system with two state variable constitutive law. *Pure and Applied Geophysics* 124: 415–444.

Blanpied ML, Lockner DA, and Byerlee JD (1991) Fault stability inferred from granite sliding experiments at hydrothermal conditions. *Geophysical Research Letters* 18: 609–612.

Blanpied ML, Lockner DA, and Byerlee JD (1995) Frictional slip of granite at hydrothermal conditions. *Journal of Geophysical Research* 100: 13045–13064.

Bodin P and Gomberg J (1994) Triggered seismicity and deformation between the Landers, California, and Little Skull

Mountain, Nevada, earthquakes. *Bulletin of the Seismological Society of America* 84: 835–843.
Bucknam RC, Plafker G, and Sharp RV (1978) Fault movement (afterslip) following the Guatemala earthquake of February 4, 1976. *Geology* 6: 170–173.
Burford RO (1976) Continued slip on the Coyote Creek fault after the Borrego Mountain earthquake, in The Borrego Mountain earthquake of April 9, 1968. *US Geological Survey Professional Paper* 787: 105–111.
Bürgmann R, Ergintav S, Segall P, *et al.* (2002) Time-dependent distributed afterslip on and deep below the Izmit earthquake rupture (*in* The Izmit, Turkey, earthquake of 17 August 1999, Toksoz) *Bulletin of the Seismological Society of America*, 92(1), 126–137.
Byerlee JD (1978) Friction of rocks. *Pure and Applied Geophysics* 116: 615–626.
Cervelli P, Segall P, Johnson K, Lisowski M, and Miklius A (2002) Sudden aseismic fault slip on the south flank of Kilauea Volcano. *Nature* 415: 1014–1018.
Cochran ES, Vidale JE, and Tanaka S (2004) Earth tides can trigger shallow thrust fault earthquakes. *Science* 306: 1164–1166.
Crescentini L, Amoruso A, and Scarpa R (1999) Constraints on slow earthquake dynamics from a swarm in central Italy. *Science* 286: 2132–2134.
Dieterich JH (1972) Time-dependent friction in rocks. *Journal of Geophysical Research* 77: 3690–3697.
Dieterich JH (1979) Modelling of rock friction. Part 1: Experimental results and constitutive equations. *Journal of Geophysical Research* 84(B5): 2161–2168.
Dieterich JH (1981) Constitutive properties of faults with simulated gouge. In: Carter NL, Friedman M, Logan JM, and Stearns DW (eds.) *Monograph 24*: *Mechanical Behavior of Crustal Rocks*, pp. 103–120. Washington, DC: American Geophysical Union.
Dieterich JH (1986) A model for the nucleation of earthquake slip. In: Das S, Boatwright J, and Scholz C (eds.) *Geophysical Monograph 37*: *Earthquake Source Mechanics*, pp. 37–47. Washington, DC: American Geophysical Union.
Dieterich JH (1987) Nucleation and triggering of earthquake slip: Effect of periodic stresses. *Tectonophysics* 144: 127–139.
Dieterich JH (1992) Earthquake nucleation on faults with rate- and state-dependent friction. *Tectonophysics* 211: 115–134.
Dieterich JH (1994) A constitutive law for rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research* 99: 2601–2618.
Dieterich JH (1995) Earthquake simulations with time-dependent nucleation and long-range interactions. *Journal of Nonlinear Processes in Geophysics* 2: 109–120.
Dieterich JH (2005) Role of Stress Relaxation in Slip of Geometrically Complex Faults, *Abstract*, *EOS*, *Transactions American Geophysical Union* (*Fall Meeting Supplement*), NG31A-04.
Dieterich JH and Conrad G (1984) Effect of humidity on time- and velocity-dependent friction in rocks. *Journal of Geophysical Research* 89: 4196–4202.
Dieterich JH, Cayol V, and Okubo P (2000) The use of earthquake rate changes as a stress meter at Kilauea Volcano. *Nature* 408: 457–460.
Dieterich JH, Cayol V, and Okubo P (2003) Stress changes before and during the Puu Oo–Kupianaha eruption. In: Heliker C, Donald AS, and Taeko J (eds.) *US Geological Survey Professional Paper 1676*: *The Puu Oo–Kupaianaha Eruption of Kilauea Volcano, Hawaii*: *The First 20 Years*, pp. 187–202. Reston, Va: US Geological Survey.
Dieterich JH and Kilgore BD (1994) Direct observation of frictional contacts: New insights for sliding memory effects. *Pure and Applied Geophysics* 143: 283–302.
Dieterich JH and Kilgore B (1996a) Implications of fault constitutive properties for earthquake prediction. *Proceedings of the National Academy of Sciences USA* 93: 3787–3794.
Dieterich JH and Kilgore B (1996b) Imaging surface contacts; power law contact distributions and contact stresses in quartz, calcite, glass, and acrylic plastic. *Tectonophysics* 256: 219–239.
Dieterich JH and Linker MF (1992) Stability of fault slip under condition of variable normal stress. *Geophysical Research Letters* 19: 1691–1694.
Dragert H, Wang K, and James TS (2001) A silent slip event on the deeper Cascadia subduction interface. *Science* 292: 1525–1528.
Dvorak JJ, Okamura AT, English TT, *et al.* (1986) Mechanical response of the south flank of Kilauea Volcano, Hawaii, to intrusive events along the rift systems. *Tectonophysics* 124: 193–209.
Evans B and Goetz C (1979) The temperature variation of hardness of olivine and its implication for polycrystalline yield stress. *Journal of Geophysical Research* 84: 5505–5524.
Freymüeller JT, Hreinsdóttir S, Zweck C, and Haeussler PJ (2002) The 1998–2002 deep megathrust slip event, Alaska, Eos Trans. AGU Fall. *Meet. Suppl.* abstract G61A–0972 .
Fujii Y (1993) Vertical crustal movement in the Boso peninsula, South Kanto, Japan, as deduced from the adjustment of a geodetic network with signals. *Tectonophysics* 218: 309–322.
Gladwin MT, Gwyther RL, Hart RHG, and Breckenridge KS (1994) Measurements of the strain field associated with episodic creep events on the San Andreas fault at San Juan Bautista, California. *Journal of Geophysical Research* 99(B3): 4559–4566.
Goldsby DL, Rar A, Pharr GM, and Tullis TE (2004) Nanoindentation creep of quartz with implications for rate- and state-variable friction laws relevant to earthquake mechanics. *Journal of Materials Research* 19(1): 357–365.
Gomberg J, Beeler N, and Blanpied M (2000) On rate-state and Coulomb failure models. *Journal of Geophysical Research* 105: 7557–7871.
Gomberg J, Beeler NM, Blanpied ML, and Bodin P (1998) Earthquake triggering by transient and static deformations. *Journal of Geophysical Research* 103: 24411–24426.
Gomberg J, Blanpied ML, and Beeler NM (1997) Transient triggering of near and distant earthquakes. *Bulletin of the Seismological Society of America* 87: 294–309.
Gu J, Rice JR, Ruina AL, and Tse ST (1984) Slip motion and stability of a single degree of freedom elasitic system rate and state dependent friction. *Journal of the Mechanics and Physics of Solids* 32: 167–196.
Hardebeck JL, Boatwright J, Dreger D, *et al.* (2004) Preliminary report on the 22 December 2003 M6.5 San Simeon, California, earthquake. *Seismological Research Letters* 75: 155–172.
Harrts RA, Simpson R, and Reasenberg PA (1995) Influence of static stress changes on earthquake locations in southern California. *Nature* 375: 221–224.
Hill DP, Reasenberg PA, Michael A, *et al.* (1993) Seismicity remotely triggered by the Magnitude 7.3 Landers, California, earthquake. *Science* 260: 1617–1623.
Hirose H, Hirahara K, Kimata F, Fujii N, and Miyazaki S (1999) A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, southwest Japan. *Geophysical Research Letters* 26: 3237–3240.
Hough S and Kanamori H (2002) Source properties of earthquakes near the Salton Sea triggered by the 16 October 1999 M 7.1 Hector Mine, California, earthquake.

Bulletin of the Seismological Society of America 92: 1281–1289.

Hsu Y-J, Bechor N, Segall P, Yu S-B, Kuo L-C, and Ma K-F (2002) Rapid afterslip following the 1999 Chi-Chi, Taiwan Earthquake. *Geophysical Research Letters* 29 (doi:10.1029/2002GL014967).

Hutton W, DeMets C, Sánchez O, Suárez G, and Stock J (2002) Slip kinematics and dynamics during and after the 1995 October 9 *Mw*=8.0 Colima–Jalisco earthquake, Mexico, from GPS geodetic constraints. *Geophysical Journal International* 146: 637.

Kodaira S, Iidaka T, Kato A, Park J-O, Iwasaki T, and Kaneda Y (2004) High pore fluid pressure may cause silent slip in the Nankai Trough. *Science* 304: 1295–1298.

Kato N and Hirasawa T (1997) A numerical study on seismic coupling along subduction zones using a laboratory-derived friction law. *Physics of the Earth and Planetary Interiors* 102: 51–68.

Kanamori H and Hauksson E (1992) A slow earthquake in the Santa Maria Basin, California. *Bulletin of the Seismological Society of America* 82: 2087–2096.

Kawasaki I (2004) Silent earthquakes occurring in a stable-unstable transition zone and impications for earthquake prediction. *Earth Planets Space* 56: 813–821.

Kawasaki I, Asai Y, and Tamura Y (2001) Space-time distribution of interplate moment release including slow earthquakes and the seismo-geodetic coupling in the Sanriku-oki region along the Japan trench. *Tectonophysics* 300: 267–283.

Kilgore BD, Blanpied ML, and Dieterich JH (1993) Velocity-dependent friction of granite over a wide range of conditions. *Geophysical Research Letters* 20: 903–906.

King GCP, Stein RS, and Lin J (1994) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

Kodaira S, Iidaka T, Kato A, Park J-O, Iwasaki T, and Kaneda Y (2004) High pore fluid pressure may cause silent slip in the Nankai Trough. *Science* 304: 1295–1298.

Kostoglodov V, Singh SK, Santiago JA, and Franco SI (2003) An large silent earthquake in the Guerroro seismic gap, Mexico. *Geophysical Research Letters* 30: 1807 (doi:10.1029/2003GL017219).

Langbein JO (1981) An interpretation of episodic slip on the Calaveras fault near Hollister, California. *Journal of Geophysical Research* 86: 4941–4948.

Lapusta N and Rice JR (2003) Nucleation and early seismic propagation of small and large events in a crustal earthquake model. *Journal of Geophysical Research* 108(B4): 2205 (doi:10.1029/2001JB000793).

Lapusta N, Rice JR, Ben-Zion Y, and Zheng G (2000) Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate- and state-dependent friction. *Journal of Geophysical Research* 105(B10): 23765–23789.

Linde AT and Sacks IS (2002) Slow earthquakes and great earthquakes along the Nankai trough. *Earth and Planetary Science Letters* 203: 265–275.

Linker MF and Dieterich JH (1992) Effects of variable normal stress on rock friction: Observations and constitutive relations. *Journal of Geophysical Research* 97: 4923–4940.

Liu Y and Rice JR (2005) Aseismic slip transients emerge spontaneously in three-dimensional rate and state modeling of subduction earthquake sequences. *Journal of Geophysical Research* 110: B08307, doi:10.1029/2004JB003424.

Lowry AR, Larson KM, Kostoglodov V, and Bilham R (2001) Transient fault slip in Guerrero, southern Mexico. *Geophysical Research Letters* 28: 3753–3756.

Nason RD and Weertman J (1973) A dislocation theory analysis of fault creep events. *Journal of Geophysical Research* 78: 7745–7751.

Marone C, Raleigh CB, and Scholz CH (1990) Frictional behavior and constitutive modeling of simulated fault gouge. *Journal of Geophysical Research* 95: 7007–7025.

Marone C (1998) Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences* 26: 643–696.

Marone C and Kilgore B (1993) Scaling of the critical slip distance for seismic faulting with shear strain in fault zones. *Nature* 362: 618–621.

Marone CJ, Scholz CH, and Bilham R (1991) On the mechanics of earthquake afterslip. *Journal of Geophysical Research* 96(B5): 8441–8452.

Miller MM, Melbourne TI, Johnson DJ, and Sumner WQ (2002) Periodic slow earthquakes from the Cascadia Subduction Zone. *Science* 295: 2423.

McHugh S and Johnston MJS (1976) Short period non-seismic tilt perturbations and their relation to episodic slip on the San Andreas fault. *Journal of Geophysical Research* 81: 6341–6346.

Miyazaki S, Segall SP, Fukuda J, and Kato T (2004) Space time distribution of afterslip following the 2003 Tokachi-oki earthquake: Implications for variations in fault zone frictional properties. *Geophysical Research Letters* 31 (doi:10.1029/2003GL019410).

Okubo PG (1989) Dynamic rupture modeling with laboratory-derived constitutive relations. *Journal of Geophysical Research* 94: 12321–12335.

Okubo PG and Dieterich JH (1986) State variable fault constitutive relations for dynamic slip. In: Das S, Boatwright J, and Scholz C (eds.) *Monograph Series 37: Earthquake Source Mechanics*, pp. 25–35. Washington, DC: AGU.

Ozawa S, Murakami M, Kaidzu M, *et al.* (2002) Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan. *Science* 298: 1009–1012.

Prejean SG, Hill DP, Brodsky EE, *et al.* (2004) Remotely triggered seismicity on the United States West Coast following the *Mw* 7.9 Denali fault earthquake. *Bulletin of the Seismological Society of America* 94: S348–S359.

Pollitz F, Bürgmann R, and Segall P (1998) Joint estimation of afterslip rate and postseismic relaxation following the 1989 Loma Prieta earthquake. *Journal of Geophysical Research* 103: 26975–26992.

Power JA, Moran SC, McNutt SR, Stihler SD, and Sanchez JJ (2001) Seismic response of the Katmai volcanoes to the 6 December 1999 Magnitude 7.0 Karluk Lake earthquake, Alaska. *Bulletin of the Seismological Society of America* 91: 57–63.

Prakash V (1998) Frictional response of sliding interfaces subjected to time varying normal pressures. *Journal of Triboiology* 120: 97–102.

Perrin G, Rice JR, and Zheng G (1995) Self-healing slip pulse on a frictional surface. *Journal of the Mechanics and Physics of Solids* 43: 1461–1495.

Reasenberg PA and Simpson RW (1992) Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake. *Science* 255: 1687–1690.

Rice JR (1983) Constitutive relations for fault slip and earthquake instabilities. *Pure and Applied Geophysics* 121: 443–475.

Rice JR (2006) Heating and weakening of faults during earthquake slip. *Journal of Geophysical Research* 111: Bo5311 (doi:10. 1029/2005JB004006).

Rice JR and Gu J-C (1983) Earthquake aftereffects and triggered seismic phenomena. *Pure and Applied Geophysics* 121: 187–219.

Rice JR and Ruina AL (1983) Stability of steady frictional slipping. *Journal of Applied Mechanics* 50: 343–349.

Rogers G and Dragert H (2003) Episodic tremor and slip on the cascadia subduction zone: The chatter of silent slip. *Science* 300: 1942–1992.

Rubin AM and Ampuero J-P (2005) Earthquake nucleation on (aging) rate and state faults. *Journal of Geophysical Research* 110: B11312 (doi:10.1029/2005JB003686).

Ruina AL (1983) Slip instability and state variable friction laws. *Journal of Geophysical Research* 88: 10359–10370.

Sacks IS, Linde AT, Snoke JA, and Suyehiro S (1981) A slow earthquake sequence following the Izu–Oshima earthquake of 1978. In: Simpson DW and Richards PG (eds.) *Maurice Ewing Series, vol 4: Earthquake Prediction: An International Review*, pp. 617–628. Washington, DC: AGU.

Sagiya T (2004) Interplate coupling in the Kanto district, central Japan, and the Boso Peninsula silent earthquake in May 1996. *Pure and Applied Geophysics* 161: 2327–2342.

Segall P and Rice JR (1995) Dilatancy, compaction, and slip instability of a fluid-infiltrated fault. *Journal of Geophysical Research* 100(B11): 22155–221121.

Segall P, Bürgmann R, and Matthews M (2000) Time-dependent triggered afterslip following the 1989 Loma Prieta earthquake. *Journal of Geophysical Research* 105(B3): 5615–5634.

Scholz CH (1998) Earthquakes and friction laws. *Nature* 391: 37–42.

Scholz CH and Engelder JT (1976) The role of asperity indentation and ploughing in rock friction. Part I: Asperity creep and stick-slip. *International Journal of Rock Mechanics and Mining Science Geomechanics Abstract* 77: 149–154.

Shibazaki B and Iio Y (2003) On the physical mechanism of silent slip events along the deeper part fo the seismogenic zone. *Geophysical Research Letters* 30: 1489–1492.

Stark MA and Davis SD (1996) Remotely triggered microearthquakes at The Geysers geothermal field, California. *Geophysical Research Letters* 23: 945–948.

Stein RS (1999) The role of stress transfer in earthquake occurrence. *Nature* 402: 605–609.

Stuart W (1988) Forecast model for great earthquakes at the Nankai Trough subduction zone. *Pure and Applied Geophysics* 126: 619–641.

Sleep NH (1995a) Ductile creep compaction and rate and state dependent friction with major fault zones. *Journal of Geophysical Research* 100: 13065–13080.

Sleep NH (1995b) Frictional heating and the stability of rate and state dependent frictional sliding. *Geophysical Research Letters* 22: 2785–2788.

Sleep NH (1997) Application of a unified rate and state dependent theory to the mechanics of fault zones with strain localization. *Journal of Geophysical Research* 102: 2875–2895.

Sleep NH and Blanpied ML (1992) Creep, compaction and the weak rheology of major faults. *Nature* 359: 687–692.

Smith SW and Wyss M (1968) Displacement on the San Andreas fault subsequent to the 1966, Parkfield earthquake. *Bulletin of the Seismological Society of America* 58: 1955–1973.

Stesky RM (1978) Rock friction – effect of confining pressure, temperature, and pore pressure. *Pure and Applied Geophysics* 116: 690–704.

Toda S, Stein RS, Reasenberg PA, Dieterich JH, and Yoshida A (1998) Stress transferred by the 1995 Mw = 6.9 Kobe, Japan, shock. *Journal of Geophysical Research* 103: 24543–24565.

Toda S, Stein RS, and Sagiya S (2002) Evidence from the AD 2000 Izu islands earthquake swarm that stressing rate governs seismicity. *Nature* 419: 58–61.

Tse ST and Rice JR (1986) Crustal earthquake instability in relation to the depth variation of frictional slip properties. *Journal of Geophysical Research* 91(B9): 9452–9472.

Tullis TE (1988) Rock friction constitutive behavior from laboratory experiments and its implications for and earthquake prediction field monitoring program. *Pure and Applied Geophysics* 126: 555–588.

Tullis TE and Weeks JD (1986) Constitutive behavior and stability of frictional sliding in granite. *Pure and Applied Geophysics* 124: 383–314.

Utsu T (1969) Aftershocks and earthquake statistics. *Journal of the Faculty of Science, Hokkaido University* 7(3): 129–195.

Walker WW and Demer LJ (1964) Effects of loading duration on indentation hardness. *Transactions of the American Institute of Mining, Metallurgical and Petroleum Engineers* 233: 425–428.

Weeks JD and Tullis TE (1985) Frictional sliding of dolomite: A variation in constitutive behavior. *Journal of Geophysical Research* 90: 7821–7826.

Wennerberg L and Sharp RV (1997) Bulk-friction modeling of afterslip and the modified Omori law. *Tectonophysics* 277: 109–136.

Wesson RL (1988) Dynamics of fault creep. *Journal of Geophysical Research* 93(B8): 8929–8951.

West M, Sanchez JJ, and McNutt SR (2005) Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake. *Science* 308: 1144–1146.

Westbrook JH and Jorgensen PJ (1965) Indentation creep of solids. *Transactions of the American Institute of Mining, Metallurgical and Petroleum Engineers* 230: 613–614.

Ziv A (2003) Foreshocks, aftershocks and remote triggering in Quasi-static fault models. *Journal of Geophysical Research* 108 (doi 10.1029/2002JB002318).

4.05 Friction of Rock at Earthquake Slip Rates

T. E. Tullis, Brown University, Providence, RI, USA

Nomenclature

c	heat capacity (J kg^{-1} °C^{-1})
c_{th}	thermal diffusivity for flash weakening (mm^2 s^{-1})
f	friction coefficient (dimensionless)
f_0	friction coefficient at velocities below V_w for flash weakening (dimensionless)
D	diameter of asperity contact (μm)
H	average thickness of the gap in hydrodynamic lubrication (μm or mm)
L	wavelength of the asperity in hydrodynamic lubrication (μm or mm or m)
L^*	length parameter in thermal pressurization analysis (mm)
p_0	pore pressure at the initiation of slip (MPa)
T_f	present background temperature of the fault for flash weakening (°C)
T_w	weakening temperature for flash weakening (°C)
V	slip velocity (m s^{-1})
V_w	velocity at which weakening occurs for flash weakening (m s^{-1})
α_{hy}	hydraulic diffusivity (mm^2 s^{-1})
α_{th}	thermal diffusivity (mm^2 s^{-1})
β_f	isothermal compressibilities of the fluid (/MPa)
β_n	pressure expansivity of the pore space (/MPa)
δ	slip displacement (mm or m)
η	fluid viscosity (Pa s)
θ	lifetime of representative contact in flash weakening analysis (s)
θ_w	weakening time for flash weakening (s)
λ_f	isobaric volumetric fluid thermal expansivity (/°C)
λ_n	isobaric volumetric pore space thermal expansivity (/°C)
ρ	material density (kg m^{-3})
ρc	specific heat of fault zone material (MPa °C)
σ_n	normal stress (MPa)
τ	shear stress (MPa)
τ_c	shear stress on asperity contacts for flash weakening (MPa)
ΔH	amplitude of the asperity in hydrodynamic lubrication (μm or mm)
Λ	undrained pressurization factor for thermal pressurization (MPa °C)

4.05.1 Introduction

Many data exist on the frictional behavior of rocks during slow interseismic, preseismic, and postseismic slip at velocities $<1\ \mathrm{mm\ s^{-1}}$, but only recently have limited data relevant to coseismic slip at velocities $\geq 1\ \mathrm{m\ s^{-1}}$ become available. Because many processes can potentially be activated at these high slip rates, we cannot assume that it is meaningful to extrapolate the small dependence of friction on slip velocity embodied in rate and state friction to seismic slip rates (e.g., Dieterich, 1972, 1978, 1979; Marone, 1998; Ruina, 1983; Tullis, 1988; Tullis, 1996; Tullis and Weeks, 1986). Several high-slip-velocity mechanisms have been proposed that may dramatically reduce the friction coefficient from typical quasi-static values of ~0.6 to values that can approach zero. This chapter briefly discusses these mechanisms, with an emphasis on those for which relevant experimental data exist. Much remains to be learned about all of these mechanisms and their relevance to coseismic slip resistance. Nevertheless, our present understanding suggests that, due to the likely operation of one or more of these mechanisms, coseismic slip resistance is likely to be dramatically lower than would correspond to a coefficient of friction of 0.6. In addition to the implications for the dynamics of earthquake rupture, this conclusion also has interesting implications for the magnitudes and orientations of stress in the upper brittle crust. These implications will be briefly discussed below.

The paucity of experimental data on the frictional properties of rocks relevant to coseismic slip in the Earth is due to the extreme conditions of earthquake slip and the difficulty of conducting experiments at these conditions. A laboratory experiment that reproduces the conditions existing during coseismic slip would have to simultaneously involve high slip rates ($1–10\ \mathrm{m\ s^{-1}}$), large slip displacements (0.1–20 m), high effective normal stress (50–200 MPa), elevated pore-fluid pressures (0.4–1 times the normal stress), and elevated temperature (ambient temperatures of 100–300°C, but as high as 1500°C in the slipping zone due to shear heating during an earthquake). Although progress is underway toward increasing experimental capability, no apparatus exists that is capable of simultaneously meeting these requirements. Consequently, existing experiments compromise on one or more of these factors.

4.05.2 Dynamic Fault Weakening Mechanisms

High-speed weakening mechanisms that have been proposed in the literature and that are discussed in this chapter are listed below. The discussion will focus on the last three mechanisms, since these are the ones for which relevant experimental data have recently been obtained:

1. normal stress reduction or loss of contact from normal interface vibrations;
2. dynamic normal stress reduction from elastic or permeability mismatch;
3. acoustic fluidization;
4. elastohydrodynamic lubrication;
5. thermal pressurization of pore fluid;
6. 'flash' heating/melting at asperity contacts;
7. interfacial lubrication by friction melt; and
8. interfacial lubrication by thixotropic silica gel.

4.05.2.1 Normal Stress Reduction or Loss of Contact from Normal Interface Vibrations

Brune *et al.* (1990, 1993) proposed that reductions in normal stress or perhaps even actual loss of contact could occur on a sliding surface during dynamic slip. If this were the case, the shear resistance during slip would be reduced, potentially to zero. That this might occur is supported by direct observations of opening during dynamic slip between foam rubber blocks (Brune *et al.*, 1993). Whether it can occur during sliding on rock surfaces is less clear. Such opening is only predicted theoretically for smooth surfaces in cases where differences in elastic properties exist across the interface (the bimaterial effect discussed below). The observed behavior of the foam rubber occurs without elastic contrast and may result from interactions of the dangling fibers that characterize a cut surface in foam rubber. Intuitively, one might expect that coseismic sliding of fault surfaces with roughness could result in fault opening or reductions in normal stress due to asperities bouncing off one another. Whether this would result in a net reduction in the time- or space-averaged shear resistance is unclear, since the dynamic contact between the asperities would presumably result in both increases in resistance from impact and decreases from rebound. Anooshehpoor and Brune (1994a, 1994b, 1999), Brown *et al.* (1991), Brune (1991, 1996), Brune *et al.* (1993, 1990), Brune and Thatcher (2002), and Castro *et al.* (1991) have suggested that unexpectedly large

high-frequency P-wave radiation observed from earthquakes may indicate that normal vibrations occur during dynamic slip. Brune (2001) has suggested that pulverized rock, which contains no fault-parallel shear features and is found along surface traces of the San Andreas Fault zone, could be the result of such vibrations, although other explanations for pulverized rock exist (e.g., Dor *et al.*, 2006a, 2006b). This potential weakening mechanism is presently not well understood and whether it is important for reducing the shear resistance during dynamic earthquake slip is unclear. It could be important.

4.05.2.2 Dynamic Normal Stress Reduction from Elastic or Permeability Mismatch

As originally analyzed by Weertman (1963, 1980), if dynamic slip occurs on an interface between two materials having different elastic properties, changes in normal stress on the interface will occur in the slipping area behind the propagating rupture tip. As shown in **Figure 1**, at the end of the rupture that has the less compliant material slipping toward it, the normal stress is increased, whereas at the opposite end of the rupture that has the more compliant material slipping toward it, there is a reduction in normal stress propagating along as a wrinkle-like pulse behind the rupture tip. This can result in two related effects, a reduction in the dynamic resistance to slip and a related tendency for the rupture to propagate at different speeds in the two different directions. In the extreme case, the rupture might propagate unilaterally. Although these effects have been studied extensively (Adams, 1995, 1998, 2001; Andrews and Ben-Zion, 1997; Andrews and Harris, 2005, 2006; Anooshehpoor and Brune, 1999; Ben-Zion, 2001, 2006; Ben-Zion and Andrews, 1998; Ben-Zion and Huang, 2002; Ben-Zion and Shi, 2005; Cochard and Rice, 2000; Harris and Day, 1997, 2005; Ranjith and Rice, 2001; Shi and Ben-Zion, 2006; Weertman, 1963, 1980, 2002; Xia *et al.*, 2005), there has been some controversy about the importance of these effects, especially concerning whether ruptures propagate unilaterally or bilaterally. Thus, some studies (Ben-Zion, 2001, 2006; Ben-Zion and Shi, 2005; Shi and Ben-Zion, 2006) emphasize the tendency for unilateral rupture to occur in bimaterials. Others (Andrews and Harris, 2005, 2006; Harris and Day, 1997, 2005) state that in realistic situations bilateral rupture will occur in the presence of material contrast, even though the propagation velocity differs in the two directions. There is agreement that if there is no reduction in friction with slip or slip velocity due, for example, to the other mechanisms discussed in this chapter, then the reduction in normal stress due to elastic mismatch can have significant effects on rupture directivity. However, it appears that if friction decreases significantly with increasing slip and/or slip velocity, then the bimaterial effects become relatively insignificant. Experiments on analog materials (Xia *et al.*, 2005) show that the ruptures propagate bilaterally, but with different velocities as predicted by theory (e.g., Harris and Day, 1997). Super-shear rupture propagation (i.e., in excess of the shear or Raleigh wave velocity) is also observed (Xia *et al.*, 2005), a result that is not necessarily related to elastic mismatch, but could be due to a large P wave generated by the explosion used to initiate rupture. Super-shear propagation is expected if the stress drop is large enough to produce a high amplitude P wave that induces rupture ahead of the shear wave front (Andrews, 1976). With more normal nucleation processes, the cause of a sufficiently large stress drop and P wave to produce super-shear rupture would be a significant decrease in friction with increasing slip and/or slip rate.

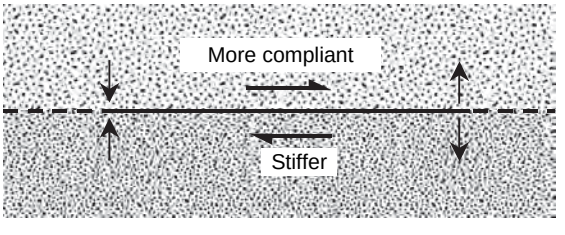


Figure 1 Schematic alterations in normal stress during dynamic rupture across a fault at a bimaterial interface between elastic materials with differing stiffness. Solid line shows section of fault currently slipping or that has slipped in this event; double arrows show fault plane ahead of rupture tip. Near the mode II rupture tips, the normal stress is reduced at the end toward which the more compliant material is slipping, and is increased at the end toward which the stiffer material is slipping. Arrows normal to fault show sense of alteration in normal stress due to this bimaterial effect.

Rudnicki and Rice (2006) discuss a different effect that can arise when a dynamic rupture propagates along a bimaterial interface (see **Figure 1**), namely the role that differences in permeability on opposite sides of a fault zone can have on the effective normal stress distribution due to changes in pore pressure. Pore pressure is increased in areas of the fault zone where the more permeable material is compressed and reduced where the more permeable material is extended. Since more slip may occur in the areas where the shear resistance is lowered due to

increased pore pressure, this mechanism might result in a net reduction in slip-integrated shear resistance. The magnitude of the changes in effective normal stress from this mechanism can be comparable to those due to elastic mismatch (Rudnicki and Rice, 2006). If both the elastic properties and the permeability differ on opposite sides of the fault, the reductions in effective normal stress from one mechanism may either add to or subtract from those due to the other mechanism.

In spite of all of these studies, it is not clear to what extent the frictional resistance during slip is reduced by elastic or permeability mismatch. One might infer that the effects are not large, since elastic mismatch has a smaller effect on the tendency for unilateral rupture propagation in the models than do modest changes in friction with slip (Andrews and Harris, 2005, 2006; Harris and Day, 1997, 2005). Furthermore, the reduction in normal stress and hence frictional resistance averaged over the total amount of slip will be less than that which occurs at the rupture tip, since much of the slip may not occur there, where the normal stress is reduced by the maximum amount. In order to determine an average effective reduction in friction, it would be necessary to integrate the product of slip and shear stress as a function of position behind the rupture tip. This does not appear to have been done in dynamic rupture models with bimaterial contrast and further work would be useful in this regard. In summary, these mechanisms could cause some reduction of dynamic shear resistance in situations where the elastic or permeability mismatch is large enough, but presently it appears that the effect will be small compared to that potentially available from other weakening mechanisms.

4.05.2.3 Acoustic Fluidization

Melosh (1979) proposed a new mechanism for the reduction of shear resistance inside a deforming granular aggregate, with possible relevance for impact crater collapse, landslides, and earthquakes. Later papers focused more specifically on the application of this mechanism to understanding crater collapse (Melosh and Ivanov, 1999), landslides (Collins and Melosh, 2003; Melosh and Girdner, 1995), and earthquakes (Melosh, 1996). In this mechanism, termed acoustic fluidization, Melosh envisions that acoustic waves reverberating inside the shearing aggregate could have sufficient intensity that they could exert large enough forces on the particles that they would be held apart or would touch with less force than would be inferred from the macroscopic stress applied to the boundaries of the aggregate and the real area of contact in the aggregate. If this were the case, then the aggregate could become fluidized and in this case would offer a reduced resistance to shearing. The process is envisioned to be self-sustaining because generation of the acoustic waves continues during shearing and so can balance the losses that will occur due to scattering out of the fault zone and to conversion of elastic energy into heat. Melosh (1996) shows that solutions to the relevant equations exist that allow this mechanism to operate, that they result in an essentially linear viscosity of the fluidized gouge (although later experiments showed highly nonlinear viscosity (Melosh and Girdner, 1995)), and that they give shear resistances that are nearly an order of magnitude smaller than the normal stress. However, this theory assumes that pressure fluctuations with frequencies greater than several kHz (Melosh, 1979) exist in the acoustic wave field and have magnitudes essentially as large as the normal stress.

While this is an interesting idea, whether such high-amplitude, high-frequency acoustic waves exist is not clear. Measuring them during seismic events would require instrumentation with high-frequency response located within a distance from the fault zone similar to the width of the shearing zone. If current thinking (e.g., see Sibson (2003) and discussion of Rice, (2006)) is correct, that the active fault shear zone in one seismic event may be thinner than 1–5 mm, perhaps even as thin as 0.2 mm, then, following Melosh (1979), the frequencies would have to be in the MHz range. In this case, the scale at which measurements would have to be made becomes impractical. Indeed, if slip is localized on such a fine scale it is possible that granular flow of a fluidized layer of gouge is not even relevant, with slip in the gouge more akin to slip on a plane. On the other hand, if slip is localized in a very narrow gouge zone, then relevant experiments are more feasible than would be the case if the thickness of the shearing zone is on the order of 1 m or more. Results of experiments exploring the relevance of this mechanism have not been published in detail, although some experimental results on masses of sand ~200 mm in dimension seem promising for application to landslides (Melosh and Girdner, 1995; Melosh, personal communication, 2006). Although this mechanism cannot be dismissed, at present its importance for earthquakes is best described as questionable.

4.05.2.4 Elastohydrodynamic Lubrication

Hydrodynamic lubrication is a well-known mechanism for reducing frictional resistance in various machine parts, notably journal bearings (Hamrock, 1994). In these applications the moving parts are immersed in oil or other fluids. The reduction in frictional resistance between the moving parts occurs because a layer of pressurized fluid separates the parts, preventing solid-to-solid contact. A high fluid pressure can be maintained by a balance between a tendency for pressure decrease due to fluid escape and a tendency for pressure increase due to the geometry of the moving parts. The balance only exists at high rates due to increased hydrodynamic resistance to fluid escape at such rates. Below some speed the viscous forces that resist the fluid escape are so small that the pressure cannot build up, but at higher speed the viscous resistance to escape is large and the parts become separated by the pressurized fluid.

Brodsky and Kanamori (2001) proposed a modification to this mechanism that might allow it to operate for a somewhat irregular sliding fault interface. The idea is illustrated in **Figure 2** in which a protrusion is assumed to exist on one side of the fault and a fluid resides in the space between the protrusion and a planar surface. When slip between the surfaces occurs at a sufficiently high velocity, pressure builds up ahead of the protrusion because the gap is closing there and viscous forces impede the fluid escape; this increased pressure tends to reduce the effective normal stress across the interface. In this geometry, the pressure behind the protrusion would be reduced, and so it would initially appear that, averaged over the entire surface, the net effect would be to make no change in the effective normal stress. However, the fluid pressures cause elastic distortion of the solid material and the resulting asymmetry results in a net increase in fluid pressure. Consequently, this process is termed elastohydrodynamic lubrication (Brodsky and Kanamori, 2001).

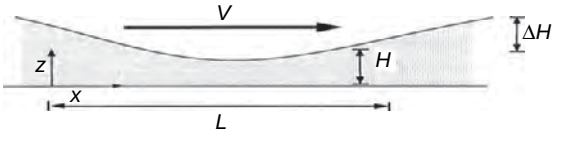


Figure 2 Geometry analyzed for hydrodynamic lubrication showing moving asperity with closing gap ahead and opening gap behind. From Brodsky EE and Kanamori H (2001). Elastohydrodynamic lubrication of faults. *Journal of Geophysical Research* 106: 16357–16374.

The geometry illustrated in **Figure 2** is an inadequate view of the interface because it is only two dimensional (2-D). It does show how a noncontacting asperity can lead to an increase in fluid pressure at the interface. However, for this mechanism to operate, many of the asperities must be in actual contact. Otherwise, if a continuous layer of fluid separates the surfaces, the shear resistance would be low without any need for this mechanism to cause weakening. The solution to this is to consider the third dimension, which allows for the presence of an interconnected fluid while at the same time some of the asperities are in solid-to-solid contact. In this case, the resistance could primarily come from these frictional contacts, but the normal and shear forces on them could be reduced by the elastohydrodynamically increased pressure in the interconnected fluid.

What increases in pressure might occur in the interface that could allow elastohydrodynamic lubrication? Brodsky and Kanamori (2001) give the excess lubrication pressure to be

$$P_{\mathrm{L}} \approx \frac{6\eta V L \Delta \mathrm{H}}{H^3}$$

where H is the average thickness of the gap, ΔH is the amplitude of the asperity, L is the wavelength of the asperity, V is the slip velocity, and η is the viscosity of the fluid (see **Figure 2**). To illustrate, assuming $V = 1\,\mathrm{m\,s^{-1}}$, $H = 20\,\mu\mathrm{m}$, $L = 1\,\mathrm{mm}$, and $\Delta H = 10\,\mu\mathrm{m}$, we find $P_{\mathrm{L}} = 7.5\eta\,\mathrm{MPa}$ when η is given in Pa s. The viscosity of water at 20°C is $\sim 10^{-3}$ Pa s (Sengers and Watson, 1986), so for water, even with gaps as small as 20 μm, $P_{\mathrm{L}} = 0.0075\,\mathrm{MPa}$, a negligible value compared to values of pore pressure arising from a hydrostatic gradient of $\sim 10\,\mathrm{MPa\,km^{-1}}$. If the calculation is done for even larger values of L and ΔH, the value of H presumably increases in proportion, and, due to the inverse dependence on H^3, P_{L} consequently gets even smaller.

What fluids with higher viscosity might allow this mechanism to cause significant lubrication pressures? Brodsky and Kanamori (2001) suggest a slurry of water and fine-grained gouge for which Major and Pierson (1992) measure the viscosity to be in the range of 0.2–7 Pa s at atmospheric pressure. However, in these experiments the effective pressure is zero since the fluid pressure is equal to the confining pressure. If this were the case during faulting, then no weakening via this mechanism is needed since the effective normal stress on the fault surface would be zero and consequently the frictional stress would be zero as well. At elevated effective pressures

all rock mechanics experiments in the presence of water suggest that a water-saturated gouge behaves as a Coulomb friction material, not as a viscous fluid (e.g., Byerlee, 1978; Zhang and Tullis, 1998). None of these experiments have been done at extremely high deformation rate, so it is possible that the behavior somehow could change from that of a Coulomb to that of a viscous fluid at high rates. There would not be time for the pressurized pore fluid to equalize throughout the fluid-saturated gouge, but this might only make it act as a Coulomb material at lower effective pressure, not as a fluid. There is certainly no evidence that fluid saturated gouge does act as a viscous fluid, so the operation of elastohydrodynamic lubrication via a slurry of water and gouge at elevated effective pressure seems unlikely. Presumably melt with a viscosity of ~10 Pa s (Spray, 1993) is a viable fluid, as might be the silica gel discussed in a later section. If a layer of relatively weak material such as melt or gel is present it is unclear whether this mechanism would reduce resistance further. If the weak layer is not very thick, so that it does not completely separate the opposing surfaces, a further reduction in shear resistance by elastohydrodynamic lubrication could be important. On the other hand, parts of the frictional melting experiments of Hirose and Shimamoto (2005) involve only an incomplete melt layer, and during this interval the frictional resistance is increasing (see discussion in Section 4.05.2.5). The resistance in these experiments seems well described by simple shearing of a viscous layer. In the melt containing experiments of Hirose and Shimamoto (2005) and Tsutsumi and Shimamoto (1997a), over the velocity range investigated, from 0.1 to 1.8 m s^{-1}, no evidence was found of a dramatic decrease with velocity as expected for elastohydrodynamic lubrication (i.e., no Stribeck curve (Brodsky and Kanamori, 2001; Hamrock, 1994) was observed). It appears that few relevant experiments exist and for the few that do, none demonstrate the operation of this mechanism. On balance, although it is possible that this mechanism could reduce coseismic frictional resistance, there is presently no reason to believe that it is an important mechanism during earthquakes.

4.05.2.5 Thermal Pressurization of Pore Fluid

Weakening results from a transient increase in pore pressure, arising from frictional heat generation and thermal expansion of the pore fluid, resulting in a reduction of the effective normal stress (Andrews, 2002; Goldsby and Tullis, 1997; Lachenbruch, 1980; Lee and Delaney, 1987; Mase and Smith, 1985, 1987; Rempel and Rice, 2006; Rice, 2006; Rice and Cocco, 2007). In order for this mechanism to be effective, the slip must be sufficiently localized that the heat is input into a relatively narrow fault zone, and rapid enough that there is minimal flow of either heat or pore fluid away from the fault. Observations of exhumed faults (Chester and Chester, 1998; Chester *et al.*, 2004, 1993; Lockner *et al.*, 2000; Noda and Shimamoto, 2005; Sulem *et al.*, 2004; Wibberley and Shimamoto, 2003, 2005) show that at least in some cases the zone of slip is very narrow, on the order of 100–300 μm (Chester and Chester, 1998), sufficient to meet the localized dissipation requirement. The constraint of low heat flow at seismic slip rates is not difficult to meet, given the low thermal conductivity of rock. However, whether the thermally expanded fluid can pressurize depends critically on the permeability of the fault zone and surrounding country rock. Given the wide range in permeability of rocks, a wide range of behavior is possible. Existing data on permeability of rocks in the vicinity of fault zones (Seront *et al.*, 1998; Wibberley, 2002; Wibberley and Shimamoto, 2003) and of experimental faults (Morrow and Lockner, 1997; Morrow *et al.*, 2001, 1984; Zhang *et al.*, 1999, 2001; Zhang and Tullis, 1998) suggest that permeabilities can be low enough to allow significant pressure buildup. Another requirement for this mechanism to be effective is that the dilatancy of the fault is minimal as slip progresses. Otherwise the pore volume occupied by the fluid could increase due to dilatancy and provide enough space for the thermally expanding fluid that no pressure increase would occur. Some initial dilatancy does occur in rock friction experiments when slip velocity is increased (Beeler and Tullis, 1997; Marone *et al.*, 1990) and implications of this for stability of slip in the absence of frictional heating have been studied (Segall and Rice, 1995). This initial dilatancy has been included in models of thermal pressurization (Rice, 2006), but it is possible that dilatancy may continue with slip due to the nonplanarity of mated fault interfaces (Junger and Tullis, 2003; Power and Tullis, 1992). Further investigation of dilatancy is needed to determine whether it may neutralize thermal pressurization.

The thermal pressurization mechanism is relatively well understood from a theoretical perspective, the most recent analysis being that of Rice (2006). He shows that for slip on a plane and at constant slip

velocity V and for constant coefficient of friction f, the shear resistance τ is given by

$$\tau = f(\sigma_n - p_0)\exp\left(\frac{\delta}{L^*}\right)\mathrm{erfc}\left(\sqrt{\frac{\delta}{L^*}}\right)$$

where δ is slip, σ_n is normal stress, p_0 is the pore pressure at the initiation of slip at velocity V following any initial dilatancy, and L^* is a length parameter proportional to $1/(Vf^2)$ with a proportionality constant that contains all of the porothermoelastic properties of the solid and fluid. L^* is given by

$$L^* = \frac{4}{f^2}\left(\frac{\rho c}{\Lambda}\right)^2 \frac{\left(\sqrt{\alpha_{hy}} + \sqrt{\alpha_{th}}\right)^2}{V}$$

where ρc is the specific heat of the fault zone material, α_{hy} and α_{th} are the hydraulic and thermal diffusivities, and Λ is an undrained pressurization factor given by $\Lambda = (\lambda_f - \lambda_n)/(\beta_f - \beta_n)$, where λ_f and λ_n are the isobaric volumetric fluid and pore space thermal expansivities, β_f the isothermal compressibility of the fluid, and β_n the pressure expansivity of the pore space. For a wide range of porothermoelastic parameter values, L^* ranges in magnitude from 2 to 50 mm. The function for τ is such that τ continues to decay with increasing slip, albeit at a decreasing rate, so that the decay cannot be characterized by a slip weakening length scale. This behavior is shown in **Figure 3**. Rice (2006) presents a 'low-end' and a 'high-end' combination of parameters that result in L^* equal to 4 and 30 mm, respectively, when $V = 1\,\mathrm{m\,s^{-1}}$ and $f = 0.25$, a value appropriate for flash heating (see discussion below). With these parameter values, the full displacement scales in the four panels of **Figure 3** range from 0.4 mm to 0.4 m for the low-end parameters and from 3 mm to 3 m for the high-end parameters, suggesting that the resistance to slip should decrease to 0.05 times its initial value within 0.4–3 m of slip. Since this is for an initial value of the friction coefficient for flash heating of $f = 0.25$, the effective friction coefficient (τ/σ_n) would be as low as 0.01 after 0.4–3 m of slip. In the absence of weakening by flash heating, the amount of frictional heating would be greater, so the displacement needed to attain a given amount of weakening is smaller. Thus if the initial friction coefficient $f = 0.6$, the above displacements of 0.4–3 m would be smaller by a factor of 0.17 (Rice, 2006), namely 68–510 mm for the low- and high-end parameters, and the effective friction coefficient would have decreased to 0.03.

Experimental confirmation of this mechanism has yet to be accomplished, although it is amenable to such study. Weakening observed in an initial attempt at exploring this mechanism in the laboratory (Goldsby and Tullis, 1997) was later found to be due to the gel-weakening mechanism described in a later section. However, since some rocks do not show weakening due to gel, it should be possible to revisit the thermal pressurization problem experimentally (Tullis *et al.*, 2007).

This mechanism remains a viable one for significant reductions in strength of faults during earthquakes. More study is needed to investigate whether the theoretical predictions are borne out by experiments and to what extent dilatancy may counteract the predicted increase in fluid pressure. If dilatancy is unimportant, with sufficient slip the shear resistance due to thermal pressurization can approach zero as shown in **Figure 3** from Rice (2006) and, for example, by Mase and Smith (1985).

4.05.2.6 'Flash' Heating/Melting at Asperity Contacts

Contact across frictional interfaces occurs over only a fraction of the nominal area of contact, and conditions at the contact spots are quite different from those on the surface as a whole. The generally accepted explanation for friction, namely the linear

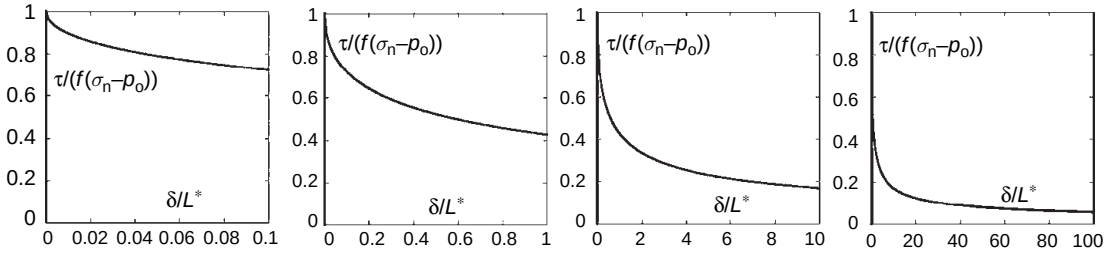


Figure 3 Progressive weakening with slip (Rice, 2006) due to thermal pressurization resulting from slip on a plane at constant slip velocity with constant coefficient of friction. Successive frames show that a characteristic decay slip distance does not exist for this mechanism. See text and Rice (2006) for meaning and values of L^*.

increase of shear resistance to sliding with increases in normal force, is that the total real area of contact increases in proportion to increases in the nominal normal stress. The real area of contact occurs where high points on the opposing surfaces called asperities touch each other, and the shear resistance to sliding arises from the average shear strength of the contacts τ_c, times the total real area of contact. The average normal stress at the contacts is much higher than the nominal normal stress by the ratio of the nominal contact area to the real area of contact. It is often assumed that the normal stresses at the contacting asperities are equal to the compressional yield strength of the material, although in hard materials the contacts may be elastic and not reach the yield stress (Johnson, 1985). During sliding the contacts are short-lived as high points on each surface move past one another, but the number and size of contacts in the population is independent of slip at constant slip velocity, assuming no fracture/comminution occurs. The lifetime θ of a representative contact is equal to the length D of the contact in the slip direction, divided by the slip velocity V.

Locally elevated temperatures occur at contacts and can result in degradation of contact strength and therefore reduction in macroscopic friction. The rate of heat dissipation at a point within the contact is the product of V and the contact shear stress τ_c. Because τ_c is much larger than the nominal macroscopic shear stress (by the ratio of the nominal area of the surface to the real area of contact), the product $\tau_c V$ is much larger than the average heat dissipation rate on the frictional interface. If V is sufficiently high, the local temperature at the contacts is much greater than the average temperature on the surface. Such elevated temperatures were measured for dissimilar metals in sliding contact using the thermocouple effect (Bowden and Ridler, 1936), and for transparent materials by studying visible (Bowden *et al.*, 1947) and infrared (Bowden and Thomas, 1954) radiation. Macroscopic weakening occurs due to thermal degradation of contact strength and at sufficiently high V melting may occur. The process is often called 'flash' heating, after small 'stars of light' observed during sliding of transparent materials (Bowden *et al.*, 1947). In cases where melting occurs, the process is called 'flash' melting.

Theoretical analyses of flash heating and weakening have been undertaken by many authors (Archard, 1958; Beeler and Tullis, 2003; Beeler *et al.*, 2007; Blok, 1937; Bowden and Persson, 1960; Bowden and Ridler, 1936; Bowden and Thomas, 1954; Ettles, 1986; Jaegar, 1942; Lim and Ashby, 1987; Lim *et al.*, 1989; Molinari *et al.*, 1999; Rice, 1999, 2006). Rice (1999) conducted an analysis of the weakening velocity and how the steady-state frictional resistance should depend on slip velocity at velocities above the weakening velocity. In Rice's analysis (Rice, 1999), since expanded (Rice, 2006), once a part of an asperity exceeds a weakening temperature T_w, it has zero strength. He obtains the weakening time θ_w, namely the time that the oldest part of an asperity contact exists before it reaches the weakening temperature, by equating the heat input (namely the heat input rate $\tau_c V$ times the age of the oldest part of the asperity $\theta = D/V$) to the thermal energy storage $\rho c(T_w - T_f)$ times an effective distance $(\pi c_{th}\ \theta)^{1/2}$. That effective distance is obtained from a 1-D heat flow analysis, where ρ is the material density, c is its heat capacity, c_{th} is the thermal diffusivity, and T_f is the present average temperature of the fault. Solving this for V and regarding it as the velocity V_w at which weakening first occurs, Rice obtains

$$V_w = (\pi c_{th}/D)[\rho c(T_w - T_f)/\tau_c]^2$$

Beeler and Tullis (2003) and Beeler *et al.* (2007) extended Rice's (1999) analysis to include a distribution of contact sizes, allowed the weakened part of the contact to have nonzero strength, and showed how the strength evolves with slip over small slip distances on the order of the contact spot size. Beeler *et al.* (2007) show that Rice's $1/V$ dependence of friction on V above the weakening velocity V_w results from variations with velocity of the contributions of the weakened older parts of a typical asperity contact (those parts older than θ_w) and the unweakened younger parts. The result of these analyses is that the friction coefficient $f = f_0(V_w/V)$, where f_0 is the friction coefficient at velocities below V_w. Appropriate values for all of the parameters in Rice's expression for V_w are not well known. The better-known values are c_{th} (1–2 $mm^2\,s^{-1}$), and the specific heat per unit volume ρc (3–4 $MJ\,m^{-3}\,°C^{-1}$). Less well known are D, which depends on the roughness of the contacting surfaces, T_w, which depends on which process causes weakening (thermal softening or melting), and τ_c, which depends on whether the contacts are at their compressive yield strength, and whether their shear strength depends on flow or fracture of the crystalline material, on melting of the contact, or on the strength of some other weak layer coating the asperities. Reasonable parameter values are $D = 5$–$20\,\mu m$, $(T_w - T_f) = 1000$–$1700\,K$, and $\tau_c = 5$–$10\,GPa$. While acknowledging the

uncertainty involved, a frequently employed estimate for V_w is 0.1 m s^{-1} (Rice, 2006).

As discussed below, data from two independent experimental studies are consistent with flash weakening, although further work is required to be more confident of the underlying mechanism. Verification of this mechanism can be accomplished by conducting experiments at elevated temperature to see if shifts in the predicted V_w with changes in $(T_w - T_f)$ match observed shifts in V_w with changes in $(T_w - T_f)$. Similarly, varying D by varying surface roughness should yield predictable shifts in V_w, according to Rice's expression for V_w above.

In unconfined rotary shear experiments by Goldsby and Tullis (2003) and Tullis and Goldsby (2003), an abrupt weakening is seen at slip velocities above 0.1 m s^{-1}, the independently predicted theoretical value of V_w. The weakening occurs after a small slip displacement of about 1 mm. That this is a larger value than contact spot sizes is expected due to the inertia of the testing machine. After the velocity is reduced below 0.1 m s^{-1} frictional strength returns to its typical quasi-static value within about 0.01 s (~2 mm of slip). These results are shown in **Figures 4** and **5** for a fine-grained pure quartz rock, novaculite. Similar behavior has been seen for granite, gabbro, and a pure feldspar rock (Roig Silva *et al.*, 2004b). Preliminary attempts to observe textural evidence on the slip surfaces that might support the flash-weakening mechanism (e.g., melted asperity tips) have revealed nothing diagnostic, although such evidence might be difficult to locate. All the characteristics of the mechanical behavior agree with expectations for flash weakening, as the fits to the data show in **Figure 5**. Further tests in which D and T_f are varied should allow a definitive conclusion to be drawn concerning whether flash weakening is responsible for the observed behavior.

Two other experimental techniques new to the geophysical research community yield low friction for rocks consistent with flash heating at small displacements and slip rates above the theoretically predicted V_w (Prakash and Yuan, 2004). Results for novaculite are shown in **Figures 6** and **7**. In **Figure 6** data from a pressure–shear experiment are shown, whereas in **Figure 7** data from a torsional Kolsky bar experiment are shown. A description of the pressure–shear and torsional Kolsky bar techniques is given by

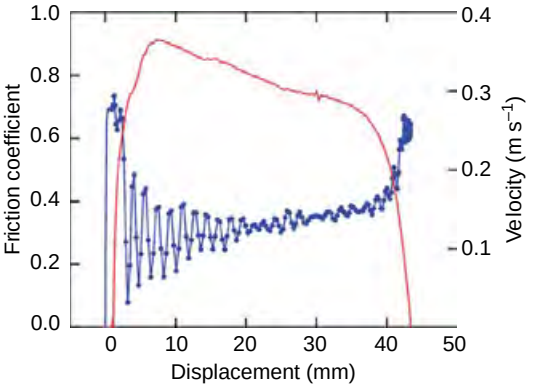


Figure 4 Friction (blue) and slip velocity (red) on novaculite vs. slip at normal stress of 5 MPa in unconfined experiment. Oscillations represent machine–sample interactions, not the intrinsic frictional behavior.

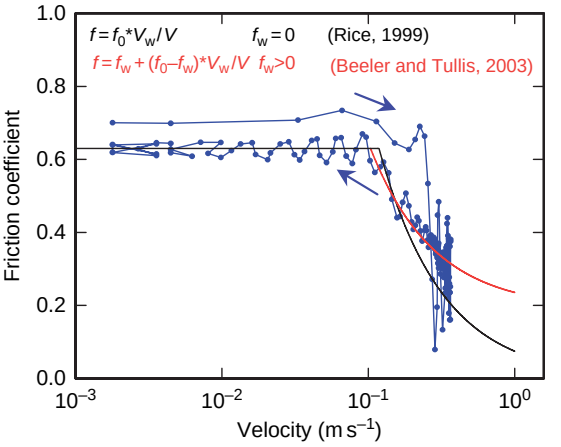


Figure 5 Data of **Figure 4** plotted as friction vs. velocity. Arrows show time sequence of data. Smooth black and red lines show predictions of theory of Rice (1999) and of Beeler and Tullis (2003) and Beeler *et al.* (2007), respectively.

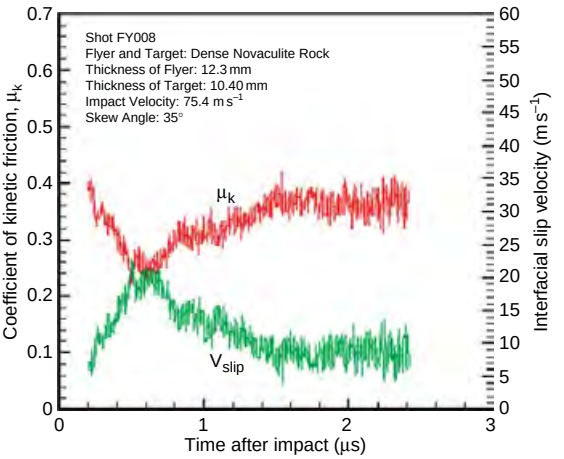


Figure 6 Friction (red) and slip velocity (green) vs. time for pressure shear experiment on novaculite. From Prakash V and Yuan F (2004).

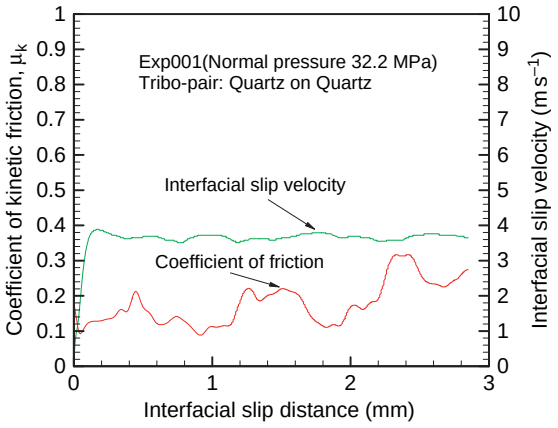


Figure 7 Friction and slip velocity for torsional Kolsky bar experiment on novaculite. From Prakash and Yuan (2004).

Prakash and Clifton (1993) and Rajagopalan and Prakash (1999), respectively. In both types of experiments, dramatic reductions in shear resistance occur with small slip displacements, to levels consistent with flash weakening. In these pilot experiments, designed in part to demonstrate the feasibility of using these dynamic testing techniques on fully dense rocks, there was no systematic attempt to vary the sliding velocity or to determine the dependence of friction on velocity. However, the techniques show promise for investigating weakening at high slip speeds (1–50 m s^{-1}) and modest slip displacements (0.5–5 mm).

Although flash weakening needs to be studied in more detail in the laboratory, it appears to be a strong candidate for causing abrupt and dramatic weakening during earthquakes. Under conditions in the upper crust where frictional behavior predominates, and consequently resistance to earthquake slip occurs over small real areas of contact on fault surfaces, it seems likely that this mechanism would be an important one.

Two factors, however, might limit the significance of the flash-weakening mechanism. One concerns the magnitude of the shear stress at contacting asperities and the other concerns the thickness of the fault zone across which the shear displacement takes place. If clay minerals dominate the fault zone, the shear stress τ_c might be considerably smaller than 5–10 GPa, increasing V_w to slip rates above those experienced seismically. Also, for flash heating calculations, it is often assumed that τ_c is equal to the theoretical strength, which is on the order of a mineral's shear modulus divided by ~30 (Cottrell, 1965; Hull, 1965), and this seems to be borne out by some experiments (Dieterich and Kilgore, 1994, 1996). On the other hand, in strong brittle materials, frictional contacts can elastically deform rather than plastically yield (Johnson, 1985), so that τ_c could be smaller than often assumed. The other factor is that, even with τ_c approaching the theoretical strength for most silicates, if shearing is distributed over many contacts in a sufficiently thick deforming layer of fault gouge, the slip velocity at every contact in the gouge could be less than V_w even if the total slip velocity across the zone was 1 m s^{-1} or higher. For example, if shearing occurred in a deforming gouge layer only 10 grains thick, the slip velocity at each contact might be reduced to 0.1 m s^{-1} even with a total slip velocity of 1 m s^{-1}, which might defeat the mechanism. These possibilities can be studied experimentally, one by rapid shearing clay minerals (which have yet to be studied in rapid slip-rate experiments), and the other by using gouge layers of varying thickness. Although it is notable that slip often tends to become very localized in gouge experiments (Beeler *et al.*, 1996; Scruggs and Tullis, 1998), as can also occur on natural faults (Chester and Chester, 1998; Chester *et al.*, 2004, 1993; Lockner *et al.*, 2000; Noda and Shimamoto, 2005; Sulem *et al.*, 2004; Wibberley and Shimamoto, 2003), localization is typically less narrow than 10 grain diameters (see discussion by Rice (2006)). How localized the deformation is in the experiments of Goldsby and Tullis (2003) and Tullis and Goldsby, (2003) is presently unclear. The experiments involve initially bare rock surfaces and the assumption is that gouge is not generated in significant thicknesses. However, some powdered rock escapes from the edges of the ring-shaped sliding surface. How much generated gouge separates the bare surfaces in these experiments during sliding, and how this might affect the observed weakening, are unknown and warrant further study.

4.05.2.7 Interfacial Lubrication by Friction Melt

If sufficiently large slip occurs at high slip velocity and with high frictional resistance, the temperature of the entire fault surface or zone can become high enough that a layer of melt develops and lubricates the entire fault surface. Melting caused by faulting was suggested as a dynamic fault-weakening mechanism by Jeffreys (1942) and was suggested as an explanation for pseudotachylytes by Goldschmidt

(1943). The origin and significance of pseudotachylytes and/or the reduction of shear resistance by shear melting has been discussed by many other authors (e.g., Bjornerud and Magloughlin, 2004; Di Toro *et al.*, 2006, 2005a, 2005b; Di Toro and Pennacchioni, 2004, 2005; Ferre *et al.*, 2005; Fialko, 2004; Fialko and Khazan, 2005; Kanamori and Heaton, 2000; Maddock, 1986; Magloughlin, 1989, 1992; Magloughlin and Spray, 1992; Mase and Smith, 1987; McKenzie and Brune, 1972; Philpotts, 1964; Rempel and Rice, 2006; Sibson, 1975; Spray, 1989a, 1989b, 1992, 1995, 2005; Spray and Thompson, 1995; Swanson, 1988, 1990, 1992; Wenk, 1978, 1979; Wenk *et al.*, 2000). Two special volumes of *Tectonophysics* were devoted to these rocks (1992, vol. 204; 2005, vol. 402). The most recent and thorough theoretical treatment is that of Fialko and Khazan (2005). In the absence of fluids, or for a fluid saturated fault of high permeability, the theory predicts a progression with slip, beginning with flash heating at asperity contacts, to a situation where the overall temperature of the fault T_f gradually increases to the point that it reaches the melting temperature. With continued input of heat, the heat of fusion is provided to more and more material in the fault zone resulting in a layer of melt along the fault surface. Continued slip results in viscous shearing of this melt layer. Although the melt layer may be significantly weaker than the initial frictional interface, thus lowering the heat input rate, continued heat may be generated by viscous dissipation in the shearing layer, perhaps along with frictional heating occurring where high points on opposite sides of the interface project through the melt layer and contact. It is known that such frictional melting occurs during some earthquakes because pseudotachylytes are observed on some faults, although they are often believed to be less abundant than expected (Sibson and Toy, 2006). Questions have been raised concerning whether all pseudotachylytes are the product of melting (Wenk, 1978), so it is best to reserve the term for rocks containing definite evidence of melting (such as presence of microlites, spherulites, clasts with embayments, amygdules, etc.), and use another term such as ultracataclasites for rocks with similar macroscopic morphology and occurrence but for which definite evidence of melting cannot be demonstrated (even though some of them may also have resulted from melting).

Although much more experimental work needs to be conducted, a series of informative experiments on shear melting have been conducted by Spray (1987, 1988, 1993) and Shimamoto and colleagues (Di Toro *et al.*, 2006; Hirose, 2002; Hirose and Shimamoto, 2003, 2005; Hirose *et al.*, 2000; Lin and Shimamoto, 1998; Shimamoto and Lin, 1994; Shimamoto and Tsutsumi, 1994; Tsutsumi and Shimamoto, 1994, 1996, 1997a, 1997b). An analysis of the observed eventual weakening in the experiments of Hirose and Shimamoto (2005) has been provided by Fialko and Khazan (2005). It is apparent from these experiments that shear melting indeed occurs at sufficiently high slip velocity and shear stress. Because these experiments have all been done without confining pressure, the normal stress is small to prevent fracturing. Therefore, the frictional shear stress prior to melting is lower than for earthquake generating faults. Thus, the slip displacements in the experiments are higher (up to 174 m) than during earthquakes in order to generate sufficient heat and raise the temperature high enough to cause melting. Since the samples are unconfined, it is not possible for any melt pressure to build up in the samples. In fact, the melt is thrown radially out of the rotary-shear-configuration samples by centrifugal force. This resembles the situation often seen in occurrences of pseudotachylytes in which only thin layers of pseudotachylyte are found along the pseudotachylyte generation zones along the faults and reservoirs of injected pseudotachylyte are found in veins connected to the generation zones.

Study of the evolution of shear resistance in these experiments (Hirose and Shimamoto, 2005) shows a complex sequence that can be related to the evolution of the microstructures in the samples. The initial mechanical response is a weakening that appears to be due to flash melting, followed by a strengthening that appears to be due to the development and growth of melt patches on the sliding surface which result in a larger real area of contact than is the case for frictional sliding prior to development of melt patches. Significant viscous resistance apparently occurs at the melt patches. The final development is a gradual weakening, which accompanies the creation of a continuous melt layer, and continues with the growth in thickness of the melt layer (Hirose and Shimamoto, 2005) and an increase in the melt fraction in the melt–clast mixture on the fault with increasing temperature (Fialko and Khazan, 2005). This sequence of correlated mechanical and textural developments is shown in **Figure 8**.

The shear resistance with a layer of melt is not as low in the experiments in Shimamoto's rotary shear

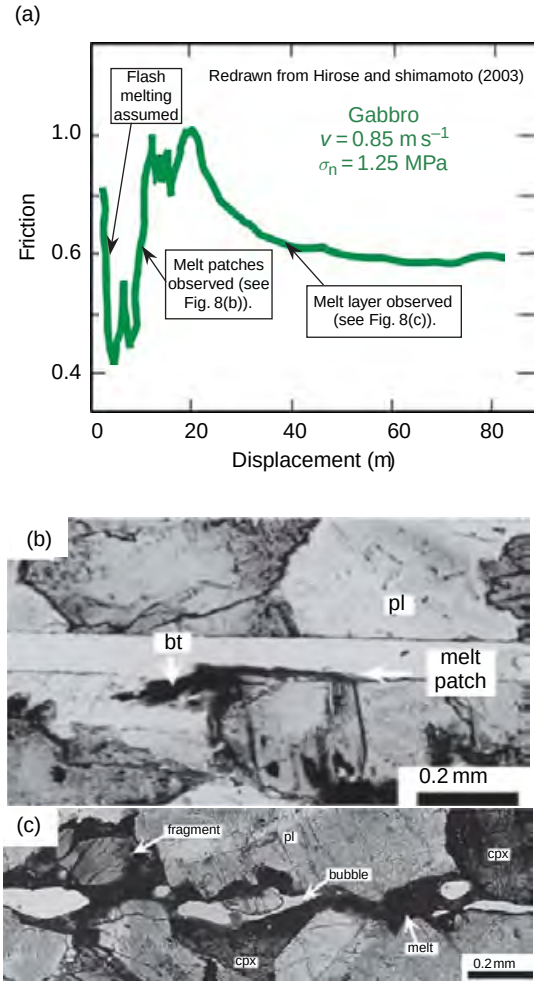


Figure 8 Results from friction experiments of (Hirose and Shimamoto, 2005, their Figures 2a, 5a, and 5e). (a) Friction vs. displacement, with initial weakening assumed due to flash melting, subsequent strengthening assumed due to increases in contact area as melt patches illustrated in (b) grow, and final weakening as melt layer illustrated in (c) thickens.

apparatus (Hirose and Shimamoto, 2005; Tsutsumi and Shimamoto, 1997a) as one might expect, namely the coefficient of friction at what appears to be nearly steady state is 0.6, a value typical of rock friction without melt in experiments with slip velocities of $1\,\mathrm{mm\,s^{-1}}$ or lower. The initial weakening shown in **Figure 8**, attributed to flash melting, results in a coefficient of friction of about 0.4, a value roughly consistent with flash melting, but perhaps slightly too high (Beeler and Tullis, 2003; Beeler *et al.*, 2007; Goldsby and Tullis, 2003, 2007a). The subsequent increase in friction to a value of the coefficient of friction of about 1.0 and the eventual decrease to values of about 0.6 do not appear to demonstrate significant weakening. However, expressing this steady-state shear resistance as a coefficient of friction is misleading. The data shown in **Figure 8** were collected at a normal stress of only 1.5 MPa, so the shear resistance is only 0.9 MPa. If this resistance is due to viscous shearing and if, as expected, the viscosity shows little or no dependence on pressure, then the shear resistance at higher pressures of perhaps 180 MPa at a depth of 10 km in the Earth might also be only 0.9 MPa. If so, the apparent coefficient of friction would be only 0.005 and the weakening would be extreme. Experiments conducted in Shimamoto's lab to somewhat higher normal stresses (up to several tens of MPa) seem to confirm this expectation, namely that the shear resistance once steady state is approached depends only weakly on normal stress (Di Toro *et al.*, 2006). This is also expected from theoretical considerations (Fialko and Khazan, 2005). Although difficult to design and conduct, experiments at elevated confining pressure in which the melt cannot escape from the fault surface would be very valuable to explore further the weakening caused by macroscopic shear melting.

The presence of frozen friction melts (pseudotachylytes) demonstrates that melting occurs during some earthquakes. The presently accepted view of many investigators is that pseudotachylytes are much less abundant than expected, leading to speculation about the cause of this apparent paucity. Expectations of ubiquitous pseudotachylyte are based on the assumption that earthquakes slip at velocities of about $1\,\mathrm{m\,s^{-1}}$ and occur at depth of a few km with coefficients of friction of ~0.6, resulting in copious amounts of frictionally generated heat (e.g., Kanamori and Heaton, 2000). Of course, one possibility is that one or more of the other weakening mechanisms discussed in this paper reduces the frictional resistance during earthquakes to such low values that insufficient heat for melting is generated on most faults.

It is also possible that pseudotachylytes are often generated, but are either seldom preserved or simply seldom recognized. Some characteristics, especially the presence of glass or clearly discernible devitrification textures, are not likely to persist over geologic timescales, since recrystallization or deformation events can obliterate such evidence. Recrystallization should not eliminate the macroscopic morphology of pseudotachylyte veins, but it might obscure the presence of layers of pseudotachylyte on their generation surfaces unless the layers are thicker than typically

observed. Deformation certainly has a good chance of destroying evidence of pseudotachylyte in a fault zone, since subsequent shearing in the fault zone is to be expected. On the other hand, a layer of pseudotachylyte can weld a fault surface together, making it stronger than before. Consequently, subsequent deformation may occur elsewhere, especially in an immature fault zone, preserving the earlier-formed pseudotachylyte adjacent to more recently active fault strands. Pseudotachylyte may be more common than is typically reported, but seldom recognized, since in many rocks they are easy to overlook. Black pseudotachylyte in white granite can be easily spotted, but in many dark rocks they can be hard to discern unless the exposures are good and the observer is experienced. They are more likely to be generated in fault zones where the normal stress and hence the shear stress is high. They might also be expected to be more common in major fault zones where the slip displacements are high. Thus, they may be more common in major fault zones at depths of 5–10 km than in shallower faults. Unfortunately, there are not many exposures of major faults exhumed to this depth, especially those that have not been overprinted by later shallower slip.

Suffice it to say that the presence of pseudotachylytes on some faults shows that shear melting sometimes occurs during earthquakes, and the available experimental evidence at low normal stress shows that interfacial lubrication by frictional melt can be an efficient weakening mechanism. In addition, energy balance calculations in natural pseudotachylytes based on thickness vs. displacement relationships (Di Toro *et al.*, 2006; Sibson, 1975) suggest that shear stress could be low in the presence of seismic melts. This is the only high-speed weakening mechanism that operates with certainty during at least some earthquakes and it warrants continued study.

4.05.2.8 Interfacial Lubrication by Thixotropic Silica Gel

This weakening mechanism was discovered in the laboratory by Goldsby and Tullis (1998, 2002) and further investigated by Titone *et al.* (2001) and Di Toro *et al.* (2004). A paper describing the role of water and the resulting textures is in preparation (Goldsby and Tullis, 2007b). Extraordinary weakening that could not be explained by any other known mechanism was observed in a variety of experiments at relatively rapid slip rats and earthquake-like sliding displacements, eventually leading to the postulation of this mechanism. The mechanism involves the generation of a layer of silica gel on a fault surface by interaction of silica in the sheared fault interface with water from the environment, either liquid water in the pore space or in fluid inclusions of the rock, or water from the atmospheric humidity of the test environment. Silica gel is a noncrystalline, nonstoichiometric mixture of silica and water (Iler, 1979). The weakening observed in the experiments only occurs when the slip rate is high, above 1 mm s^{-1}, and the slip displacement is large, greater than several hundred mm (**Figure 9**). The steady-state friction coefficient drops from ~0.6 to ~0.25 as the slip velocity increases from 1 to 100 mm s^{-1} as is shown in **Figure 10**. Following cessation of slip at speeds above 1 mm s^{-1} the friction coefficient recovers to ~0.6 as a function of time over a time interval from 300 s (**Figure 9**) to a few thousand seconds, depending on the experimental details. As is illustrated in **Figure 11**, SEM examination of the sliding surface shows a layer about 3 μm thick that was clearly soft and flowed during the experiments, but is brittle after the experiments. These observations can be explained if the silica gel behaves as a thixotropic material (Shaw, 1980), a behavior that has been documented independently for silica gel (Kalinin *et al.*, 2002; Serdobintseva and Kalinin, 2001).

The reason that silica gel can be weak when it experiences large strain at a high rate and strong when is it deformed slowly is that its strength depends on a competition between strain-dependent weakening and time-dependent strengthening (Ferguson and Kemblowski, 1991). Bonding between

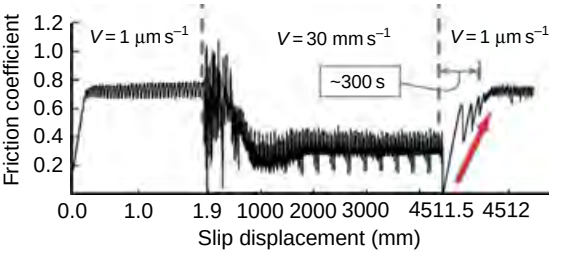


Figure 9 Friction vs. displacement at low and high speed (Di Toro *et al.*, 2004, Fig 1a). The oscillations at 1 μm s^{-1} are due to stick slip behavior, whereas those at 30 mm s^{-1} are artifacts due to rectification of frequent reversals of sliding direction. This high-speed weakening due to gel lubrication requires a slip displacement of 0.5–1 m, whereas strength recovery at either low or zero velocity occurs as a function of time, not slip displacement.

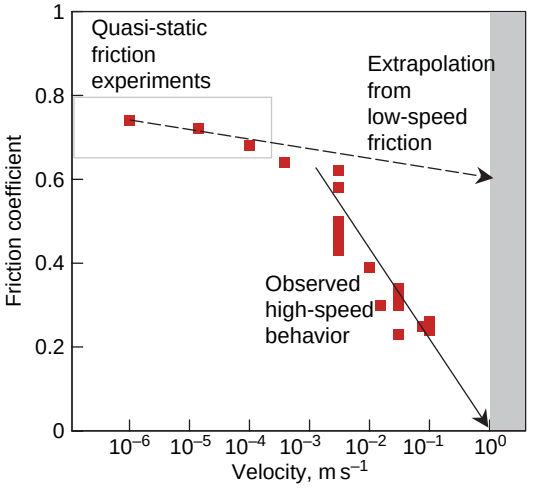


Figure 10 Friction vs. slip velocity. The dramatic weakening above 1 mm s^{-1} is due to the formation of a thixotropic layer of silica gel, illustrated in **Figure 11**. The large reduction in resistance from gel weakening is in contrast to the extrapolated weakening expected from rate and state friction at low speed. Gray area shows typical seismic slip rates. Modified from fig. 3 of Di Toro *et al.* (2004).

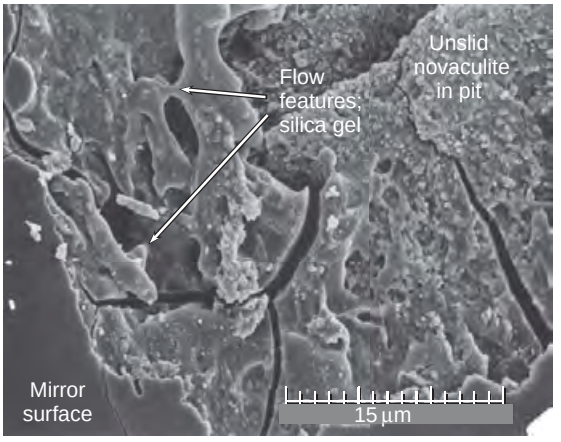


Figure 11 SEM micrograph of portion of surface of novaculite sample slid 63 m at 3 mm s^{-1} in rotary shear experiments similar to those of Di Toro *et al.* (2004) and Goldsby and Tullis (2002). The field of view is at the edge of one of the many small pits in the surface that result from porosity in the sample. Silica gel has been smoothed over most of the surface and at the edges of the pits it shows flow features as it is smeared over the edge of the pit. The illustrated features show the gel had the ability to flow, but the fractures show that subsequently its viscosity greatly increased.

adjoining SiO_2 entities is disrupted by strain and this may be aided by the presence of water that can hydrolyze Si–O bonds. With no strain, the bonding is strengthened in a time-dependent way, perhaps accompanied by a rearrangement of the water. In the silica literature (e.g., Iler, 1979) the SiO_2 entities are described as being particles, similar to the µm to sub-µm sized particles that make up the regular structure of opal (Graetsch, 1994; Serdobintseva *et al.*, 1998). However, it also seems possible that the SiO_2 entities also could be individual SiO_4^{-4} tetrahedra or groups of tetrahedra (Khavryuchenko and Sheka, 1994) and that the differences between the structure of the rapidly deformed silica gel and the slowly deformed silica gel might be the degree of polymerization, similar to the difference between the structure of a wet and a dry granite melt, respectively (Baker, 1998; Burnham, 1979; Shaw, 1972).

Weakening due to silica gel clearly depends on the presence of SiO_2 in appreciable amounts in the rock in the fault zone. Experiments employing rapid slip rates and large slips show that the weakening occurs, not just for pure quartz rock, but also for silicate rocks with less silica. Gabbro, having only ~50 wt.% SiO_2, does not show the weakening behavior, but granite as well as an albite feldspar rock, both with ~68 wt.% SiO_2, reveal weakening (Roig Silva *et al.*, 2004a, 2004b). It is also worth noting that many rocks in nature having low percentages of SiO_2 nevertheless are cut by faults that contain abundant deformed quartz, introduced by hydrothermal vein deposition. An example is the fault zone of the Dixie Valley Fault in Nevada (Power and Tullis, 1989). This fault zone has experienced large historical earthquakes, in particular, the M 7.1 Fairview Peak and M 6.8 earthquakes of 1954 (Caskey *et al.*, 1996) and the M 7.6 Pleasant Valley earthquake of 1915 (Doser, 1988). At the well-exposed 'mirror locality', north of where the 1954 earthquakes occurred, the fault shows abundant quartz mineralization, even though the country rock is a gabbro (Power and Tullis, 1989). Thus, dramatic weakening during earthquakes may still occur during faulting in low silica rocks due to the presence of mineralized quartz.

This weakening mechanism could be important for earthquakes. It is possible that at elevated temperature the transition to thixotropic weakening might be shifted to velocities above 1 m s^{-1}, since the kinetics of gel dehydration might be enhanced. On the other hand, in hydrothermal environments as may be common on faults, elevated temperature in the presence of pressurized H_2O might promote gel formation. These speculations can be investigated by experiment. Verifying the operation of this weakening mechanism on faults may be difficult, because

silica gel is even less likely to leave diagnostic traces than shear melting. The gel may lose its water and crystallize to chalcedony or other fine-grained quartz rocks. Very fine-grained deformed quartz layers in fault zones such as Dixie Valley (Power and Tullis, 1989) may plausibly have formed as a result of crystallization of silica gel, but proving this may not be possible.

4.05.3 Friction Resulting from High-Speed Weakening Mechanisms

For many of these weakening mechanisms we do not know to what value the shear resistance might be reduced during coseismic slip, whereas for others we have estimates of how low the shear stress might become. Recognizing that there is still uncertainty in the friction estimates and in whether these weakening mechanisms can operate during earthquakes, **Table 1** presents our current understanding of the effective friction coefficient possible from these mechanisms. Values are being presented to only one decimal point since our current knowledge does not warrant greater precision. These estimates are presented as effective coefficients of friction for reasons of convenience of comparison between mechanisms and for comparison with the commonly used friction coefficient for rock of ~0.6. A friction coefficient is not strictly an appropriate representation for some of the mechanisms. Examples include interfacial lubrication by frictional melt, for which the resistance is better understood as a viscous resistance, and any mechanism that involves dynamic reductions of the effective normal stress. In such cases, the effective friction coefficients should be considered to be the shear stress that resists dynamic sliding, divided by the effective normal stress that is expected to operate across a fault at a depth of about 10 km.

Table 1 Plausible values of shear resistance for proposed dynamic weakening mechanisms

Weakening mechanism	Effective coefficient of friction (τ/σ_n)
Dynamic normal stress reduction or loss of contact from normal interface vibrations	0.0–0.6
Dynamic normal stress reduction from elastic or permeability mismatch	?
Acoustic fluidization	?
Elastohydrodynamic lubrication	?
Thermal pressurization of pore fluid	0.0–0.6
'Flash' heating/melting at asperity contacts	~0.2
Interface lubrication by friction melt	0.0–0.6
Interface lubrication by thixotropic silica gel	<0.2

4.05.4 Implications of Low Dynamic Friction for Earthquake Stress Drops and for Orientations and Magnitudes of Tectonic Stress

A complete discussion of the implications of low dynamic friction during earthquakes due to the various weakening mechanisms presented in **Table 1** is beyond the scope of this chapter, but some discussion is warranted. Low dynamic friction can result in slip occurring as self-healing pulses (Heaton, 1990) rather than in a more conventional crack-like manner (Rosakis, 2002). Some particularly important implications exist for the stress drops of earthquakes and/or for the levels of tectonic stresses typical of the Earth's crust and their orientations.

If the differential stress (the greatest principal compressive stress magnitude minus the least principal compressive stress magnitude) in the crust typically is 'high', namely approaching the differential stress required for frictional slip on optimally oriented faults with a static coefficient of friction of 0.6, as *in situ* stress measurements suggest (Hickman, 1991), and if shear stresses on faults during coseismic slip are essentially zero as **Table 1** suggests, then dynamic stress drops would be in the range of about 50–100 MPa for strike-slip earthquakes (for example) at depths of 5–10 km, respectively. However, stress drops are in the 1–10 MPa range for large earthquakes (Kanamori and Allen, 1986; Kanamori and Anderson, 1975).

Consequently, it seems possible that one or more of the three observations referred to in the above paragraph, namely high *in situ* measured differential stress, low dynamic friction (shear stress divided by effective normal stress, namely τ/σ_n), and low stress drops, are in error. It is certainly possible that earthquake slip does not occur at shear stresses approaching zero and much more work on the weakening mechanisms that lead to this suggestion is unquestionably needed. On the other hand, given the number of potential weakening mechanisms that may reduce dynamic friction to nearly zero, it would

not be surprising if one or more of them commonly operated during earthquakes. If this were the case, how can the dilemma presented by the combination of high *in situ* differential stress measurements, low dynamic friction, and estimates of low earthquake stress drops be resolved?

One point to make is that high shear stresses (namely half of the differential stress) do not need to exist everywhere in order to overcome high static frictional strength as an earthquake rupture propagates. (Here we will refer to the static friction or static frictional strength as the value of the friction coefficient – namely the shear stress on the fault plane at which slip begins, divided by the effective normal stress on the fault plane – which must be overcome to initiate significant slip.) The value of this friction coefficient is typically in the range of 0.6–0.8. It is typically only somewhat higher (~0.1) than the steady-state friction coefficient for sliding at slip velocities less than about 1 mm s^{-1} (see this velocity range in **Figure 10** for representative values of friction coefficient as a function of velocity). (This static friction might be regarded as the unlabeled peak of the solid curve in figure 3 in Chapter 4.03.) Only in the nucleation area of the earthquake is it necessary that static friction be overcome quasi-statically, possibly by having locally high shear stresses, low static friction, or low effective normal stress. Once the rupture begins to propagate, static friction can be overcome by the dynamic stress concentration at the tip of the propagating rupture (Lapusta and Rice, 2004), allowing a crust with relatively low average stress values to experience earthquakes with low stress drops and high static friction (see discussion following equation (39) in Chapter 4.03).

The difficulty with this model would seem to be the relatively high measured values of *in situ* differential stress. However, a possible resolution of this dilemma would be for the orientations of the principal stresses to be such that they are always at high angles to faults on which the stress drops are low and on which the dynamic friction is low. In this case, although the magnitudes of the differential stresses can be high and consistent with *in situ* stress measurements, the shear stresses on the fault would be low, permitting both low stress drops and low dynamic friction. This relationship between fault and stress orientations is of course one reason we believe the San Andreas fault is weak (Mount and Suppe, 1987; Rice, 1992; Zoback *et al.*, 1987). This relationship might only apply to major mature faults and in fact there is a tendency for higher stress drops to occur on intraplate earthquakes and those with longer repeat times than on interplate earthquakes (Kanamori and Allen, 1986; Kanamori and Anderson, 1975). Thus, for some small to moderate earthquakes on immature faults, observed stress drops do obtain values in the 30–60 MPa range (Chung and Gao, 1997; Dewberry and Crosson, 1996; Fukuyama *et al.*, 1991; Huang and Yeh, 1999; Ide and Takeo, 1996; Kanamori *et al.*, 1992; Ma and Kanamori, 1994; Mori, 1996; Takeo *et al.*, 1993), namely values as large as we would expect for high *in situ* differential stresses and low dynamic friction. Although the nearly orthogonal orientation relationship between major faults and principal stresses required by this model might seem to represent an infrequent situation, this in fact may be what one would expect to develop if faults are weak when they slip (i.e., have low dynamic friction), because the faults relax the shear stresses parallel to them, but not shear stress on other planes (Tullis, 2007).

Thus, an interesting possibility is that faults with low stress drops (i.e., in the 1–3 MPa range) tend to be the well-developed long-lived ones that have been able to relax the shear stresses parallel to them, whereas those with high stress drops (i.e., in the 10–60 MPa range) are less mature and have not yet slipped enough over many earthquake cycles to relax the shear stress on them, and so tend to have the maximum and minimum principal stress orientations closer to 45° to them. This opens the possibility that for some earthquakes on appropriately oriented faults the stress drops and the resulting ground motions might be substantially higher than is typically the case.

4.05.5 Conclusions

This brief review of our current understanding of high-speed weakening mechanisms that may result in low resistance to slip during earthquakes shows that many such mechanisms potentially exist. It is of course possible that further theoretical and experimental study will discover new dynamic weakening mechanisms. It is also clear that much more remains to be understood about all of these weakening mechanisms. Although some of these mechanisms have been described in the literature for years, it is only recently that they have received renewed attention, so there is hope that it may not be too long before much more is known and this review will be out of date.

At present it appears that many of these mechanisms may result in quite low shear resistance during coseismic slip (**Table 1**). The resulting low dynamic stress can be compatible with low stress drops and high magnitudes of *in situ* stress if principal stresses are oriented at high angles to faults with low stress drops.

Acknowledgments

This work was supported by USGS grants 05-HQGR0087 and 06-HQGR0068 and by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. The SCEC contribution number for this paper is 1113. The author thanks David Goldsby for providing **Figures 4**, **5**, and **11**, Vikas Prakash for providing **Figures 6** and **7**, and Giulio Di Toro for providing the materials for **Figure 8**.

References

Adams GG (1995) Self-excited oscillations of two elastic half-spaces sliding with constant coefficient of friction. *Journal of Applied Mechanics* 62: 867–872.

Adams GG (1998) Steady sliding of two elastic half-spaces with friction reduction due to interface stick-slip. *Journal of Applied Mechanics* 65: 470–475.

Adams GG (2001) An intersonic slip pulse at a frictional interface between dissimilar solids. *Journal of Applied Mechanics* 68: 81–86.

Andrews DJ (1976) Rupture velocity of plane strain shear cracks. *Journal of Geophysical Research* 81: 5679–5687.

Andrews DJ (2002) A fault constitutive relation accounting for thermal pressurization of pore fluid. *Journal of Geophysical Research* 107: 12.

Andrews DJ and Ben-Zion Y (1997) Wrinkle-like slip pulse on a fault between different materials. *Journal of Geophysical Research* 102: 553–571.

Andrews DJ and Harris RA (2005) The wrinkle-like slip pulse is not important in earthquake dynamics. *Geophysical Research Letter* 32: L23303 (doi:10.1029/2005GL023996).

Andrews DJ and Harris RA (2006) Reply to comment by Y. Ben-Zion on 'The wrinkle-like slip pulse is not important in earthquake dynamics'. *Geophysical Research Letters* 33(6): L06311.

Anooshehpoor A and Brune JN (1994a). Dynamics and frictional heat generation in a foam rubber model of earthquake stick-slip. In: *USGS Red Book NEHRP Open-File Report 94-228, Proceedings of Workshop LXIII The Mechanical InVol.vement of Fluids in Faulting*. June 6–10, 1993, Fish Camp, CA, pp. 443–459.

Anooshehpoor A and Brune JN (1994b) Frictional heat generation and seismic radiation in a foam rubber model of earthquakes. *Pure and Applied Geophysics* 142: 735–747.

Anooshehpoor A and Brune JN (1999) Wrinkle-like Weertman pulse at the interface between two blocks of foam rubber with different velocities. *Geophysical Research Letters* 26(13): 2025–2028.

Archard JF (1958) The temperature of rubbing surfaces. *Wear* 2: 438–455.

Baker DR (1998) Granitic melt viscosity and dike formation. *Journal of Structural Geology* 20(9-10): 1395–1404.

Beeler NM, Tullis TE, and Goldsby DL (2007). Constitutive relationships and physical basis of fault strength due to flash-heating. *Journal of Geophysical Research* 112, (in press), paper 2007JB004988R.

Beeler NM and Tullis TE (1997) The roles of time and displacement in velocity-dependent volumetric strain of faults. *Journal of Geophysical Research* 102: 22595–22609.

Beeler N and Tullis T (2003) Constitutive relationships for fault strength due to flash-heating. In *2003 SCEC Annual Meeting Proceedings and Abstracts*, pp. 66, Southern California Earthquake Center, Palm Springs, CA.

Beeler NM, Tullis TE, Blanpied ML, and Weeks JD (1996) Frictional behavior of large displacement experimental faults. *Journal of Geophysical Research* 101: 8697–8715.

Ben-Zion Y (2001) Dynamic ruptures in recent models of earthquake faults. *Journal of the Mechanics and Physics of Solids* 49(9): 2209–2244.

Ben-Zion Y (2006) Comment on ''The wrinkle-like slip pulse is not important in earthquake dynamics'' by D. J. Andrews and R. A. Harris. *Geophysical Research Letter* 33: L06310 (doi:10.1029/2005GL025372).

Ben-Zion Y and Andrews DJ (1998) Properties and implication of dynamic rupture along a material interface. *Bulletin of the Seismological Society of America* 88: 1085–1094.

Ben-Zion Y and Huang Y (2002) Dynamic rupture on an interface between a compliant fault zone layer and a stiffer surrounding solid. *Journal of Geophysical Research* 107(B2): 2042 (doi:10.1029/2001JB000254).

Ben-Zion Y and Shi Z (2005) Dynamic rupture on a material interface with spontaneous generation of plastic strain in the bulk. *Earth and Planetary Science Letters* 236: 486–496 (doi:10.1016/j.epsl.2005.03.025).

Bjornerud A and Magloughlin JF (2004) Pressure-related feedback processes in the generation of pseudotachylytes. *Journal of Structural Geology* 26(12): 2317–2323.

Blok H (1937) Measurement of temperature flashes on gear teeth under extreme pressure conditions. In: *Proc. general discussion on lubrication and lubricants*, vol. 2, pp. 14–20. London: Institute of Mechanical Engineers.

Bowden FB and Persson PA (1960) Deformation heating and meltingof solids in high speed friction. *Proceedings of the Royal Society of London. Series A* 260: 433–458.

Bowden FP and Ridler KEW (1936) Physical properties of surfaces. Part III: The surface temperature of sliding metals; The temperature of lubricated surfaces. *Proceedings of the Royal Society of London. Series A* 154: 640–656.

Bowden FP, Stone MA, and Tudor GK (1947) Hot spots on rubbing surfaces and the detonation of explosives by friction. *Proceedings of the Royal Society of London. Series A* 188: 329–349.

Bowden FP and Thomas PH (1954) The surface temperature of sliding solids. *Proceedings of the Royal Society of London. Series A* 223: 29–40.

Brodsky EE and Kanamori H (2001) Elastohydrodynamic lubrication of faults. *Journal of Geophysical Research* 106: 16357–16374.

Brown SR, Olsson WA, and Brune JN (1991) Experimental observation of interface separation waves during stick-slip. *EOS Transactions* 72: 326.

Brune JN (1991) Seismic source dynamics, radiation and stress. *Reviews of Geophysics, Supplement* 688–699.

Brune JN (1996) Particle motions in a physical model of shallow angle thrust faulting. *The Proceedings of the Indiana Academy of Science* 105(2): L197–L206.

Brune J (2001) Fault-normal dynamic loading and unloading: An explanation for 'nongouge' rock powder and lack of fault-parallel shear bands along the San Andreas Fault. *EOS Transactions* 82(47).

Brune JN, Brown S, and Johnson PA (1993) Rupture mechanism and interface separation in foam rubber models of earthquakes: A possible solution to the heat flow paradox and the paradox of large overthrusts. *Tectonophysics* 218: 59–67.

Brune JN, Johnson PA, and Slater C (1990) Nucleation, predictability, and rupture mechanism in foam rubber models of earthquakes. *Journal of Himalayan Geology* 1: 155–166.

Brune JN and Thatcher W (2002) Strength and energetics of active fault zones. *International Handbook of Earthquake and Engineering Seismology* 81A: 569–588.

Burnham CW (1979) The importance of volatile constituents. In: Yoder HS, Jr. *The Evloution of Ignesus rocks: Fiftieth Anniversary Perspectives*, pp. 439–482. Princeton: Princeton University Press.

Byerlee JD (1978) Friction of rocks. *Pure and Applied Geophysics* 116: 615–626.

Caskey SJ, Wesnousky SG, Zhang P, and Slemmons DB (1996) Surface faulting of the 1954 Fairview Peak (MS 7.2) and Dixie Valley (MS 6.8) earthquakes, central Nevada. *Bulletin of the Seismological Society of America* 86(3): 761–787.

Castro RR, Anderson JG, and Brune JN (1991) Origin of high P/S spectral ratios from the Guerrero accelerograph array. *Bulletin of the Seismological Society of America* 81(6): 2268–2288.

Chester FM and Chester JS (1998) Ultracataclasite structure and friction processes of the Punchbowl fault, San Andreas system, California. *Tectonophysics* 295: 199–221.

Chester FM, Chester JS, Kirschner DL, Schulz SE, and Evans JP (2004) Structure of large-displacement strike-slip fault zones in the brittle continental crust. In: Karner GD, Taylor B, Driscoll NW, and Kohlstedt DL (eds.) *Rheology and Deformation in the Lithosphere at Continental Margins*, pp. 223–260. New York: Columbia University Press.

Chester FM, Evans JP, and Biegel RL (1993) Internal structure and weakening mechanisms of the San Andreas Fault. *Journal of Geophysical Research* 98(B1): 771–786.

Chung WY and Gao H (1997) The Greenland earthquake of 11 July 1987 and postglacial fault reactivation along a passive margin. *Bulletin of the Seismological Society of America* 87(4): 1058–1068.

Cochard A and Rice J (2000) Fault rupture between dissimilar materials: Ll-posedness, regularization, and slip-pulse response. *Journal of Geophysical Research* 105: 25891–25907.

Collins GS and Melosh HJ (2003) Acoustic fluidization and the extraordinary mobility of sturzstroms. *Journal of Geophysical Research-Solid Earth* 108(B10).

Cottrell AH (1965) *Dislocations and Plastic Flow in Crystals*, 223pp. London: Oxford University Press.

Dewberry SR and Crosson RS (1996) The M(D) 5.0 earthquake of 29 January 1995 in the Puget lowland of western Washington: An event on the Seattle fault? *Bulletin of the Seismological Society of America* 86(4): 1167–1172.

Di Toro G, Goldsby DL, and Tullis TE (2004) Friction falls toward zero in quartz rock as slip velocity approaches seismic rates. *Nature* 427: 436–439.

Di Toro G, Hirose T, Nielsen S, Pennacchioni G, and Shimamoto T (2006) Natural and experimental evidence of melt lubrication of faults during earthquakes. *Science* 311(5761): 647–649.

Di Toro G, Nielsen S, and Pennacchioni G (2005a) Earthquake rupture dynamics frozen in exhumed ancient faults. *Nature* 436(7053): 1009–1012.

Di Toro G and Pennacchioni G (2004) Superheated friction-induced melts in zoned pseudotachylytes within the Adamello tonalites (Italian Southern Alps). *Journal of Structural Geology* 26(10): 1783–1801.

Di Toro G and Pennacchioni G (2005) Fault plane processes and mesoscopic structure of a strong-type seismogenic fault in tonalites (Adamello batholith, Southern Alps). *Tectonophysics* 402(1–4): 55–80.

Di Toro G, Pennacchioni G, and Teza G (2005b) Can pseudotachylytes be used to infer earthquake source parameters? An example of limitations in the study of exhumed faults. *Tectonophysics* 402(1–4): 3–20.

Dieterich JH (1972) Time-dependent friction in rocks. *Journal of Geophysical Research* 77: 3690–3697.

Dieterich JH (1978) Time-dependent friction and the mechanics of stick slip. *Pure and Applied Geophysics* 116: 790–806.

Dieterich JH (1979) Modeling of rock friction, Part 1: Experimental results and constitutive equations. *Journal of Geophysical Research* 84: 2161–2168.

Dieterich JH and Kilgore BD (1994) Direct observation of frictional contacts: new insights for state-dependent properties. *Pure and Applied Geophysics* 143: 283–302.

Dieterich JH and Kilgore BD (1996) Imaging surface contacts: power law contact distributions and contact stresses in quartz, calcite, glass, and acrylic plastic. *Tectonophysics* 256: 219–239.

Dor O, Ben-Zion Y, Rockwell TK, and Brune JN (2006a) Pulverized rocks in the Mojave section of the San Andreas fault zone. *Earth and Planetary Science Letters* 245: 642–654 (doi:10.1016/j.epsl.2006.03.034).

Dor O, Rockwell TK, and Ben-Zion Y (2006b) Geologic observations of damage asymmetry in the structure of the San Jacinto, San Andreas and Punchbowl faults in southern California: A possible indicator for preferred rupture propagation direction. *Pure and Applied Geophysics* 163: (doi 10.1007/s00024-005-0023-9).

Doser DI (1988) Source parameters of earthquakes in the Nevada Seismic Zone, 1915–1943. *Journal of Geophysical Research* 93(B12): 15001–15015.

Ettles CMM (1986) The thermal control of friction at high sliding speeds. *Journal of Tribology-Transactions of the Asme* 108(1): 98–104.

Ferguson J and Kemblowski Z (1991) *Applied Fluid Rheology*, 323pp. New York: Elsevier.

Ferre EC, Allen JL, and Lin AM (2005) Pseudotachylytes and seismogenic friction: an introduction to current research. *Tectonophysics* 402(1–4): 1–2.

Fialko Y (2004) Temperature fields generated by the elastodynamic propagation of shear cracks in the Earth. *Journal of Geophysical Research-Solid Earth* 109(B1).

Fialko Y and Khazan Y (2005) Fusion by earthquake fault friction: Stick or slip? *Journal of Geophysical Research* 110(B12,407): doi:10.1029/2005JB003869.

Fukuyama E, Kinoshita S, and Yamamizu F (1991) Unusual High-Stress Drop Subevent During the M5.5 Earthquake, the Largest Event of the 1989 Ito-Oki Swarm Activity. *Geophysical Research Letters* 18(4): 641–644.

Goldsby DL and Tullis TE (1997) Shear heating induced pressurization of pore fluid as a dynamic fault weakening mechanism. *Eos Transactions* 78: 472.

Goldsby D and Tullis TE (1998) Experimental observations of frictional weakening during large and rapid slip. *Eos Transactions* 97: F610.

Goldsby D and Tullis TE (2002) Low frictional strength of quartz rocks at subseismic slip rates. *Geophysical Research Letters* 29(17): 1844 (doi:10.1029/2002GL015240).

Goldsby DL and Tullis T (2003) Flash heating/melting phenomena for crustal rocks at (nearly) seismic slip rates. In: *2003 SCEC Annual Meeting Proceedings and Abstracts*, pp. 98–90. Palm Springs, CA: Southern California Earthquake Center.

Goldsby DL and Tullis TE (2007a) Flash weakening in high speed rock friction experiments. *Journal of Geophysical Research* in preparation.

Goldsby DL and Tullis TE (2007b) The role of thixotropic silica gel as a lubricating layer during rapid fault slip. *Journal of Geophysical Research* in preparation.

Goldschmidt VM (1943) Om friksjonsglass (pseudo-tachylitt) i fjellkjeden. *GFF* 65: 83–84.

Graetsch H (1994) Structural characteristics of opaline and microcrystalline silica minerals. In: Heaney PJ, Prewitt CT, and Gibbs GV (eds.) *Silica, Physical Behavior, Geochemistry, and Materials Applications, Reviews in Mineralogy, 29*, pp. 209–232. Washington, DC: Mineralogical Society of America.

Hamrock BJ (1994) *Fundamentals of Fluid Film Lubrication*. New York: McGraw-Hill.

Harris RA and Day SM (1997) Effects of a low-velocity zone on a dynamic rupture. *Bulletin of the Seismological Society of America* 87(5): 1267–1280.

Harris RA and Day SM (2005) Material contrast does not predict earthquake rupture propagation direction. *Geophysical Research Letters* 32, L23301.

Heaton TH (1990) Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors* 64: 1–20.

Hickman SH (1991) Stress in the lithosphere and the strength of active faults, U.S. Nat. Rep. Int. Union of Geod. and Geophys. 1987–1990. *Reviews of Geophysics* 29: 759–775.

Hirose T (2002). *Experimental and Field Studies on the Frictional Melting along Faults and Implications to the Earthquake Generation Processes*. PhD thesis, Kyoto University. Kyoto: Japan.

Hirose T and Shimamoto T (2003) Fractal dimension of molten surfaces as a possible parameter to infer the slip-weakening distance of faults from natural pseudotachylites. *Journal of Structural Geology* 25: 1569–1574.

Hirose T and Shimamoto T (2005) Growth of a molten zone as a mechanism of slip weakening of simulated faults in gabbro during frictional melting. *Journal of Geophysical Research* 110: B05202 (doi:10.1029/2004JB003207).

Hirose T, Tsutsumi A, and Shimamoto T (2000) High-velocity slip weakening on gabbro during frictional melting. *EOS Transactions* 81(48): Abstract T12B-09.

Huang WG and Yeh YT (1999) Characteristics of seismic source spectra from the Chia-Yi and Tai-Nan area of Taiwan. *Terrestrial Atmospheric and Oceanic Sciences* 10(2): 415–446.

Hull D (1965) *Introduction to Dislocations*, 259pp. Oxford: Pergamon Press.

Ide S and Takeo M (1996) The dynamic rupture process of the 1993 Kushiro-oki earthquake. *Journal of Geophysical Research-Solid Earth* 101(B3): 5661–5675.

Iler RK (1979) *The Chemistry of Silica*, 720 pp. New York: Wiley.

Jaegar JC (1942) Moving sources of heat and temperature at a sliding contact. *Journal and Proceedings of the Royal Society of New South Wales* 76: 203–224.

Jeffreys H (1942) On the mechanics of faulting. *Geological Magazine* 79: 291–295.

Johnson KL (1985) *Contact Mechanics*, 452 pp. Cambridge, UK: Cambridge University Press.

Junger JA and Tullis TE (2003) Fault roughness and matedness suggest significant fault-interface dilatancy with slip. *EOS Transactions* 84(46): Fall Meeting Suppl., Abstract S51B-03.

Kalinin DV, Serdobintseva VV, and Kuznetsova YV (2002) Thixotropy in supramolecular crystallization and genesis of macrostructures of precious opal. *Geologiya I Geofizika* 43(11): 1002–1008.

Kanamori H and Allen CR (1986) Earthquake repeat times and average stress drop. In: Das S, Boatwright J, and Scholz CH (eds.) *Geophysical Monograph 37*, pp. 227–235. Washington, DC: American Geophysical Union.

Kanamori H and Anderson DL (1975) Theoretical basis of some emperical relations in seismology. *Bulletin of the Seismological Society of America* 65(5): 1073–1095.

Kanamori H and Heaton TH (2000) Microscopic and Macroscopic Physics of Earthquakes. In: Rundle J, Turcotte DL, and Klein W (eds.) *Geocomplexity and the Physics of Earthquakes*, pp. 147–163. Washington, DC: American Geophysical Union.

Kanamori H, Thio HK, Dreger D, Hauksson E, and Heaton T (1992) Initial Investigation of the Landers, California, Earthquake of 28 June 1992 Using Terrascope. *Geophysical Research Letters* 19(22): 2267–2270.

Khavryuchenko VD and Sheka EF (1994) Computational Modeling of Amorphous Silica .3. Modeling the Initial Structures – Silica-Gel. *Journal of Structural Chemistry* 35(3): 299–304.

Lachenbruch AH (1980) Frictional heating, fluid pressure, and the resistance to fault motion. *Journal of Geophysical Research* 85: 6249–6272.

Lapusta N and Rice JR (2004) Earthquake sequences on rate and state faults with strong dynamic weakening. *EOS Transactions* 85(47): T22A–05.

Lee TC and Delaney PT (1987) Frictional Heating and Pore Pressure Rise Due to a Fault Slip. *Geophysical Journal of the Royal Astronomical Society* 88(3): 569–591.

Lim SC and Ashby MF (1987) Wear mechanism maps. *Acta Metall* 35: 1–24.

Lim SC, Ashby MF, and Brunton JH (1989) The Effects of Sliding Conditions on the Dry Friction of Metals. *Acta Metallurgica* 37(3): 767–772.

Lin A and Shimamoto T (1998) Selective melting processes as inferred from experimentally generated pseudotachylytes. *Journal of Asian Earth Sciences* 16: 533–545.

Lockner D, Naka H, Tanaka H, Ito H, and Ikeda R (2000) Permeability and strength of core samples from the Nojima fault of the 1995 Kobe earthquake. In: Ito H, *et al.* (eds.) *Proceedings of the International Workshop on the Nojima Fault Core and Borehole Data Analysis*. Tsukuba, Japan, Nov 22–23, 1999, *U.S. Geol. Surv. Open File Rep.*, 00-129, pp. 147–152.

Ma KF and Kanamori H (1994) Broad-Band Wave-Form Observation of the 28 June 1991 Sierra-Madre Earthquake Sequence (M(L) = 5.8). *Bulletin of the Seismological Society of America* 84(6): 1725–1738.

Maddock R (1986) Frictional melting in landslide-generated frictionites (hyalomylonites) and fault-generated pseudotachylytes – discussion. *Tectonophysics* 128: 151–153.

Magloughlin JF (1989) The Nature and Significance of Pseudotachylite from the Nason Terrane, North Cascade Mountains, Washington. *Journal of Structural Geology* 11(7): 907–917.

Magloughlin JF (1992) Microstructural and chemical-changes associated with cataclasis and frictional melting at shallow crustal levels – the Cataclasite pseudotachylyte connection. *Tectonophysics* 204(3–4): 243–260.

Magloughlin JF and Spray JG (1992) Frictional melting processes and products in geological-materials – Introduction and discussion. *Tectonophysics* 204(3–4): 197–206.

Major JJ and Pierson TC (1992) Debris flow rheology – Experimental-analysis of fine-grained slurries. *Water Resources Research* 28(3): 841–857.

Marone CJ (1998) Laboratory-derived friction constitutive laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences* 26: 643–696.

Marone C, Raleigh CB, and Scholz CH (1990) Frictional behavior and constitutive modeling of simulated fault gouge. *Journal of Geophysical Research* 95: 7007–7025.

Mase CW and Smith L (1985) Pore-fluid pressures and frictional heating on a fault surface. *Pure and Applied Geophysics* 122: 583–607.

Mase CW and Smith L (1987) Effects of frictional heating on the thermal hydrologic and mechanical response of a fault. *Journal of Geophysical Research* 92: 6249–6272.

McKenzie DP and Brune JN (1972) Melting of fault planes during large earthquakes. *Geophysical Journal of the Royal Astronomical Society* 29: 65–78.

Melosh HJ (1979) Acoustic fluidization: A new geologic process? *Journal of Geophysical Research* 84: 7513–7520.

Melosh HJ (1996) Dynamic weakening of faults by acoustic fluidization. *Nature* 379: 601–606.

Melosh HJ and Girdner KK (1995) Rheology of vibrated granular materials: Application to long runout landslides. *EOS Transactions* 76(46): F270.

Melosh HJ and Ivanov BA (1999) Impact crater collapse. *Annual Review of Earth and Planetary Sciences* 27: 385–415.

Molinari A, Estrin Y, and Mercier S (1999) Dependence of the coefficient of friction on the sliding conditions in the high velocity range. *Journal of Tribology-Transactions of the Asme* 121(1): 35–41.

Mori J (1996) Rupture directivity and slip distribution of the M 4.3 foreshock to the 1992 Joshua Tree earthquake, Southern California. *Bulletin of the Seismological Society of America* 86(3): 805–810.

Morrow CA and Lockner DA (1997) Permeability and porosity of the Illinois UPH 3 drillhole granite and a comparison with other deep drillhole rocks. *Journal of Geophysical Research-Solid Earth* 102(B2): 3067–3075.

Morrow CA, Moore DE, and Lockner DA (2001) Permeability reduction in granite under hydrothermal conditions. *Journal of Geophysical Research-Solid Earth* 106(B12): 30551–30560.

Morrow CA, Shi LQ, and Byerlee JD (1984) Permeability of fault gouge under confining pressure and shear stress. *Journal of Geophysical Research* 99: 3193–3200.

Mount VS and Suppe J (1987) State of stress near the San Andreas fault: Implications for wrench tectonics. *Geology* 15: 1143–1146.

Noda H and Shimamoto T (2005) Thermal pressurization and slip-weakening distance of a fault: An example of the Hanaore fault, southwest Japan. *Bulletin of the Seismological Society of America* 95(4): 1224–1233.

Philpotts AR (1964) Origin of pseudotachylites. *American Journal of Science* 262: 1008–1035.

Power W and Tullis T (1992) The contact between opposing fault surfaces at Dixie Valley, Nevada, and implications for fault mechanics. *Journal of Geophysical Research* 97: 15425–15435.

Power WL and Tullis TE (1989) The relationship between slickenside surfaces in fine-grained quartz and the seismic cycle. *Journal of Structural Geology* 11: 879–893.

Prakash V and Clifton RJ (1993) Time resolved dynamic friction measurements in pressure-shear. In: Ramesh KT (ed.) *Experimental Techniques in the Dynamics of Deformable Bodies*, pp. 33–47. New York: ASME.

Prakash V and Yuan F (2004) Results of a pilot study to investigate the feasibility of using new experimental techniques to measure sliding resistance at seismic slip rates. *EOS Transactions* 85(47): Abstract T21D-02.

Rajagopalan S and Prakash V (1999) A modified Kolsky bar for investigating dynamic friction. *Experimental Mechanics* 39(4): 295–303.

Ranjith K and Rice J (2001) Slip dynamics at an interface between dissimilar materials. *Journal of the Mechanics and Physics of Solids* 49: 341–361.

Rempel AW and Rice JR (2006) Thermal pressurization and onset of melting in fault zones. *Journal of Geophysical Research-Solid Earth* 111(B9): B09314 (doi:10.1029/2006JB004314).

Rice JR (1992) Fault stress states, pore pressure distributions and the weakness of the San Andreas Fault. In: Evans B and Wong T (eds.) *Fault Mechanics and Transport Properties of Rocks*, pp. 475–503. San Diego, CA: Academic Press.

Rice JR (1999) Flash heating at asperity contacts and rate-dependent friction. *EOS Transactions* 80: F681.

Rice JR (2006) Heating and weakening of faults during earthquake slip. *Journal of Geophysical Research-Solid Earth* 111(B5): doi:10.1029/2005JB004006.

Rice JR and Cocco M (2007) Seismic fault rheology and earthquake dynamics. In: Handy MR (ed.) *The Dynamics of Fault Zones*, pp. 99–137. Cambridge, MA: MIT Press.

Roig Silva C, Goldsby DL, Di Toro G, and Tullis TE (2004a) The role of silica content in dynamic fault weakening due to gel lubrication. In: *2004 SCEC Annual Meeting Proceedings and Abstracts*, 150 pp. Palm Springs, CA: Southern California Earthquake Center.

Roig Silva C, Goldsby DL, Toro GD, and Tullis TE (2004b) The role of silica content in dynamic fault weakening due to gel lubrication. *EOS Transactions* 85(47): T21D–07.

Rosakis AJ (2002) Intersonic shear cracks and fault ruptures. *Advances in Physics* 51(4): 1189–1257.

Rudnicki JW and Rice JR (2006) Effective normal stress alteration due to pore pressure changes induced by dynamic slip propagation on a plane between dissimilar materials. *Journal of Geophysical Research-Solid Earth* 111(B10).

Ruina AL (1983) Slip instability and state variable friction laws. *Journal of Geophysical Research* 88: 10359–10370.

Scruggs VJ and Tullis TE (1998) Correlation between velocity dependence of friction and strain localization in large displacement experiments on feldspar, muscovite and biotite gouge. *Tectonophysics* 295: 15–40.

Segall P and Rice JR (1995) Dilatancy, compaction, and slip instability of a fluid infiltrated fault. *Journal of Geophysical Research* 100: 22155–22173.

Sengers JV and Watson JTR (1986) Improved International Formulations for the viscosity and thermal-conductivity of water substance. *Journal of Physical and Chemical Reference Data* 15(4): 1291–1314.

Serdobintseva VV and Kalinin DV (2001) Mechanism of reversible thixotropic transformation in concentrated suspensions of monodisprses spherical silica particles: analysis in terms of collodial crystal formation and growth. *Collodial Journal* 63(5): 626–629.

Serdobintseva VV, Kalinin DV, and Vosel SV (1998) The forms of colloid silica participating in the formation of noble opal and the mechanism of silicification of its gel crystals. *Geologiya I Geofizika* 39(8): 1116–1120.

Seront B, Wong TF, Caine JS, Forster CB, Bruhn RL, and Fredrich JT (1998) Laboratory characterization of hydromechanical properties of a seismogenic normal fault system. *Journal of Structural Geology* 20(7): 865–881.

Shaw DJ (1980) *Introduction to Colloid and Surface Chemistry*, 273 pp. London: Butterworths.

Shaw HR (1972) Viscosities of magmatic silicate liquids: An empirical method of prediction. *American Journal of Science* 272: 870–889.

Shi Z and Ben-Zion Y (2006) Dynamic rupture on a bimaterial interface governed by slip-weakening friction. *Geophysical Journal International* 165: doi:10.1111/j.1365-246X.2006.02853.x.

Shimamoto T and Lin A (1994) Is frictional melting equilibrium melting or non-equilibrium melting? (in Japanese with English abstract). *Structrual Geology* 39: 79–84.

Shimamoto T and Tsutsumi A (1994) A new rotary-shear high-velocity frictional testing machine: Its basic design and scope of research (in Japanese with English abstract). *Structural Geology* 39: 65–78.

Sibson RH (1973) Interactions between temperature and fluid pressuer during earthquake faulting – A mechanism for partial or total stress relief. *Nature* 243: 66–68.

Sibson RH (1975) Generation of pseudotachylyte by ancient seismic faulting. *Geophysical Journal of the Royal Astronomical Society* 43: 775–794.

Sibson RH (2003) Thickness of the seismic slip zone. *Bulletin of the Seismological Society of America* 93: 1169–1178.

Sibson RH and Toy V (2006) The habitat of fault-generated pseudotachylyte: Presence vs. absence of friction melt. In: Abercrombe R, McGarr A, Kanamori H, and DiToro G (eds.) *Geophysical Monograph Series: Radiated Energy and the Physics of Earthquake Faulting*, pp. 153–166. Washington, DC: AGU.

Spray JG (1987) Artificial generation of pseudotachylite using friction welding apparatus: simulation of melting on a fault plane. *Journal of Structural Geology* 9: 49–60.

Spray JG (1988) Generation and crystallization of an amphibolite shear melt – An investigation using radial friction welding apparatus. *Contributions to Mineralogy and Petrology* 99(4): 464–475.

Spray JG (1989a) Slickenside formation by surface melting during the mechanical excavation of rock. *Journal of Structural Geology* 11(7): 895–905.

Spray JG (1989b) Special Issue – Friction Phenomena in Rock – An Introduction. *Journal of Structural Geology* 11(7): 783–785.

Spray JG (1992) A physical basis for the frictional melting of some rock forming minerals. *Tectonophysics* 204: 205–221.

Spray JG (1993) Viscosity determinations of some frictionally generated silicate melts - implications for fault zone rheology at high-strain rates. *Journal of Geophysical Research* 98: 8053–8068.

Spray JG (1995) Pseudotachylyte controversy – Fact or friction. *Geology* 23(12): 1119–1122.

Spray JG (2005) Evidence for melt lubrication during large earthquakes. *Geophysical Research Letter* 32: doi:10.1029/2004GL022,293, Art. No. L07,301.

Spray JG and Thompson LM (1995) Friction melt distribution in a multiring impact basin. *Nature* 373(6510): 130–132.

Sulem J, Vardoulakis I, Ouffroukh H, Boulon M, and Hans J (2004) Experimental characterization of the thermo-poro-mechanical properties of the Aegion Fault gouge. *Comptes rendus Geoscience* 336(4–5): 455–466.

Swanson MT (1988) Pseudotachylyte-bearing strike-slip duplex structures in the Fort Foster brittle zone, S Maine. *Journal of Structural Geology* 10(8): 813–828.

Swanson MT (1990) Extensional duplexing in the York Cliffs strike-slip-fault system, Southern Coastal Maine. *Journal of Structural Geology* 12(4): 499–512.

Swanson MT (1992) Fault structure, wear mechanisms and rupture processes in pseudotachylyte generation. *Tectonophysics* 204: 223–242.

Takeo M, Ide S, and Yoshida Y (1993) The 1993 Kushiro-Oki, Japan, earthquake – a High stress-drop event in a subducting slab. *Geophysical Research Letters* 20(23): 2607–2610.

Titone B, Sayre K, Di Toro G, Goldsby DL, and Tullis TE (2001) The role of water in the extraordinary frictional weakening of quartz rocks during rapid sustained slip. *EOS Transactions* 82: T31B–0841.

Tsutsumi A and Shimamoto T (1994) An attempt to measure the temperature of frictional melts of rocks produced during rapid fault motion (in Japanese with English abstract). *Structural Geology* 39: 103–114.

Tsutsumi A and Shimamoto T (1996) Frictional properties of monzodioriteand gabbro during seismogenic fault motion. *Journal of the Geological Society of Japan* 102: 240–248.

Tsutsumi A and Shimamoto T (1997a) High-velocity frictional properties of gabbro. *Geophysical Research Letters* 24: 699–702.

Tsutsumi A and Shimamoto T (1997b) Temperature measurements along simulated faults during seismogenic fault motion. In: Hong Y (ed.) *Contemporary Lithospheric Motion Seismic Geology, Proceedings of the 30th International Geological Congress, vol. 5*, pp. 223–232. Leiden: VSP.

Tullis TE (1988) Rock friction constitutive behavior from laboratory experiments and its implications for an earthquake prediction field monitoring program. *Pure and Applied Geophysics* 126: 555–588.

Tullis TE (1996) Rock friction and its implications for earthquake prediction examined via models of Parkfield earthquakes. In: Knopoff L (ed.) *Earthquake Prediction: The Scientific Challenge, Proceedings of the National Academy of Sciences, United States of America*, pp. 3803–3810. Washington, DC: National Academy of Sciences.

Tullis TE (2007) Low strength of major faults may control the orientation of stresses. *Nature* in preparation.

Tullis TE, Burgmann R, Cocco M, *et al.* (2007) Group Report: Rheology of fault rocks and their surroundings. In: Handy MR, Hirth G, and Hovius N (eds.) *Tectonic Faults: Agents of Change on a Dynamic Earth*, (Dahlem Workshop Report 95, Berlin, January 2005, on The Dynamics of Fault Zones), pp. 183–204. Cambridge, MA, USA: The MIT Press.

Tullis TE and Goldsby DL (2003) Flash melting of crustal rocks at almost seismic slip rates. *EOS Transactions* 84(46): Abstract S51B-05.

Tullis TE and Weeks JD (1986) Constitutive behavior and stability of frictional sliding of granite. *Pure Applied Geophys* 124: 10–42.

Weertman J (1963) Dislocations moving uniformly on the interface between isotropic media of different elastic properties. *Journal of the Mechanics and Physics of Solids* 11: 197–204.

Weertman J (1980) Unstable slippage across a fault that separates elastic media of different elastic constants. *Journal of Geophysical Research* 85: 1455–1461.

Weertman J (2002) Subsonic type earthquake dislocation moving at approximately square root 2 $\times$ shear wave velocity on interface between half spaces of slightly different elastic constants. *Geophysical Research Letter* 29(10): doi:10.1029/2001GL013916.

Wenk HR (1978) Are pseudotachylites products of fracture or fusion? *Geology* 6(8): 507–511.

Wenk HR (1979) Are pseudotachylites products of fracture or fusion – Reply. *Geology* 7(4): 162–163.

Wenk HR, Johnson LR, and Ratschbacher L (2000) Pseudotachylites in the eastern peninsular ranges of California. *Tectonophysics* 321(2): 253–277.

Wibberley CAJ (2002) Hydraulic diffusivity of fault gouge zones and implications for thermal pressurization during seismic slip. *Earth Planets Space* 54: 1153–1171.

Wibberley CAJ and Shimamoto T (2003) Internal structure and permeability of major strike slip fault zones: The median tectonic line in mid prefecture, southwest Japan. *Journal of Structural Geology* 25: 59–78.

Wibberley CAJ and Shimamoto T (2005) Earthquake slip weakening and asperities explained by thermal pressurization. *Nature* 436: 689–692.

Xia K, Rosakis AJ, Kanamori H, and Rice JR (2005) Laboratory earthquakes along inhomogeneous faults: directionality and supershear. *Science* 308: 681–684.

Zhang S, Tullis T, and Scruggs V (1999) Permeability anisotropy and pressure dependency of permeability in experimentally sheared gouge materials. *Journal of Structural Geology* 21: 795–806.

Zhang S and Tullis TE (1998) The effect of fault slip on permeability and its anisotropy in an artificial quartz gouge. *Tectonophysics* 295: 41–52.

Zhang S, Tullis TE, and Scruggs VJ (2001) Implications of permeability and its anisotropy in a mica gouge for pore pressures in fault zones. *Tectonophysics* 335: 37–50.

Zoback MD, Zoback ML, Mount J, *et al.* (1987) New evidence on the state of stress of the San Andreas fault system. *Science* 238: 1105–1111.

4.06 Dynamic Shear Rupture in Frictional Interfaces: Speeds, Directionality, and Modes

A. J. Rosakis, California Institute of Technology, Pasadena, CA, USA

K. Xia, University of Toronto, Toronto, ON, Canada

G. Lykotrafitis, Massachusetts Institute of Technology, Cambridge, MA, USA

H. Kanamori, Seismological Laboratory, Caltech, Pasadena, CA, USA

4.06.1 Introduction

The dramatic progress of research on earthquake rupture in the last decades along with advances in the various branches of geophysics has decisively improved our understanding of seismic source events. A tremendous scientific activity including theoretical seismomechanics studies, field data processing, and laboratory investigations have answered many outstanding questions on earthquake dynamics.

Other problems yet remain to be solved whereas new questions have recently emerged. One of these questions is related to the ability of laboratory experiments to faithfully 'mimic' natural earthquake rupture events. In particular, this question is concerned with the 'scalability' of the laboratory results and observations. Can conclusions regarding the nature of dynamic rupture (e.g., modes, speeds, directionality etc.), obtained from such experiments, be accurately transferred from the laboratory scale (meters) to the natural earthquake rupture scale (hundred of kilometers)? Another question relates to similar issues in modeling. Do mechanics models and numerical simulations of rupture also suffer from scaling issues and do they need to first be validated by direct comparison to experiments before their results are extrapolated to the scale of rupture processes in the Earth's crust? These are some of the questions which, although only partially addressed here, have motivated much of the work reviewed in the present chapter.

In this chapter, we first review past work on experimental simulations of earthquake rupture in the laboratory. We describe the basic traditional experimental techniques used in the last few decades and we then focus on a new experimental design philosophy introduced by the authors. This new approach to laboratory earthquakes has resulted in an experimental configuration particularly suited to the study of fast rupture events.

Following the description of the design philosophy, the chapter also concentrates on experimental results which demonstrate the attainability of intersonic rupture speeds (i.e., speeds in the interval between the shear and the dilatational wave speeds) and elucidate the sub-Rayleigh to supershear transition phenomenon in homogeneous fault systems. This transition is discussed in the light of well-established theoretical and numerical models based on the principles of fracture mechanics. In addition, recent field evidence of intersonic rupture that has occured during natural earthquakes is presented and compared to the experimental observations.

The issue of the directionality of ruptures along faults separating weakly dissimilar materials is currently in the center of a very active debate between seismologists. In the present chapter, theoretical analyses together with field evidence and recent experimental and numerical results are utilized to provide a complete picture of the state of contemporary knowledge on this problem.

Finally, the various rupture modes (crack-like, pulse-like, wrinkle-like) which have been proposed on the basis of field evidence and theoretical analyses are critically discussed. Experimental results, which show that the activation of these modes in the laboratory scale is possible, are presented.

Here, we have made the conscious decision not to include an extensive general introduction to this article. Instead, we have chosen to strengthen the introductory comments of individual sections and to blend reference to previous and current work within each section.

4.06.2 Experimental Techniques for Creating Earthquakes in the Laboratory

4.06.2.1 Early Experimental Studies

The first efforts of simulating earthquakes in the laboratory have utilized the classic block-slider configuration. Based on valuable observations gathered through this configuration, Brace and Byerlee (1966) have proposed that stick-slip may be an important mechanism for shallow earthquakes occurring along pre-existing faults. The basic assumption of the block-slider model is that the sliding rock blocks are rigid so that edge effects, leading to an initially nonuniform stress distribution along the interface, can be ignored. Under the rigidity assumption, friction along the interface can be assumed uniform and as a result only the displacement history of the slider needs to be measured. This assumption is clearly a very strong one since the Earth's crust is far from being rigid during earthquake faulting. Although the block-slider model experiments do provide useful insight into some aspects of earthquake mechanics, such models are more appropriate for providing frictional constitutive information instead of modeling the detailed mechanics of earthquake rupture. Indeed, based on the pioneering experimental results of Dieterich (1979), the famous state- and rate-dependent friction law was formulated by Ruina (1983) and Rice (1983).

There are two configurations suitable for producing laboratory earthquakes, namely those of direct shear and of biaxial compression (Brune, 1973; Johnson and Scholz, 1976; Wu *et al.*, 1972). In order to extract meaningful rupture dynamics information from these configurations, the assumption of rigidity should be abandoned. Indeed, if strain gauges are attached along the fault in a two-dimensional (2-D) domain, the dynamic strain history resulting from the rupture can be recorded at various points. With

this modification, rupture velocities can also be estimated (Brune, 1973; Johnson and Scholz, 1976; Wu *et al.*, 1972) in addition to the dynamic friction-slip relation (Dieterich and Kilgore, 1994; Ohnaka and Shen, 1999; Okubo and Dieterich, 1984). However, strain gauges have approximate sizes of few millimeters, and this limits the spatial resolution of these discrete measurements. As a result, the data are pointwise, averaged, and are arguably difficult to interpret. For example, although supershear ruptures were suggested by this type of experiment (Johnson and Scholz, 1976; Okubo and Dieterich, 1984; Wu *et al.*, 1972), this evidence was considered far from being conclusive due to the 'averaged' quality of the data. Furthermore, since the diagnostics system (the Wheatstone bridge and oscilloscope) was triggered by one of the strain gauge signals and there was no information on the exact time and location of the nucleating point of earthquake (hypocenter), the sampling rate was necessarily set too low to capture the whole process. This has limited the time resolution of the measurements. Despite their shortcomings, these early studies have produced very valuable information concerning the nature and modes of frictional sliding. Indeed, by using this type of setup, the two important phenomena of frictional healing and velocity weakening were first observed by Scholz *et al.* (1972) and by Dieterich (1972).

Recently, two important modifications of the original block-slider model were introduced. The first one involved introducing a granular layer into the sliding interface to simulate the fault core (Gu and Wong, 1994). The second one involved using soft materials (e.g., foam rubber) and multipoint measurements performed by burying accelerometers inside meter-sized blocks (Brune, 1973; Brune *et al.*, 1993). The first type of test usually ignored the dynamic features of the earthquake, and is thus beyond the scope of this chapter. As to the second type of test, namely, the 'foam model', the usage of slow wave speed material has some distinct advantages over using rocks in that: (1) the whole process can be captured without the need of high sampling rates or high-speed photography and (2) 3-D effects can be more effectively addressed, since it is relatively easy to bury gauges inside the model. However, some concerns exist regarding the model material itself. These include its nonlinear and viscous nature and the special frictional properties of the foam rubber (the coefficient of friction can be larger than unity). Regarding their constitutive behavior, foam rubbers can sustain large deformations and are not well described by the theory of linear elasticity. The presence of material nonlinearities (both kinematic and constitutive) raises concerns regarding the scalability of the results from the laboratory scale to naturally occurring earthquake ruptures. This together with their known viscoelastic response makes foam models unlikely candidates to simulate earthquakes, which occur in much stiffer and more brittle materials (i.e., rocks) and also involve sliding processes extending over lengths that are four to five orders of magnitude longer than in the laboratory. Furthermore, like all other direct shear-type experiments, the 'foam model' has also suffered from unavoidable edge effects as first pointed out by Scholz *et al.* (1972). Despite the few shortcomings described above, it should be emphasized that the foam rubber model has produced very interesting observations whose scientific value has been to elucidate some of the basic physics governing the sliding of frictional interfaces (Anooshehpoor and Brune, 1994, 1999; Brune and Anooshehpoor, 1997; Brune *et al.*, 1993).

At least three important simultaneously acting prerequisites are needed in order for an earthquake to occur either in the laboratory or in the Earth's crust. These are: proper tectonic loading, a pre-existing fault, and the occurrence of a certain type of triggering mechanism. The first two prerequisites were very well simulated and controlled in the previously mentioned block slider and foam rubber models. However, the triggering of the rupture was mainly left uncontrolled and was dominated by the configuration edge effects discussed above. The triggering mechanism is important for both the earthquake process and its subsequent evolution (Lu *et al.*, 2005). In experiments, control of the triggering mechanism is crucial since it effects the synchronization of high-speed diagnostics and also serves as an important boundary/initial conditions that may affect the repeatability and the interpretation of the results. Except for one particular study specifically targeting the problem of the dynamic triggering of earthquakes by incoming Rayleigh waves (Uenishi *et al.*, 1999), this aspect has largely been ignored by the previous literature. In this important investigation, two photoelastic polymer plates of different sizes were held together at their convex sides to form a fault (line of frictional contact) with an extended free surface. A Rayleigh wave pulse was subsequently generated on the free surface by a point explosion and this wave propagated along the free surface toward the contact part (simulated fault). The incoming Rayleigh pulse was used to trigger dynamic slip along the fault in a controlled and repeatable fashion. In addition to

controlling rupture initiation, these experiments were also the first ones to feature the use of full-field optical diagnostics to study the dynamic rupture process in 'incoherent' (frictional) interfaces.

In addition to the experimental studies described above, there also exists a class of dynamic shear fracture experiments originating from the engineering community which are closely related to the phenomenon of geological fault rupture. As reviewed by Rosakis (2002), fracture mechanicians are recently paying more and more attention to the dynamic shear rupture propagating along material interfaces which are becoming common in modern engineering structures and composite materials. These interfaces usually feature finite strength and toughness and can resist both opening and shear loading. They are often referred to as 'coherent' interfaces. Since they are in general weaker than the constituent matrix material, these interfaces often trap dynamic ruptures, of both the opening and shear types and serve as sites of catastrophic failure. On a much larger length scale, crustal faults provide natural weak frictional interfaces where shear-dominated earthquake ruptures occur under the superimposed action of tectonic pressure. Experimental methods for dynamic fracture mechanics, such as photoelasticity and coherent gradient sensing (CGS) combined with high-speed photography, have been applied successfully in the experimental study of shear ruptures propagating dynamically along coherent interfaces in engineering composites of various types (Coker and Rosakis, 2001; Coker *et al.*, 2003; Lambros and Rosakis, 1995; Rosakis *et al.*, 1998, 1999; 2000). These two diagnostic techniques, which are 2-D in nature, are able to provide very valuable full-field and real-time stress information. Specifically, the photoelastic method, which measures maximum in-plane shear stress contours, is an attractive candidate for the study of shear-dominated processes such as earthquake ruptures.

Further discussion of experimental methodologies and results pertaining to the investigation of various modes of rupture can be found in Section 4.06.5.1. In the following section, we present the principles that have guided the design of simple experiments involving incoherent (frictional) interfaces whose purpose has been to simulate earthquake rupture in a highly controlled laboratory setting.

4.06.2.2 Experimental Design Philosophy

One goal in designing dynamic frictional experiments simulating earthquake rupture has been to create a testing environment or platform which could reproduce some of the basic physics governing the rupture dynamics of crustal earthquakes while preserving enough simplicity so that clear conclusions can be obtained by pure observation. Another goal has been to ensure that this platform would be versatile enough to accommodate increasing levels of modeling complexity. We have concentrated in creating experiments that are well instrumented and feature diagnostics of high temporal and spatial resolution. In order to keep the design as simple as possible, we have decided to first concentrate on experiments involving essentially 2-D processes which may be thought of as 'experimental' equivalents of the multiplicity of 2-D numerical models of rupture which exist in the open literature. Because of its predominant two dimensionality this configuration is perhaps better suited to mimic processes corresponding to larger earthquake events which have already saturated the seismogenic zone and as such, involve primarily 2-D, mode-II rupture growth. The design of the generic experimental configuration, the earthquake triggering mechanism, and the choice of model materials and diagnostics are guided by the following principles:

4.06.2.2.1 Simulating tectonic loading

As shown in **Figure 1**, the crust is simulated by a 'large' photoelastic plate with a thickness of 9.5 mm and 150 mm × 150 mm in 2-D plane dimensions. The plate is large enough so that during the time window of observation there are no reflected wave arrivals. The plate is cut into two identical quadrilaterals, which are then put together, and the frictional interface is used to simulate a fault. The angle of the fault line to the horizontal is denoted by α while the uniaxial pressure acting at the top and the bottom ends of the sample is denoted by P. Both P and α are system variables. Following these definitions, the resolved shear traction τ and the resolved normal traction σ along the fault can be expressed in terms of angle α and pressure P as

$$\tau = P \sin\alpha \cos\alpha \quad \text{and} \quad \sigma = P \cos^2\alpha \qquad [1]$$

To make the connection with commonly used geophysics terminology, a dimensionless factor s, which is often used by seismologists to describe the loading along faults with respect to the strength of the fault (Scholz, 2002), is introduced. By invoking the slip-weakening frictional model (Palmer and Rice, 1973) which is shown schematically in **Figure 1(c)**, and by denoting the maximum strength of the fault by τ^y and

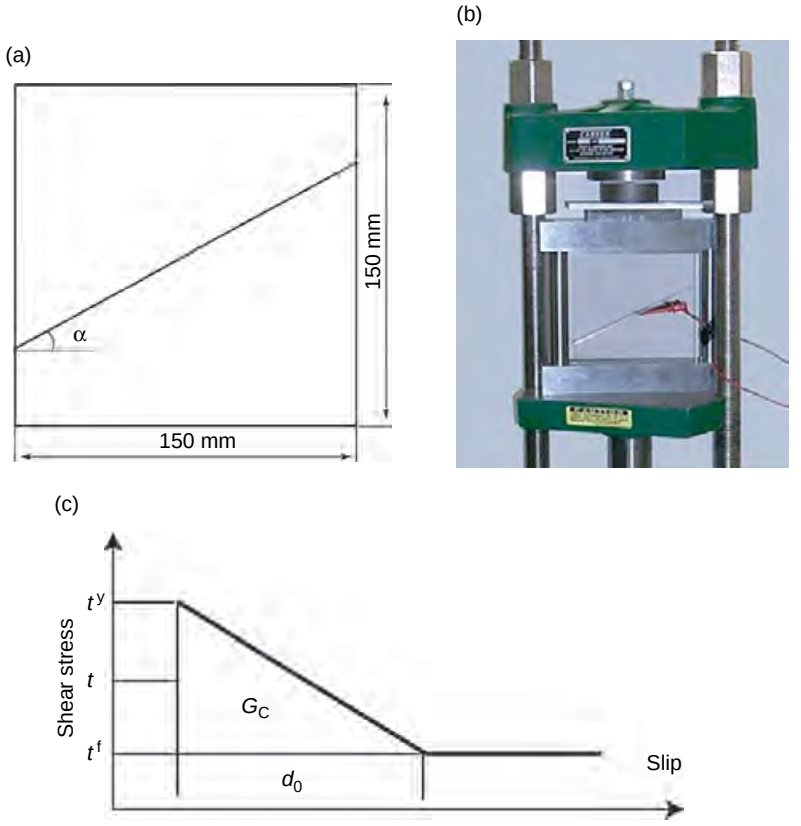


Figure 1 Laboratory fault model (a) and the loading fixture inside a hydraulic press (b). Slip-weakening frictional law (c).

the final strength of the fault by τ^f, this loading factor (or dimensionless driving force) s is defined by $s = (\tau^y - \tau)/(\tau - \tau^f)$. According to the linear slip-weakening fault strength model, τ^y is the static frictional strength while and τ^f is the dynamic frictional strength of the interface. By denoting the static and dynamic coefficients of frictions by μ^s and μ^d, respectively, and by using eqn [1], the loading factor s corresponding to the geometry of the experiment of **Figure 1** becomes

$$s = (\mu^s \cos\alpha - \sin\alpha)/(\sin\alpha - \mu^d \cos\alpha) \qquad [2]$$

As is obvious from eqn [2], s, and through it the earthquake rupturing process, can be controlled by varying the fault angle α for given frictional properties of the fault in the experiments. The magnitude of the uniaxial pressure P does not appear in s. However, it controls the total amount of deformation and the total slip.

4.06.2.2.2 Simulating the nucleation process

Researchers have tried to understand the nucleation process of earthquakes (triggering earthquake rupture) and indeed, they have shown that nucleation can be understood by invoking proper friction relations (Dieterich, 1992; Lapusta and Rice, 2003; Uenishi and Rice, 2003). Unfortunately, it is impractical to explore the detailed physics of the nucleation event experimentally and instead, we have decided to mimic the way that triggering of rupture is achieved in numerical simulations. The triggering of natural earthquakes can be achieved either by an increase of the shear loading or by a decrease of the fault strength at a specific location. In the first case triggering due to a loading increase can occur dynamically if the requisite jump of shear stress is provided by stress waves generated by a nearby earthquake source. It can also occur quasi-statically due to the presence of inhomogeneities or stress concentrations along a fault (e.g., kinks or forks) which, as slip accumulates, may result in a critical increment of the local resolved shear stress on the interface. In the second case, triggering occurs through a local decrease of fault strength which may result, among other things, from the flow of pore fluid into the fault interstice generating a local release of fault pressure. Both mechanisms (pre-stress

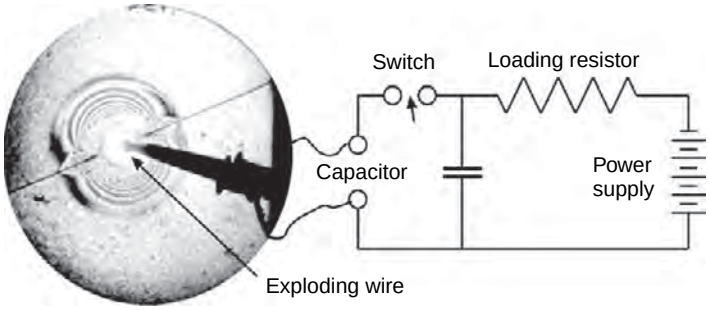


Figure 2 Schematic drawing of the exploding wire system coupled with a photoelastic fault model. Isochromatic fringes due to the explosion are visible.

and pressure release) have been applied in numerical simulations of earthquake rupture dynamics (Aagaard *et al.*, 2001; Andrews, 1976; Andrews and Ben-Zion, 1997; Cochard and Rice, 2000; Fukuyama and Madariaga, 1998).

In the laboratory earthquake models described in this chapter, the triggering mechanism is a local pressure release, which is achieved by the exploding wire technique shown in **Figure 2**. A capacitor is charged by a high-voltage power supply. The charging time is determined by the resistance of the charging resistor and the capacity of the capacitor. Upon closing the switch, the electric energy stored in the capacitor causes a high current in a thin metal wire (buried inside a hole of 0.1 mm in diameter in the interface) for a short duration. The high current turns the metal wire into high-pressure, high-temperature plasma in approximately 10 μs. The expansion of the high-temperature, high-pressure plasma causes a local pressure release. The adjustable power supply can provide electric potential in a wide range and different intensities and time durations of explosion can be obtained easily.

Before the explosion, the shear traction along the fault is less than the maximum static frictional strength. At the simulated hypocenter, after the explosion, the local normal traction on the fault is reduced and so is the static frictional strength. As a result, the applied shear traction, which is initially smaller than the static frictional strength and unaffected by the isotropic explosion, can be momentarily larger than the reduced frictional strength. The resulting net driving force, defined by the difference between the shear traction and the frictional resistance, drives the slip along the interface. Furthermore, the slip or the slip velocity also reduces the coefficient of friction as described by either a slip-weakening, or a velocity-weakening friction law. In other words, the frictional strength changes from static to dynamic. If the original shear traction is larger than the dynamic friction, the slip continues to propagate away from the explosion site (corresponding to the hypocenter of an earthquake) where normal traction reduction due to the explosion is not important any more. In this way a spontaneous rupture or a laboratory earthquake is triggered.

4.06.2.2.3 Choosing appropriate model materials and diagnostics

The diagnostic method of choice is dynamic photoelasticity. This technique is a classical method which measures the maximum shear stress in transparent, birefringent plates (Dally and Riley, 1991). As such the technique is very well suited for the study of shear-dominated processes. Two photoelastic solids, namely Homalite-100 and Polycarbonate, are used in all of the experiments described in the various sections of this chapter. The two materials have been chosen because of their enhanced birefringence relative to other transparent polymers such as polymethyl methacrylate (PMMA). Relevant properties of several photoelastic materials are listed in **Table 1**.

Table 1 Summary of optical and mechanical properties of Homalite-100 and Polycarbonate

Material property	Homalite 100	Polycarbonate
Young's Modulus E (MPa)	3860	2480
Poisson's Ratio ν	0.35	0.38
Stress fringe value f_σ (kN m^{-1})	23.6	7.0
Yielding Stress σ_Y (MPa)	48.3	34.5
P Wave Speed C_P (km s^{-1})	2.498	2.182
S Wave Speed C_S (km s^{-1})	1.200	0.960
Density ρ (kg m^{-3})	1230	1129

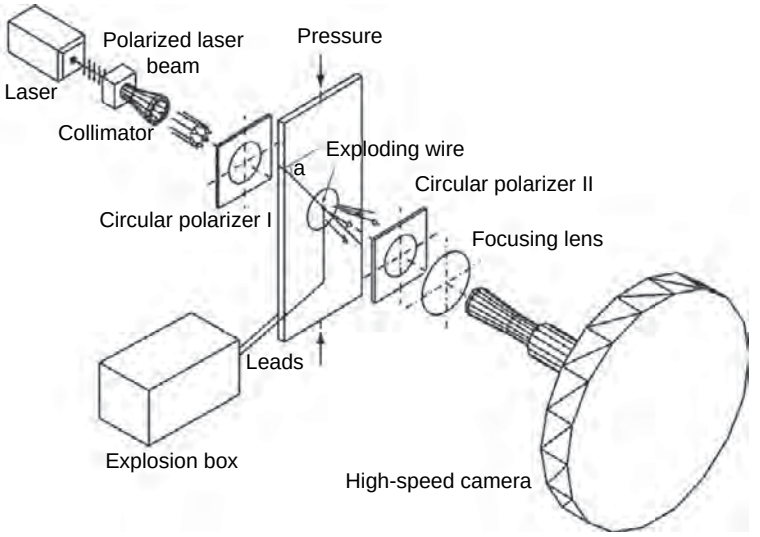


Figure 3 The setup of dynamic photoelasticity combined with high-speed photography for laboratory earthquake studies.

The stress fringe values quoted in the table are for green light at a wave length 525 nm. The static elastic properties listed in **Table 1** are from the literature (Dally and Riley, 1991), while the dynamic elastic properties (wave speeds) were measured using 5 MHz ultrasonic transducers.

A typical of dynamic photoelasticity setup is shown in **Figure 3**. A polarized laser provides a high-intensity beam continuously at a power level of a few watts. The beam is then expanded by a collimator to a size of 100 mm or 130 mm in diameter. The large beam is transmitted through a combination of circular polarizers and the transparent photoelastic specimen, and it is arranged so that an isochromatic fringe pattern is obtained and focused into the camera. The isochromatic fringe pattern obtained from the two-dimensional photoelastic model gives fringes along which the in-plane principal stress difference $\sigma_1 - \sigma_2$ is equal to a constant. When the fringe order N is known, the in-plane principal stress difference can be computed as follows:

$$\sigma_1 - \sigma_2 = Nf_\sigma/h \qquad [3]$$

where h is the plate thickness of the model material. The high-speed camera (Cordin 220) used can take pictures at speeds up to 10^8 frames/second. In most experiments an interframe time of 2–4 μs was used.

Another and perhaps the most important reason of using brittle polymers in our experiments is the small rupture nucleation length, L_C, resulting when such materials are frictionally held together by loads on the order of 10–20 MPa. As one can see by using eqn [1] in Section 4.06.3.3, these lengths are on the order of 5 mm and unlike rocks allow for relatively small laboratory specimen sizes. If rock samples were to be used instead of stiff polymers, the resulting rupture nucleation lengths would be approximately 30 times larger under similar far-field loading and surface finish. Finally, both Homalite and Polycarbonate are very brittle and are known to deform elastically at strain rates exceeding $10^2\ s^{-1}$.

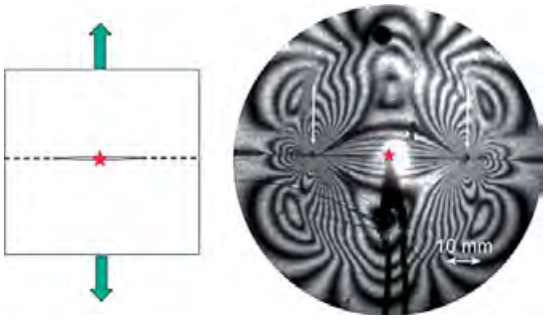


Figure 4 Specimen configuration and photoelastic fringe pattern for self-similar crack growth of a mode-I (opening) crack along a coherent interface. White arrows indicate crack tips.

An example of a dynamic photoelastic pattern corresponding to a 'coherent' (glued) and horizontal ($\alpha = 0$) interface subjected to far-field tension is shown in **Figure 4**. This configuration which is of relevance to dynamic self-similar fracture of the mode-I (opening) type has been discussed in the recent work by Xia *et al.* (2006).

4.06.3 Supershear and Sub-Rayleigh to Supershear Transition in Homogeneous Fault Systems

The question of whether natural earthquake ruptures can ever propagate at supershear speeds was until recently a subject of active debate within the seismological community. This is because of the often insufficient field data, as well as the limited resolution and nonuniqueness of the inversion process. However, a recent experimental study (Xia *et al.*, 2004) has almost settled this question by demonstrating that under controlled laboratory conditions supershear frictional rupture can occur under conditions similar to those present in some natural settings. A widespread view in seismology speaks of crustal earthquake ruptures mainly propagating at sub-Rayleigh speeds between 0.75 and 0.95 C_R (Kanamori, 1994). However, the multiplicity of independently collected evidence warrants further investigations of the mechanics of supershear rupture (velocity faster than the shear wave speed of the rock, C_S) propagation. Whether and how supershear rupture occurs during earthquakes has an important implication for seismic hazard because the rupture velocity has a profound influence on the character of near-field ground motions (Aagaard and Heaton, 2004).

In the 8 August 2003 issue of *Science*, Bouchon and Vallee (2003) took advantage of the unusual length of the rupture event and used both seismic waves and geologically observed total slip distribution to infer the rupture velocity history of the great Ms 8.1 (Mw 7.8) central Kunlunshan earthquake that occured in Tibet on 14 November 2001. This is an extraordinary event from the point of view of both earthquake dynamics and dynamic rupture mechanics. The rupture occurred over a very long, near-vertical, strike-slip fault segment of the active Kunlunshan fault and featured an exceptionally long (400 km) surface rupture zone and large surface slip (Lin *et al.*, 2002). Although it may not be unique, their modeling suggests speeds that are close to the Rayleigh wave speed of the crust, C_R, for the first 100 km of rupture growth, and transitioning to a supershear speed for the remaining 300 km of propagation. Their results were later corroborated by Das (private communication, 2006).

Recently, several other seismological reports also point to the possibility of supershear ruptures. These reports concentrate on events such as the 1979 Imperial Valley earthquake (Archuleta, 1984; Spudich and Cranswick, 1984), the 1992 Landers earthquake (Olsen *et al.*, 1997), and most recently the 2002 Denali earthquake in Alaska (Ellsworth *et al.*, 2004). The 1999 Izmit earthquake in Turkey (Bouchon *et al.*, 2001) is yet another event featuring a very long segment of supershear rupture growth. It should be noted here that for most of those examples mentioned above (Izmit and Denali excluded), the supershear ruptures occurred only on short patches along the whole rupture length and the results are not conclusive. Bouchon and Vallee's (2003) paper together with Das's work on Kunlunshan, as well as the work of Ellsworth *et al.* (2004) on Denali are the most recent of a series of investigations reporting supershear rupture growth occurring during 'large' earthquake events. Moreover, they present the first seismological evidence for transition from sub-Rayleigh to supershear.

Classical dynamic fracture theories of growing shear cracks have many similarities to the earthquake rupture processes (Broberg, 1999; Freund, 1990; Rosakis, 2002). Such theories treat the rupture front as a distinct point (sharp tip crack) of stress singularity. These conditions are close to reality in cases that feature strong 'coherent' interfaces of finite intrinsic strength and toughness. The singular approach ultimately predicts that dynamic shear fracture is allowed to propagate either at a sub-Rayleigh wave speed or at only one supershear speed, which is $\sqrt{2}$ times the shear wave speed. As a result, it excludes the possibility of a smooth transition of a steady-state rupture from sub-Rayleigh to supershear speed for a steady-state rupture. The introduction of a distributed rupture process zone has allowed fracture mechanics to better approximate the conditions that exist during real earthquake events (Ida, 1972; Palmer and Rice, 1973). Based on this so-called cohesive zone fracture mode, there is a forbidden speed range between C_R and C_S (Burridge *et al.*, 1979; Samudrala *et al.*, 2002a, 2002b). In the sub-Rayleigh speed range all speeds are admissible, but only the Rayleigh wave speed is a stable speed; in the supershear speed range all speeds are admissible, but only speeds larger than $\sqrt{2}C_S$ are stable (Samudrala *et al.*, 2002b). Ruptures with unstable speeds will accelerate to a stable speed as determined by the fault strength and the loading conditions. The theoretical results of the cohesive zone rupture model ultimately predict that earthquake ruptures can propagate either at Rayleigh wave speed or at supershear speeds lying within the interval between $\sqrt{2}C_S$ and C_P.

Early theoretical results by Burridge and coworkers (Burridge, 1973; Burridge *et al.*, 1979), along with numerical results by Andrews (1976) and Das and Aki (1977) have predicted the possibility of supershear rupture and have alluded to a mechanism (Rosakis, 2002) for transition from the sub-Rayleigh to the supershear rupture velocity regime. This is often referred to as the 'Burridge–Andrews' mechanism. According to the 2-D Burridge–Andrews mechanism, a shear rupture accelerates to a speed very close to C_R soon after its initiation. A peak in shear stress is found sitting at the shear wave front and is observed to increase its magnitude as the main rupture velocity approaches C_R. At that point, the shear stress peak may become strong enough to promote the nucleation of a secondary microrupture whose leading edge propagates at a supershear speed. Shortly thereafter, the two ruptures join up and the combination propagates at a speed close to C_P, the longitudinal wave speed of the solids. It is interesting that this transition was also clearly visualized by recent 2-D, atomistic calculations (Abraham and Gao, 2000; Gao *et al.*, 2001) of shear rupture in the microscale, which provided an impressive demonstration of the length scale persistence (14 orders of magnitude difference between the nanoscale and the scale of natural earthquake rupture) of this sub-Rayleigh to supershear rupture transition mechanism. The Burridge–Andrews mechanism is also known as the mother–daughter mechanism in mechanics literature.

For mixed-mode (tensile and shear) ruptures, a different transition model has also been suggested (Geubelle and Kubair, 2001; Kubair *et al.*, 2002, 2003). Based on numerical simulation, they suggest that a mix-mode rupture can speed up and cross the forbidden speed range between C_R and C_S continuously. Finally, recent numerical investigations of frictional rupture have identified alternate, asperity based, mechanisms that provide a 3-D rationalization of such a transition (Day, 1982; Dunham *et al.*, 2003; Madariaga and Olsen, 2000). In this case, 3-D effects play an important role in the transition. The rupture front focusing effect provides extra driving force to speed up the spontaneous rupture in the supershear range.

The experimental confirmation of the possibility of supershear (intersonic) fracture followed many years after the first theoretical predictions. Indeed, the long series of experiments summarized in a recent review by Rosakis (2002) demonstrate that intersonic crack growth in constitutively homogenous systems featuring coherent interfaces (interfaces with inherent strength) is possible and may also occur in various combinations of bimaterial systems. However, in all of the various cases discussed by Rosakis (2002), the cracks were nucleated directly into the intersonic regime and there was no observation of a transition from sub-Rayleigh to supershear speeds. This was due to the nature of the impact induced stress wave loading which was imparted in the absence of pre-existing static loading and the nature of the relatively strong coherence of the interface (provided by adhesive). The major differences between the conditions during earthquake rupture and those early fracture experiments have left questions regarding the plausibility of spontaneously generated intersonic rupture in frictionally held, incoherent interfaces unanswered. The recent work by Xia *et al.* (2004) on frictionally held interfaces has addressed this issue in detail. In the following sections, we will summarize their most important findings and we will attempt to relate these findings to field measurements related to recent natural earthquake events.

4.06.3.1 Purely Sub-Rayleigh and Purely Supershear Earthquake Ruptures

In this section, experimental parameter ranges leading to either exclusively sub-Rayleigh or exclusively supershear ruptures will be discussed. The configuration of interest is the one described in Section 4.06.2.2. The physics governing possible transitions between these two speed regimes within a single rupture event will be examined later. **Figure 5** shows two different experiments featuring sub-Rayleigh speed ruptures. In all the photographs, we can clearly see the circular shear wave front emitted from the simulated hypocenter. Rupture tips, characterized by the stress concentration (fringe concentrations in the photographs), are identified just behind the shear wave front. The P-wave is long gone out of the field of view (100 mm). The ruptures as shown in the photograph feature right lateral slip while the rupture tips are equidistant from the hypocenter (symmetric bilateral slip). From the rupture length history, it is possible to estimate the rupture velocity. In these two cases, the rupture velocities are very close to the Rayleigh wave speed of the material.

All experiments featuring lower inclination angles α and lower magnitude of uniaxial compression pressure P involve exactly the same features and ruptures speeds. This observation is consistent with many observations of earthquake rupture velocities in the

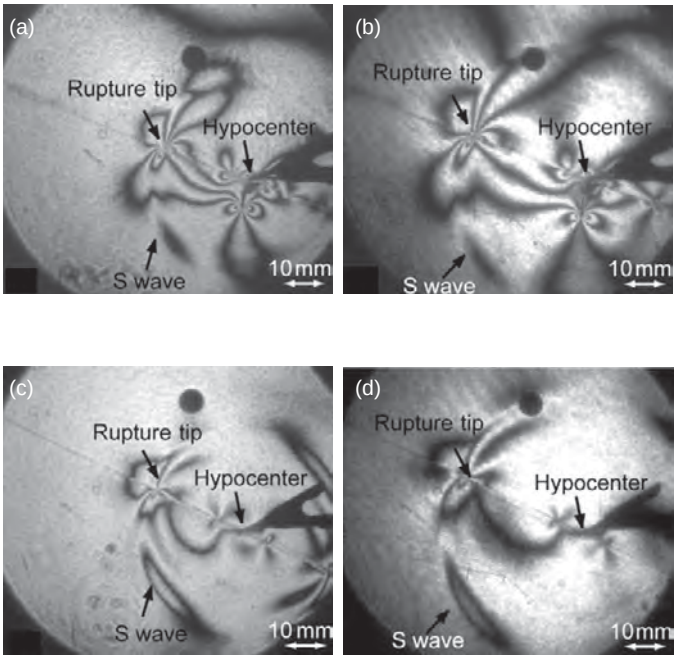


Figure 5 Experimental results of purely sub-Rayleigh cases. (a) and (b) are from one experiment with a pressure of $P = 13$ MPa and angle $\alpha = 20^\circ$ at time instants of 28 and 38 μs, respectively. (c) and (d) are from an experiment with a pressure of $P = 7$ MPa and angle $\alpha = 25^\circ$ at the time instants of 28 and 38 μs, respectively. For (a) and (b), we can also identify two mode-I cracked in the lower half of the sample caused by the explosion itself. We expect that the effect of these cracks is localized.

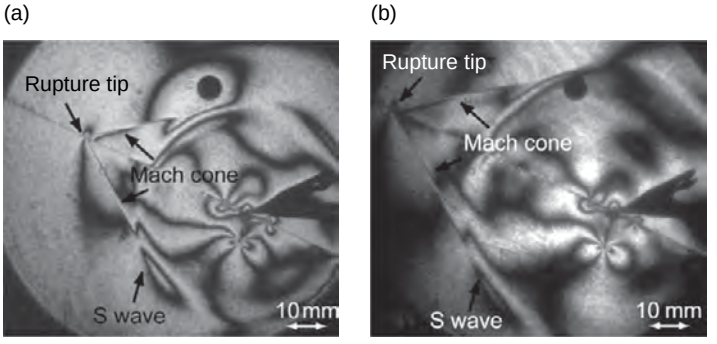


Figure 6 Experimental results for the purely supershear case. (a) and (b) both correspond to an experiment with the pressure of $P = 15$ MPa and angle $\alpha = 25^\circ$ at the two time instants of 28 and 38 μs, respectively.

field. As we have discussed earlier, in the sub-Rayleigh speed range the only stable rupture velocity is the Rayleigh wave speed according to various cohesive zone models of rupture. The experiments of Xia *et al.* (2004) described in this section confirm this observation.

In **Figure 6**, the inclination angle was kept at 25° while the pressure was increased to 15 MPa. For comparison purposes, the same time instants (28 and 38 μs after nucleation) are displayed.

In this case, the circular traces of the shear wave are also visible and are at the same corresponding locations as in **Figure 6**. However, in front of this circle supershear disturbances (propagating to the left, marked in the photograph as the rupture tip and featuring a clearly visible Mach cone) are shown. The formation of the Mach cone is due to the fact that the rupture front (tip) is propagating faster than the shear wave speed of the material. For this case, the sequence of images, other than

those at 28 and 38 μs, have a very similar form and reveal a disturbance that was nucleated as supershear. The speed history $v(t)$ is determined independently by either differentiating the rupture length history record or measuring the inclination angle, δ, of the shear shocks with respect to the fault plane, and using the relation $v = C_S/\sin\delta$. The speed was found to be almost constant and very close to the plane stress P-wave speed C_P of the material. These experiments provide the first evidence of supershear growth of a spontaneous shear rupture propagating along a frictional interface. The supershear rupture initiated right after the triggering of the earthquake rupture. This is determined from the fact that the Mach cones are nearly tangential to the shear wave front which was itself created at the time of rupture nucleation.

In previous experiments involving strong, coherent interfaces and impact induced stress wave loading, stable rupture velocities slightly exceeding $\sqrt{2}C_S$ were observed (Rosakis *et al.*, 1999). This apparent discrepancy can be explained by referring to the rupture velocity dependence on the available energy per unit shear rupture advance within the supershear regime (Samudrala *et al.*, 2002b). This dependence was obtained by means of an analytical model of a dynamic steady state, shear rupture featuring a shear cohesive zone of the velocity weakening type. The initial, static, strength of the interface was represented by the parameter τ_0. For a given normalized far-field loading level, σ_{12}^D/τ_0, this energy, shown here in normalized form (**Figure 7**) as a function of rupture speed, attains a maximum value at speeds closer to $\sqrt{2}C_S$ for interfaces that are 'strong'. This is consistent with the purely singular prediction which corresponds to idealized interfaces of infinite strength. To this idealized case, ($\tau_0 \to \infty$), the only value of rupture velocity corresponding to finite G is exactly $\sqrt{2}C_S$. For weaker interfaces (or higher driving stresses), this maximum moves toward C_P. In the situation described in this section, the interface is 'weak' and the driving force (resolved shear minus dynamic frictional force) is relatively large and constant. Hence, a rupture velocity close to C_P is expected.

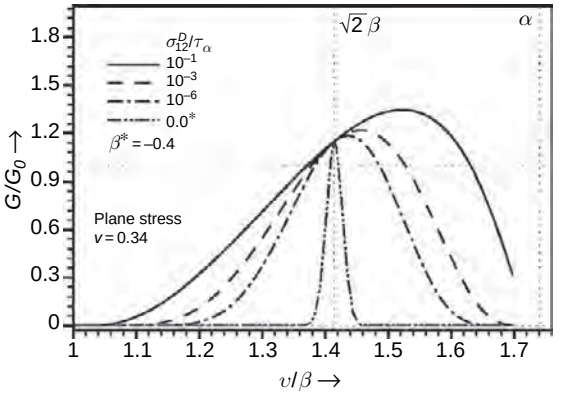


Figure 7 Normalized energy release rate as a function of intersonic rupture speed. The figure illustrates that the energy flux available to the tip has a maximum at or above $v = \sqrt{2}C_S$.

4.06.3.2 The Experimental Visualization of the Sub-Rayleigh to Supershear Earthquake Rupture Transition

As discussed earlier, investigating how supershear ruptures are nucleated experimentally is of great interest. For the experimental cases described in the above section, the supershear rupture event was nucleated immediately after triggering of the rupture event. Since the rupture velocity is controlled by both the inclination angle α and the magnitude of uniaxial compression P, it is possible to vary both of them carefully to suppress or perhaps delay the appearance of supershear rupture. Specifically, the inclination angle α was fixed at 25° while P was decreased in order to induce and capture the nucleation process of a supershear rupture.

Figures 8(a)–(c) corresponds to a case with an intermediate far-field pressure compared to the ones displayed in **Figures 5(c)** and **5(d)** and **Figure 6**. Here, the angle is kept the same (25°) and the pressure is decreased to 9 MPa in an attempt to visualize a transition within our field of view (100 mm). Three different time instances of the same rupture event are displayed. In **Figure 8(a)**, the circular traces of both P- and S-waves are visible, followed by a rupture propagating at C_R. In **Figure 8(b)**, a small secondary rupture appears in front of the main rupture and propagates slightly ahead of the S-wave front. In **Figure 8(c)**, the two ruptures coalesce and the leading edge of the resulting rupture grows at a speed of 1970 m s^{-1} which is very close to C_P. A well-defined 'shear' Mach cone is visible in **Figure 8(c)**. Unlike **Figure 6**, the Mach lines are not tangent to the shear wave circle since the supershear rupture was not generated at initiation. **Figure 8(d)** displays the length versus time of the two ruptures, in which the length is directly read from the figure. The figure also compared the slopes to the characteristic wave speeds of the material before and after their coalescence. A magnified view

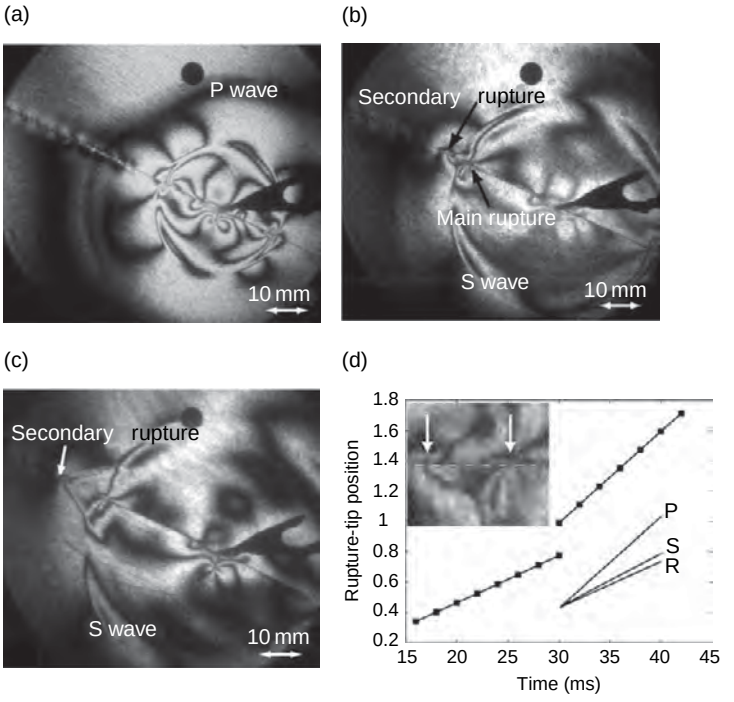


Figure 8 Visualization of the sub-Rayleigh to supershear rupture transition.

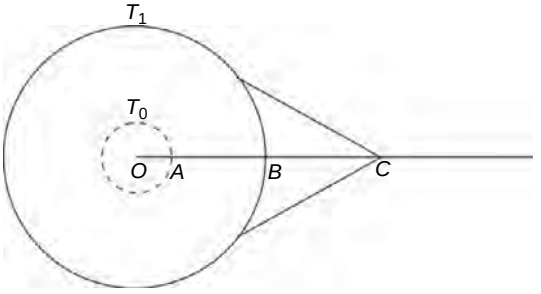


Figure 9 Method of transition length L estimation.

of the secondary rupture as it nucleates in front of the main rupture is shown in the inset. Both ruptures are indicated by arrows. The transition length L for this case is approximately 20 mm, as is visually obvious by **Figure 8(b)**.

In cases where the transition length cannot be measured directly if it is too small, an indirect method is developed. As shown in **Figure 9**, at time T_0, the supershear rupture is nucleated at the intersection of the shear wave front and the fault line (point A). If the transition length $L = \overline{OA}$ is very small, spatial resolution may not be good enough to measure it accurately. Alternatively, a photograph may not have been taken at that instant. Assuming that at all times before transition the rupture speed is very close to C_S and that the shear wave position (A) and supershear rupture-tip position (B) at a later time instance T_1 can be measured, the transition length can be inferred by pure geometry. To do so we observe that relations $\overline{OA} = C_S T_0$, $\overline{OB} = C_S T_1$, and $\overline{AC} = V(T_1 - T_0)$ hold provided that the supershear rupture tip also grows at a constant speed $V > C_S$ (as confirmed by **Figure 8**).

A simple manipulation gives $\overline{AC}/\overline{OB} = (1 - T_0/T_1)V/C_S$, which leads to $T_0 = [1 - (C_S/V)(\overline{AC}/\overline{OB})]T_1$. Multiplication of both sides of the last relation by C_S gives $L = \overline{OA} = C_S T_0 = [\overline{OB} - \overline{OC} \times C_S/V]$. This relation allows for the estimation of the supershear rupture transition length, L, even if no photographic records have been obtained at or before the time instance of the transition. Several results listed in this chapter were either obtained or verified by using this method.

4.06.3.3 A Theoretical Model for the Sub-Rayleigh to Supershear Transition

The above physical picture is comparable with the Burridge–Andrews mechanisms already described in the introduction of this chapter. Andrews (1976,

1985) quantified this transition in a parameter space spanned by a normalized supershear transition length L/L_C and the nondimensional driving stress parameter s [$s=(\tau^y-\tau)/(\tau-\tau^f)$]. The parameters τ, τ^y, and τ^f are the resolved shear stress on the fault, the static frictional strength, and the dynamic strength of the fault, respectively; they describe the linear slip-weakening frictional law (Ida, 1972; Palmer and Rice, 1973) used in Andrews' computations.

The Andrews result can be symbolically written as $L=L_C f(s)$. The function $f(s)$ has been given numerically by Andrews as an increasing function of s, and can be well approximated by the equation $f(s)=9.8(1.77-s)^{-3}$. The normalizing length L_C is the critical length for unstable rupture nucleation and is proportional to the rigidity G and to d_0, which is defined as the critical or breakdown slip of the slip-weakening model. L_C can then be expressed as

$$L_C=\frac{1+v}{\pi}\frac{(\tau^y-\tau^f)}{(\tau-\tau^f)^2}Gd_0 \qquad [4]$$

By adopting the above relation to the geometry of the experimental configuration and by using eqn [1], the transition length L is found to be inversely proportional to the applied uniaxial pressure P and to be governed by the following general functional form:

$$L=f(s)\frac{1+v}{\pi}\frac{(\mu^s-\mu^d)}{(\tan\alpha-\mu^d)^2}\frac{Gd_0}{P} \qquad [5]$$

Figure 10 displays the dependence of the transition length L on pressure from a set of experiments corresponding to the same inclination angle of 25° ($s=0.5$) and identical surface finish (roughness is about 17 μm). The static frictional coefficient was measured to be $\mu^s=0.6$ using the traditional inclined plane method. In this method, one block was placed on top of an inclined plane and the inclination angle until the block slides. By this way the critical inclination angle Ψ_C was measured. The static coefficient of friction was then determined from the relation $\mu^s=\tan^{-1}(\Psi_C)$. To estimate the average, steady state, dynamic frictional coefficient, α was gradually increased from 10° to a critical angle α_c at which slip was initiated under the action of far-field loads and the dynamic triggering. It was assumed that the shear traction was approximately equal to the dynamic frictional strength at this critical angle α_c. This angle was found to be between 10° and 15°, from which the average coefficient of dynamic friction was estimated to be $\mu^d=0.2$. This low value of dynamic friction coefficient is very consistent with recent experimental measurements conducted by Tullis and his research group at slip velocities approaching seismic slip rates (Di Toro *et al.*, 2004; Tullis and Goldsby, 2003). Using $\mu^s=0.6$ and $\mu^d=0.2$, the experiments can be compared to the predictions of eqn [5] of Andrews' theory as shown in **Figure 10**. Although the theory qualitatively captures the decreasing trends of the experiments, the data exhibit a dependence on pressure that is visibly stronger than P^{-1}.

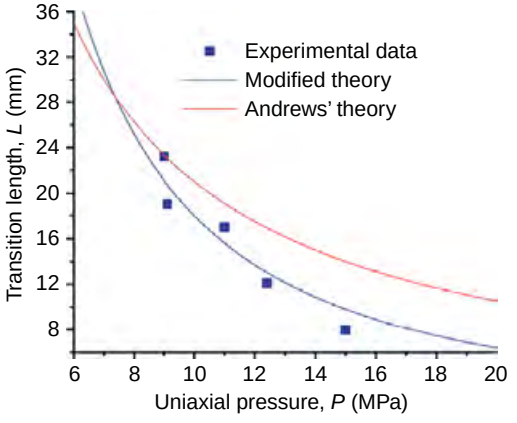


Figure 10 Transition length as a function of the far-field load.

A natural way to modify Andrews' results and to introduce some scaling into the pressure dependence of L is to consider the effect of pressure on the critical breakdown slip d_0. As pointed out by Ohnaka, on the basis of frictional experiments on rocks, and as also supported by dimensional analysis, there exists a linear relation between a characteristic surface length (half-distance between contacting asperities, denoted as D in this case) and the critical slip distance d_0 as (Ohnaka, 2003)

$$d_0=c[(\tau^y-\tau^f)/\tau^f]^M D \qquad [6]$$

where c and M are experimentally determined constants. In addition, D depends on the normal stress, σ, applied on the fault (which is $\sigma=P\cos^2\alpha$ in this case). As shown in **Figure 11**, a classical plastic contact model may be used to establish this dependence. In this model the average radius of n contacting asperities is a_0 (assumed as a constant in this model). As the pressure over a macroscopic contact area A ($=n\pi D^2$) is increased, the number of contacts,

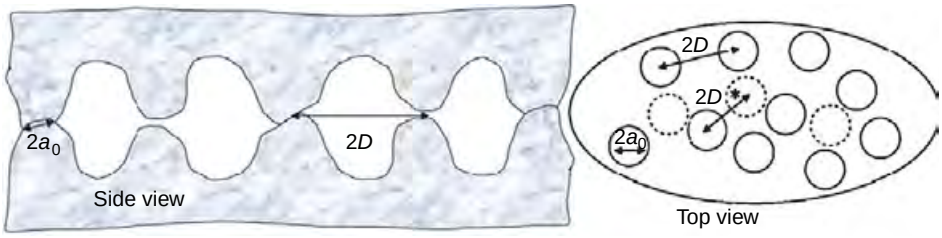


Figure 11 Schematic drawing of the microcontact based frictional model. The top figure is the side view of the contact and the bottom figure is the top view. As the normal force increases, the number of contacts, n, increase.

n, as well as the real contact area $A_r\left(=n\pi a_0^2\right)$, also increases.

By defining the hardness H as the ratio of the total normal force N to the real contact area A_r (Bowden and Tabor, 1986), N can be expressed as

$$N = HA_r = Hn\pi a_0^2 = \sigma A = AP\cos^2\alpha \qquad [7]$$

Substitution of A and A_r in terms of D and a_0 gives $D = \sqrt{H}a_0\cos\alpha P^{-1/2}$. Using the linear relation between D and d_0, the breakdown slip is further found to depend on the pressure as $d_0 \sim P^{-1/2}$. By substituting the above relations into the expression relating L and d_0/P, discussed above, a modified expression of transition length to pressure that features a stronger dependence ($L \sim P^{-3/2}$) on pressure emerges. For the loading conditions and geometry of the experimental configuration presented in **Figure 1**, this expression becomes

$$L = f(s)\frac{1+v}{\pi}G\frac{\mu^s-\mu^d}{\left(\sin\alpha-\mu^d\cos\alpha\right)^2}2c\left(\frac{\mu^s-\mu^d}{\mu^s}\right)^M \times\sqrt{H}a_0P^{-3/2}\cos^{-1}\alpha \qquad [8]$$

As shown in **Figure 10**, this modified relation agrees very well with the experimental data presented in this paper for appropriate choices of material and geometrical parameters of the micromechanics contact model. In plotting **Figure 10**, the hardness and the elastic constants of Homalite were used while the friction coefficients were taken to be $\mu^s = 0.6$ and $\mu^d = 0.2$ as estimated above. The values for the constants c and μ were from the experiments of Ohnaka (2003). The asperity radius a_0 was taken to be equal to 1.78 μm, for a best fit to the experimental data. This value was found to be consistent to averaged measurements of asperity radii obtained through microscopic observations of the sliding surfaces. Relation [8] can also easily be generalized to biaxial loading conditions. Indeed, if $P_1 = P$ and $P_2 = bP$ denote the horizontal and vertical components of the applied loads such a situation, the transition length can be expressed as

$$L = f[s(\alpha)]\frac{1+v}{\pi}G\frac{2c(\mu^s-\mu^d)^{1+M}(\cos^2\alpha+b\sin^2\alpha)^{1.5}}{[(1-b)\sin\alpha\cos\alpha-\mu^d(\cos^2\alpha+b\sin^2\alpha)]^2} \times(\mu^s)^{-M}\sqrt{H}a_0P^{-1.5} \qquad [9]$$

In addition to tectonic loading magnitude and direction relative to fault inclination, eqn [9] involves the material properties G, v, and H, the frictional properties μ^s and μ^d, as well as fault related length scale a_0. Because of this, it can perhaps be used as a simple scaling relation connecting laboratory observations to reports of supershear transition in 'large' natural earthquake rupture events such as these related to 2001 Kunlushan event in Tibet.

For making order of magnitude level contact with seismological applications, we rewrite the general form of eqn [5] in terms of the effective stress $\tau^e = \tau - \tau^f$, a commonly used and estimated seismological parameter. Equation [5] can be rewritten as

$$L = f(s)\frac{1+v}{\pi}(1+s)\frac{Gd_0}{\tau^e} \qquad [10]$$

Application of this equation to both seismic faulting and laboratory data allows us to scale the transition length L from laboratory to seismological conditions. The effective stress τ^e in our experiment is chosen to be of the same order of magnitude as that measured in seismology. The ratio of rigidity of the Earth's crust to Homalite is about 25. From the experiment described in **Figure 8** ($P = 9$ MPa and $\alpha = 25°$) the transition length is estimated to be $L = 20$ mm from which $d_0 = 10$ μm is obtained using eqn [4]. The values of d_0 for large earthquakes are often estimated as 50 cm to 1 m (Ide and Takeo, 1997). If s is approximately the same under laboratory and crustal conditions, the transition length for earthquakes can be estimated to be in the range between 25 and 50 km. Because s can be different and the estimate of d_0 for earthquakes is very uncertain at present, this value

should only be taken as an order of magnitude estimate. Nevertheless, it is of the same order as that inferred for Kunlunshan (100 km) and for Denali (75 km). The large transition length required for supershear is perhaps one of the reasons that relatively few earthquake events have been observed to feature such high rupture velocities and that all of them correspond to large magnitude earthquakes. If the tectonic stress is well below the static fault strength (i.e., large s), the transition length becomes too large for many earthquake ruptures to attain supershear. The observation that during several large earthquakes the rupture velocity became very fast, possibly supershear, suggests that the tectonic stress is fairly close to the static fault strength (i.e., small s), which has important implications for the evolution of rupture in large earthquakes.

4.06.3.4 Discussion

A final word of caution is in order here. It has recently been shown in recent preliminary numerical simulations of the homogeneous experiments described in this section (Lu *et al.*, 2005) that the detailed variation of the transition length on the load P may also strongly depend on the nature of the rupture nucleation process. The severity of this effect is still unknown. If this dependence proves to be significant, an experimental investigation involving various pressure release histories (intensities, rise times, and durations) will be necessary in order to complete the experimental investigation of supershear transition presented above. A related issue regarding the influence of nucleation conditions in a bimaterial setting (see Section 4.06.4) was recently discussed by Shi and Ben-Zion (2006).

4.06.4 Directionality of Ruptures along Faults Separating Weakly Dissimilar Materials: Supershear and Generalized Rayleigh Wave Speed Ruptures

In most mature fault systems, the elastic properties vary across the fault plane (Magistrale and Sanders, 1995; Peltzer *et al.*, 1999) and also the shear wave speeds may vary by as much as 30% (Ben-Zion and Huang, 2002; Cochard and Rice, 2000). Recently, Rubin and Gillard (2000) studied several thousands of pairs of consecutive earthquakes that occurred on a segment of the central San Andreas fault, south of the Loma Prieta rupture. Among the second events of each pair, they found that over 70% more occurred to the northwest than to the southwest. They interpret this asymmetry as being a result of the contrast in material properties across the fault. Indeed, at this location of the San Andreas Fault, the rock body is more compliant northeast of the fault than it is southwest (Eberhart-Phillips and Michael, 1998).

Early theoretical and numerical studies of rupture that employ various frictional laws with a constant coefficient of friction which would be independent of slip or slip rate (Adams, 1995; Andrews and Ben-Zion, 1997; Cochard and Rice, 2000; Harris and Day, 1997; Heaton, 1990; Ranjith and Rice, 1999; Rice *et al.*, 2001; Weertman, 1980) imply that if rupture occurs on the boundary between two frictionally held solids, having different elastic properties and wave speeds, such a rupture may preferentially propagate in a particular direction. This is the same direction as the direction of slip in the lower wave speed solid. This direction is often referred to as the 'preferred' direction (Ben-Zion 2001). Since such implied directionality of fault rupture may have a profound influence on the distribution of damage caused by earthquake ground motion, it would be extremely useful if this behavior could be confirmed under controlled laboratory conditions. While many of the physical aspects of dynamic rupture (including supershear) are recently becoming progressively clearer in relation to homogeneous faults (Ben-Zion, 2001; Rice, 2001; Rosakis, 2002; Xia *et al.*, 2004), the behavior of spontaneously nucleated ruptures in inhomogeneous faults, separating materials with different wave speeds, has until recently (Xia *et al.*, 2005) remained experimentally unexplored. In this section we elaborate on the results of this experimental investigation and we discuss their implications in relation to the theoretical concept of a 'preferred' rupture direction in the presence of bimaterial contrast.

The recent large earthquakes (1999 Izmit and Düzce) and the seismic migration history along the North Anatolian Fault may represent a unique field example of the effect of the material contrast across the fault. The 1999 Izmit and Düzce events featured both supershear and sub-Rayleigh rupture branches (Bouchon *et al.*, 2001). Most significantly, they are the last of a series of large ($M \geq 6.8$) earthquakes that have occurred since 1934 in the North Anatolian Fault. These earthquakes have occurred on a rather long and allegedly inhomogeneous fault system (Zor *et al.*, 2006; Le Pichon *et al.*, 2003) that has hosted tens of major migrating earthquakes in the past century. Following the work of Stein *et al.* (1997) and

of Parsons *et al.* (2000), tens of large ($M \geq 6.8$) earthquakes occurred over 1000 km along the North Anatolian Fault between the 1939 earthquake at Ercinzan and the 1999 Izmit and Düzce earthquakes. Such a long series of earthquakes are believed to be textbook examples of how the transfer of stress from a recent nearby event can trigger the next major event in due time. This presumably happens by adding or transferring stress to the fault segment, which is adjacent to the tips of a segment that has last failed. The stress distribution is highly nonuniform and it occurs in addition to the long-term stress renewal and to the pre-existing stress inhomogeneities. However, as much as this model seems to be complete and convincing, a few questions remain that need to be resolved and are of relevance to the work described here.

4.06.4.1 Two Types of Ruptures along Inhomogeneous Fault Systems

Inhomogeneous faults separate materials with different wave speeds. When such faults experience spontaneous rupture, the equibilateral symmetry, expected in the homogeneous case, is broken. This leads to various forms and degrees of rupture directionality. Dynamic rupture along bimaterial interfaces is known to involve substantial coupling between slip and normal stress (Ben-Zion, 2001; Rice, 2001; Weertman, 1980). As a consequence, the relative ease or difficulty for a rupture to propagate in a specific direction along a bimaterial interface should be closely related to the degree of mismatch in wave speeds. This phenomenon is also related to the fault's frictional characteristics which play a dominant role in facilitating the phenomenon of normal shear coupling. For bimaterial contrast with approximately less than 35% difference in shear wave speeds (as in the case of most natural faults), generalized Rayleigh waves can be sustained. These waves are waves of frictionless contact propagating at a speed, C_{GR}, called the generalized Rayleigh wave speed (Rice, 2001).

The 1980 rupture solution by Weertman (1980) involves a dislocation like sliding pulse propagating subsonically with a velocity equal to C_{GR} along an interface governed by Amonton–Coulomb friction. However, the classical Amonton–Coulomb's description has been shown to be inadequate for addressing fundamental issues of sliding (Ranjith and Rice, 2001), since sliding becomes unstable to periodic perturbations. Instability, in the above sense, implies that periodic perturbations to steady sliding grow unbounded for a wide range of frictional coefficient and bimaterial properties (Adams, 1995; Renardy, 1992). The growth rate is proportional to the wave number. This issue has been discussed in relation of both homogeneous and bimaterial systems (Rice, 2001). For bimaterial systems and in particular, when generalized Rayleigh waves exist, Ranjith and Rice (2001) demonstrate that unstable periodic modes of sliding appear for all values of the friction coefficient. Mathematically, instability to periodic perturbations renders the response of a material interface to be ill-posed (no solution exists to the problem of growth of generic, rather than periodic, self-sustained perturbations to steady sliding). The problem is regularized by utilizing an experimentally based frictional law (Prakash and Clifton, 1993), in which shear strength in response to an abrupt change in normal stress evolves continuously with time (Cochard and Rice, 2000; Ranjith and Rice, 2001). In such a case, the problem becomes well-posed and generic self-sustained pulse solutions exist while numerical convergence through grid size reduction is achieved (Cochard and Rice, 2000; Coker *et al.*, 2005). However, despite the fact that this special frictional law provides regularization, self-sustained slip pulses may still grow in magnitude with time. This is a phenomenon that has been demonstrated numerically by Ben-Zion and Huang (2002). Moreover, self-sustained pulses were found to exist and to propagate at discrete steady velocities and at specific directions along the inhomogeneous interface by analytical (Ranjith and Rice, 2001) and numerical means (Andrews and Ben-Zion, 1997; Cochard and Rice, 2000).

Two types of such steady, self-sustained pulses were discovered by Ranjith and Rice (2001) theoretically. Consistent with Weertman's analysis (Weertman, 1980), the first type corresponds to rupture growth in the direction of sliding of the lower wave speed material of the system. This direction is referred to in the literature (Ben-Zion, 2001; Rice, 2001) as the 'positive' direction and sometimes as the 'preferred' direction (Ben-Zion, 2001). The rupture pulses belonging to this type are subshear and always propagate with a steady velocity $V = +C_{GR}$, where the plus denotes growth in the 'positive' direction. Thus, in this work these rupture pulses will be referred to as 'positive' generalized Rayleigh ruptures and will be abbreviated as '+GR' ruptures. The second type of self-sustained rupture corresponds to growth in the direction opposite to that of

sliding in the lower wave speed material of the bimaterial system. This direction is often referred to as the 'negative' direction or 'opposite' direction (Cochard and Rice, 2000). Such ruptures are supershear and they always propagate with a steady velocity that is slightly lower than the P-wave speed of the material with the lesser wave speed $\left(V = -C_P^2\right)$. Such ruptures are generated for sufficiently high values of the coefficient of friction (Ranjith and Rice, 2001) and are less unstable than the '+GR' ruptures described above (Cochard and Rice, 2000). In the present paper such ruptures will be abbreviated as '$-P_{SLOW}$' ruptures. This second type of rupture was later studied by Adams (2001), who showed that the leading edges of these supershear (intersonic) ruptures are weakly singular, a result which is consistent with numerical analysis (Cochard and Rice, 2000).

From the point of view of numerics, the early work of Andrews and Ben-Zion (1997) has brought to light the persistence and interesting properties of rupture pulses of the '+GR' type. This was possible even in the ill-posed context of sliding governed by Amontons–Coulomb friction before much of the theoretical concepts were at hand. In their work, the sliding '+GR' pulses were triggered by a local release of interfacial pressure spread out over a finite region at the interface and over finite time. No pulses of the second type ('$-P_{SLOW}$' pulses) were excited in these early simulations despite the fact that the coefficient of friction was high enough to have permitted their existence as suggested by the modal analysis of Ranjith and Rice (Cochard and Rice, 2000). Perhaps this is not altogether surprising if one considers the fact that in these early simulations, nucleation was achieved by a 'biased' localized stress drop having a favored propagation direction. In particular, the 'positive' direction of growth was artificially seeded in the nucleation process (Ben-Zion, 2006). The subsequent numerical simulations of Cochard and Rice (2000), which utilized the modified Prakash and Clifton Law, still featuring a constant coefficient of friction, were able to sequentially excite regularized self-sustained pulses of both types. This was achieved by introducing small changes in the parameters of the friction law and in the geometry of the nucleation zone. At the same time, no simultaneous excitation of both modes was reported. Moreover, the '$-P_{SLOW}$' pulses were found to be slightly more difficult to excite than the '+GR' pulses. However, the degree of relative difficulty was not examined in detail. In partial agreement to the above 2-D numerical studies, Harris and Day (1997) demonstrated the simultaneous existence of both types of sliding modes, propagating in opposite directions during the same rupture event. They considered various bimaterial and trilayered configurations featuring modest wave speed mismatch and a slip-weakening frictional law. As first pointed out by Xia *et al.* (2005), the inconsistency between the various numerical studies is most probably due to the different friction laws utilized. Indeed, up to very recently, all studies except this of Harris and Day (1997) have assumed a constant coefficient of friction. The need for experimental analysis becomes clear at this point since only experiments can be used to judge the physical relevance of various assumed friction laws and to validate various proposed numerical methodologies. As emphasized by Xia *et al.* (2005), the goal of some of the early theoretical and numerical studies (Adams, 1995, 1998; Andrews and Ben-Zion, 1997; Cochard and Rice, 2000; Ranjith and Rice, 2001; Weertman, 1980) was to investigate what kind of unstable slip would develop on a surface which, as judged from conventional friction notions, was superficially stable, in the sense that its friction coefficient, μ, did not decrease, or vary otherwise, with slip and/or slip rate. For most brittle solids, however, ample evidence exists that, μ does decrease with increase of slip and/or slip rate (or, more fundamentally, μ varies with slip rate and contact state). As a result, a proper model for natural faulting along a bimaterial interface should include both a weakening of μ and the slip-normal stress coupling effects of the bimaterial situation. Indeed, various models for μ are expected to strongly influence the effectiveness of the bimaterial contrast in either enhancing or retarding rupture growth. Such a weakening model was first included by Harris and Day (1997) in a bimaterial context. Given the above, Xia *et al.* (2005) first pointed out that it would be an invalid interpretation of the results of the earlier set of papers (Adams, 1995, 1998; Andrews and Ben-Zion, 1997; Cochard and Rice, 2000; Ranjith and Rice, 2001; Weertman, 1980) to conclude that the rupture scenarios (including preference for specific rupture mode) which they predict constitute the full set of possibilities available to a real earthquake, of which μ decreases with increasing slip and/or slip rate. Indeed, according to Xia *et al.* (2005) the consistently bilateral nature of rupture predicted by Harris and Day (1997) is an indication of the effect of including a strongly slip-weakening frictional law in their calculations. In addition, recent refined calculations by Harris and Day (2005) have reconfirmed their conclusions

within a fully 3-D setting. These conclusions, however, have recently been partially challenged by Shi and Ben-Zion (2006) on the basis of an extensive parameter study. A more complete analysis of this on-going discussion will be presented in Section 4.06.4.5.

In the immediately following section, we elaborate on recent experiments (Xia *et al.*, 2005) designed to investigate some of the issues discussed above. We also briefly summarize the recent debate on the existence or nonexistence of a preferable rupture direction which has been triggered by these experiments.

4.06.4.2 Experimental Setup

The experiments described in this section mimic natural earthquake rupture processes in fault systems, where bimaterial contrast between intact rock masses seldom feature more than a 30% difference in shear wave speeds (Rice, 2001).

The laboratory fault model is similar to the one described in **Figure 1(a)** and is shown in **Figure 12**. The figure shows a Homalite-100 plate (material 1, top) and a polycarbonate plate (material 2, bottom) that are held together by far-field load, P. The higher wave speed material at the top (Homalite-100) has a shear wave speed $C_S^1 = 1200\,\mathrm{m\,s^{-1}}$ and a longitudinal wave speed $C_P^1 = 2498\,\mathrm{m\,s^{-1}}$. The lower wave speed material at the bottom (Polycarbonate) has a shear wave speed $C_S^2 = 960\,\mathrm{m\,s^{-1}}$ and a longitudinal wave speed $C_P^2 = 2182\,\mathrm{m\,s^{-1}}$. The fault is simulated by a frictionally held contact interface forming an angle to the applied load that is varied to mimic a wide range of tectonic load conditions. Spontaneous rupture is triggered at the hypocenter through the exploding wire mechanism described in Chapter 1.02. The static compressive load P is applied through a hydraulic press. By arbitrary convention, the fault line runs in the east–west direction with the lower wave speed solid located at the south side. This choice was motivated in order to eventually facilitate comparisons with the 1999 Izmit earthquake in Turkey. According to the work of Le Pichon *et al.* (2003), the material to the south of the North Anatolian Fault (at its western end near the sea of Marmara) is the lower wave speed solid. As viewed from the camera, a rupture will produce right lateral slip.

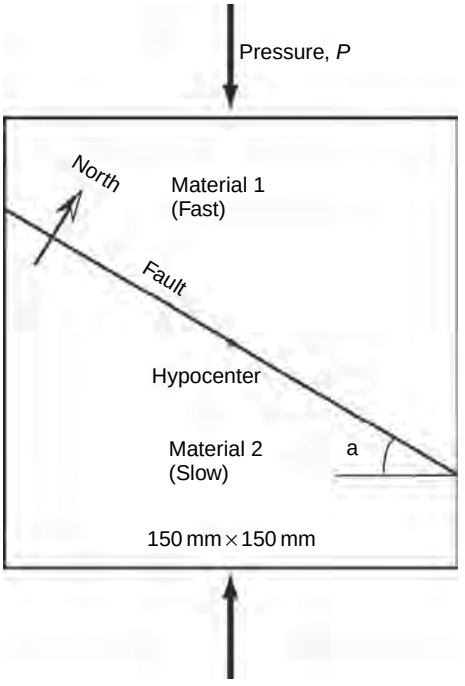


Figure 12 Laboratory earthquake fault model composed of two photoelastic plates of the same geometry.

The ratio of shear wave speeds, $C_S^1/C_S^2 = 1.25$, was chosen to be within the naturally occurring bimaterial range so that the interfacial phenomena can be applied to some field observations. In particular, the bimaterial difference is big enough to allow for a high-enough growth rate of sliding instabilities (Rice, 2001) and to permit the clear distinction between various wave speeds. Within roughly the same range generalized Rayleigh waves exist as well. The shear wave speeds were directly measured for each material by following the shear wave fronts through high-speed photography and photoelasticity. The listed P wave speeds were calculated by using measured values of Poisson's ratios ($v^1 = 0.35$, $v^2 = 0.38$) and by using the listed shear wave speeds. An independent measurement of the P-wave speeds in the plates using ultrasonic transducers confirmed these listed values to within 5%. The value of C_{GR} can be determined (Rice, 2001) from the equation:

$$f(V) = \left(1 - b_1^2\right) a_1 G_2 D_2 + \left(1 - b_2^2\right) a_2 G_1 D_1 = 0$$

where $a_n = \sqrt{1 - V^2/(C_P^n)^2}$, $b_n = \sqrt{1 - V^2/(C_S^n)^2}$, $D_n = 4a_n b_n - \left(1 - b_n^2\right)^2$, V is the rupture speed, G_n are rigidities of the two materials, and $n = 1,2$. Substituting the material constants for Homalite-100 and Polycarbonate into the equation, the generalized Rayleigh wave speed is calculated as $C_{GR} = 950\,\mathrm{km\,s^{-1}}$. This is a value that is extremely close to the shear wave speed of Polycarbonate.

4.06.4.3 Experimental Results

More than 30 experiments featuring different angles, α (20°, 22.5°, and 25°), and far-field loading, P (10–18 MPa), were performed and the repeatability of rupture events was confirmed. The higher level of angles was limited by the static frictional characteristics of the interface. Depending on P and α, three distinct and repeatable rupture behaviors were observed. In all cases, the two separate, semicircular traces of the shear waves in the two materials were clearly visible as discontinuities in the maximum shear stress field. The ruptures were always bilateral and became progressively asymmetric with time, within the time window of all experiments. As shown in Figure 13, two distinct rupture tips, one moving to the west and the other moving to the east, with velocities V_E and V_W, respectively, were identified by a distinct concentration of fringe lines. For this case (case-1), both tips propagate at subshear velocities $V_E < V_W < C_S^2 < C_S^1$. Differentiation of the rupture length–time histories, obtained from a series of high-speed images, allows for the estimation of the rupture velocity histories. On the one hand, the rupture moving to the west is the one propagating in the direction of sliding of the lower wave speed material (positive direction). Within experimental error this rupture was found to grow at a constant velocity equal to the speed of the generalized Rayleigh waves ($V_W = 950\,\mathrm{m\,s^{-1}} \approx +C_{GR}$). The rupture moving to the east, on the other hand, was the one propagating in the direction opposite to that of sliding in the lower wave speed material (opposite direction). This rupture grew at an almost constant sub-Rayleigh velocity of $V_E = -900\,\mathrm{m\,s^{-1}}$, which is clearly slower than the Rayleigh wave speed, C_R^2, of the slower wave speed material. The observations were very similar for smaller angles, α, and compressive loads, P, as well. The rupture to the west (positive direction) always propagated with $+C_{GR}$. The rupture to the east remained sub-Rayleigh $\left(V_E < C_R^2 < C_R^1\right)$. However, its velocity varied across experiments with different load levels and angles. In particular, smaller angles of α (or smaller values of the s factor described earlier) and lower P resulted in V_E being lower fractions of C_R^2. Judging from the number of near-tip fringes per unit area, the eastward moving rupture resulted in a visibly smaller level of stress drop than the one moving to the west. This observation is consistent with predictions by Cochard and Rice (2000) and Adams (2001) who also predict a weaker singularity for ruptures moving in the negative direction.

A very distinct but equally repeatable rupture case (case-2) was observed for higher values of α and P. These conditions correspond to much higher values of driving stress or to conditions closer to incipient uniform sliding of the entire interface (smaller values of s). A typical example corresponding to $\alpha = 25°$ and $P = 17$ MPa is shown in Figure 14. In this case the rupture is still bilateral with a westward tip trailing

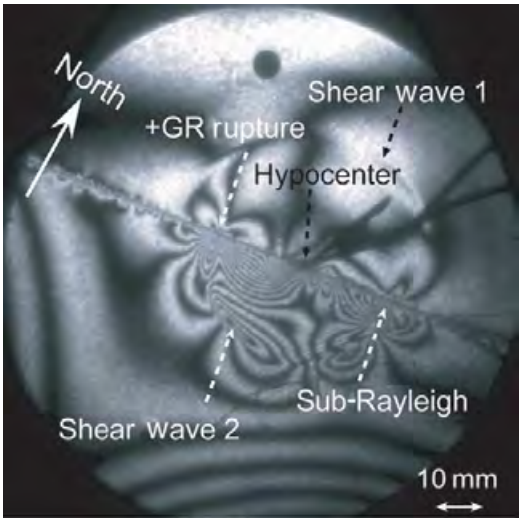


Figure 13 The photoelastic patterns for an experiment with $\alpha = 22.5$, $P = 17$ MPa, and smooth surface. Both ruptures to the east and the west are subshear (case-1).

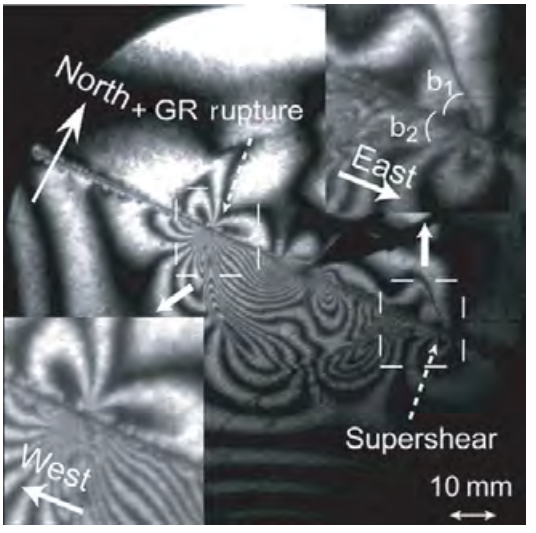


Figure 14 For $\alpha = 25°$, $P = 17$ MPa, and smooth surface finish the bilateral rupture features two distinct tips. The one moving to the west (positive direction) has a velocity $V_W \approx +C_{GR}$, while the one moving to the east (opposite direction) is supershear. (case-2).

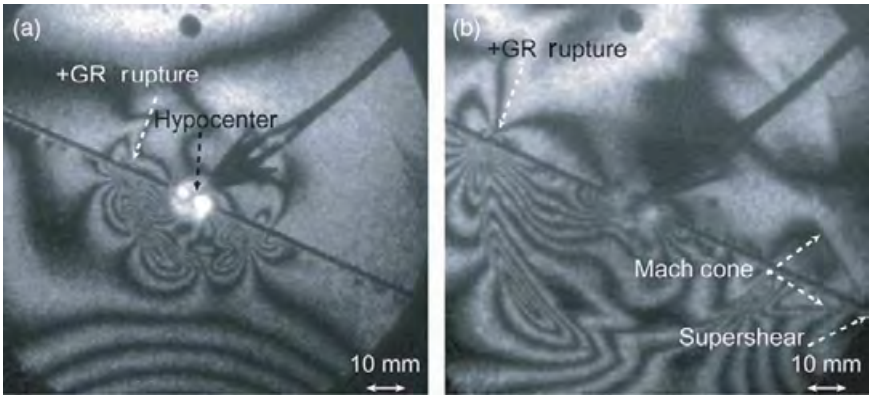


Figure 15 Experimental results for $\alpha = 25°$, $P = 13$ MPa, and rough surface showing transition of the eastward moving rupture to supershear. The westward rupture retains a constant velocity $V_W \approx +C_{GR}$ (case-3).

behind both shear wave traces. This tip moves at a constant velocity $V_W \approx +C_{GR}$ along the 'positive' direction. This observation is identical to the situation described above in relation to lower values α and P. The eastward moving tip however is clearly different from the previously described case. Its tip is moving with a velocity faster than both the shear wave speeds. Moreover, its structure, shown in detail in the upper insert of **Figure 14**, is distinctly different to the structure of the sub-Rayleigh, westward moving rupture shown in the lower inset.

As conclusive proof of its supershear velocity, two distinct shear shock waves are clearly visible. The magnitude of the velocity of the eastward rupture V_E is 1920 m s^{-1}, which is approximately 12% less than the P-wave speed, C_P^2, of the lower wave speed material. V_E is also equal to 1.6 C_S^1, or is slightly higher than $\sqrt{2}C_S^1$. The upper inset in **Figure 14** shows two clear lines of discontinuity in the maximum shear contours of photoelasticity. Each of these lines (shear shock waves) is located at two different angles $\beta_1 = 41°$ and $\beta_2 = 30°$, to the north and to the south of the fault respectively. The two angles β_n ($n = 1, 2$) are related to the shear wave speeds C_S^n and to the rupture velocity V_E, by $\beta_n = \sin^{-1}\left(V_E/C_S^n\right)$. This relation provides independent means of estimating V_E from each individual frame of the high-speed camera record without reliance on the less accurate rupture length history. Both methods yield consistent values of $V_E = -1920$ m s^{-1}.

Both cases described above feature westward moving ruptures that are of the '+GR' type. Irrespective of the values of α and P, these ruptures have a constant speed $V_W \approx +C_{GR}$, and they propagate in the 'positive' direction. However, those two cases also feature eastward ruptures that are distinctly different in nature. For

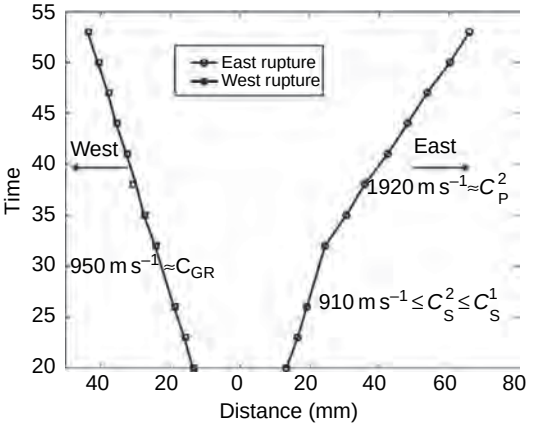


Figure 16 Rupture length plot of an experiment for $\alpha = 25°$, $P = 13$ MPa, and rough surface finish.

sufficiently low P and α (or large s), the eastward ruptures, which propagate in the opposite direction, are purely sub-Rayleigh within the time window of our experiments. For large-enough P and α however, eastward ruptures propagate in the opposite direction with a constant supershear velocity, which is slightly less than C_P^2 and are thus of the '$-P_{SLOW}$' type.

To visualize an intermediate situation and a controlled transition from one case to the other within the field of view, P was reduced to 13 MPa. For this case (case-3), **Figure 15** shows a smooth transition from case-1 to case-2 within the same experiment. While the westward rupture remains of the '+GR' type throughout the experiment, the eastward rupture jumps from a constant sub-Rayleigh velocity (-910 m s^{-1}) to a constant supershear velocity (-1920 m s^{-1}), and thus transitions to the '$-P_{SLOW}$' type. The rupture length plot of **Figure 16** also

shows the abrupt transition of the eastward rupture from a sub-Rayleigh velocity to a velocity slightly less than C_p^2. This happens at a transition length, L, which is approximately equal to 25 mm. However, the westward rupture retains its constant $+C_{GR}$ velocity throughout the experiment. The eastward transition behavior of case-3 is qualitatively similar to the one we have discussed in relation to homogeneous interfaces (Xia *et al.*, 2004), while the transition length, L, is also a decreasing function of α and P. Most important to the discussion of the present paper is the observation that the ruptures that propagate to the opposite direction require a certain minimum rupture length before they become supershear. This observation provides a clear intuitive link between super shear growth in the 'opposite' direction and large earthquakes. In contrast, no such transition was observed for 'positively' growing '+GR' ruptures irrespective of α, P, and rupture length. As a result, the experiments do not provide an obvious link between 'positively' growing ruptures and large earthquakes.

In other words, a certain minimum length of rupture growth in the negative direction is required before such a transition can be observed. This suggests that in smaller earthquake events, such as the 2004 Parkfield rupture such a transition may not have had the chance to happen before the negative moving rupture was arrested. (for more on Parkfield, see Section 4.06.4.5).

In Section 4.06.3.3, we have discussed the dependence of the transition length L on the uniaxial pressure P. In the homogeneous case there is a very well-defined point for transition, while in the inhomogeneous case the transition point is not always so clear. This difference again is due to the presence of a material contrast. In the homogeneous case, there is an energetically forbidden velocity zone between C_R and C_S (Rosakis, 2002). As a result, the secondary crack is initiated exactly at the shear wave front. In contrast for inhomogeneous fault systems, the forbidden zone no longer exists (Rosakis, 2002) and the subshear crack in the opposite direction accelerates to the supershear speed in a smoother way (**Figure 16**). Nevertheless, we can still define the transition length where a rapid speed change occurs. The plot of transition length L is in **Figure 17**. L has a weaker dependence on P ($L \sim P^{-0.4}$) than the homogeneous case ($L \sim P^{-1.5}$). This is expected because for the inhomogeneous case, and for subshear rupture speeds, the coupling between the slip and the normal traction causes a local rupture-tip compression. This compression increases the resistance to slip for rupture growing in the negative direction. On the other hand, the shear traction driving the rupture is constant and hence it takes a longer slip distance for a rupture in bimaterial interface to reach the supershear velocity.

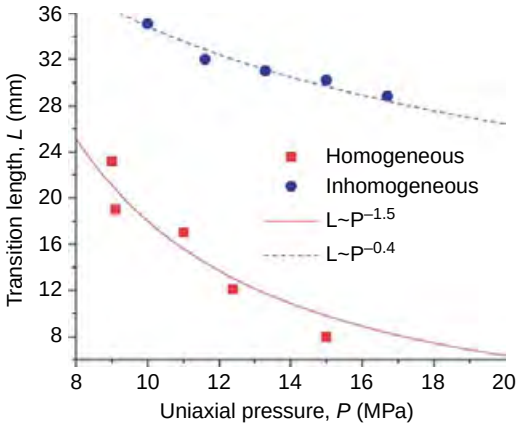


Figure 17 Transition length as a function of pressure P for experiments with $\alpha = 25°$. The data and fitting curves for the homogeneous case are included for comparison.

4.06.4.4 Comparison of the Experimental Results to Early Numerical and Theoretical Studies

The experiments by Xia *et al.* (2004, 2005) described above provide the first full-field and real-time visualization of dynamic frictional rupture events occurring along inhomogeneous interfaces, which feature low-wave-speed mismatch such that the generalized Rayleigh wave speed can be defined. While it is very difficult to access whether the ruptures are pulse-like, crack-like, or a mixture of the two, the observations confirm the existence of two distinct self-sustained and constant speed rupture modes. These bare strong similarities to the ones that have been theoretically and numerically predicted over the recent years (Andrews and Ben-Zion, 1997; Ben-Zion, 2001; Ben-Zion and Huang, 2002; Cochard and Rice, 2000; Rice, 2001; Weertman, 1980). In particular, a '+GR' type of rupture mode is always excited instantaneously along the 'positive' direction of sliding. Besides the '+GR' rupture mode, a '$-P_{SLOW}$' mode is observed as long as the rupture propagating in the 'opposite' direction is allowed to grow to sufficiently long distances from the hypocenter. The triggering of the '$-P_{SLOW}$' mode is always preceded by a purely sub-Rayleigh, crack-like

rupture whose velocity depends on loading, on geometry, and on the bimaterial characteristics. Therefore, the existence of this preliminary and apparently transient stage is one of the main differences with the early numerical predictions (Andrews and Ben-Zion, 1997; Cochard and Rice, 2000). However, its existence does not contradict early theoretical studies (Adams, 2001; Ranjith and Rice, 2001), which can only predict stable rupture events whose constant velocities relate to the wave speeds of the bimaterial system.

A far more striking difference to some of the early numerical predictions (Andrews and Ben-Zion, 1997; Ben-Zion and Huang, 2002; Cochard and Rice, 2000) is the consistent experimental observation of bilateral slip. In contrast to the experiments, the above numerical predictions seem capable only of exciting one or the other of the two self-sustained rupture modes (Cochard and Rice, 2000), giving rise to purely unilateral rupture events. They also seem to primarily favor the triggering of the '+GR' mode in low-wave-speed mismatch bimaterial systems (Andrews and Ben-Zion, 1997). This kind of preference has led to the labeling of the 'positive' direction as the 'favored' rupture direction and of $+C_{GR}$ as the 'favored' rupture velocity. These numerical results indirectly support the closely related notion of ruptured directionality (McGuire *et al.*, 2002). A notable exception to this rule is provided by the early numerical analysis by Harris and Day (1997), as well as their subsequent work (Harris and Day, 2005; Andrews and Harris, 2005) which consistently reports asymmetric bilateral rupture growth in a variety of low-speed contrast, inhomogeneous fault systems. These results are qualitatively very similar to the experimental observations of cases 1 and 2. In particular, the latter 3-D results by Harris and Day (2005) also clearly report on sub-Rayleigh to supershear transition for the rupture propagating along the negative direction. As briefly discussed by Cochard and Rice (2000) and Xia *et al.* (2005), the excitation of various modes or their combinations should be related to the details of the numerically or experimentally implemented triggering mechanisms. In an attempt to further reconcile the observed differences between various models and the experiments, Xia *et al.* (2005) have noted that unstable slip rupture propagation has also been observed (Xia *et al.*, 2004) on homogeneous Homalite/Homalite and Polycarbonate/Polycarbonate interfaces. Such unstable rupture growth would be possible only if there was a substantial reduction of the friction coefficient with slip and/or slip rate, and hence such reduction must be a property of both these model materials when sliding against each other. It is then plausible to assume that a similar reduction of friction coefficient occurs along the Homalite/Polycarbonate interface, and to thus infer that its rupture behavior should not be expected to fully correspond to the idealized models of a dissimilar material interface with constant coefficient of friction (Adams, 1995, 1998; Andrews and Ben-Zion, 1997; Cochard and Rice, 2000; Ranjith and Rice, 2001; Weertman, 1980). Indeed, the stronger the weakening becomes, the reduced frictional resistance is expected to further neutralize the effect of bimaterial contrast on normal shear coupling thus rendering the 'preferable' rupture direction less preferable than originally thought.

The present experiments neither support exclusivity nor show a strong preference for rupture direction. Although they support the idea that frictional ruptures that grow in the positive direction will always do so at a specific constant velocity ($V = + C_{GR}$), they still allow for a significant possibility of self-sustained supershear ruptures growing in the opposite direction. This possibility becomes significant, provided that their transient, sub-Rayleigh precursors grow for a large-enough length and are not arrested prior to transitioning to supershear. The requirement of a critical transition length along the 'opposite' direction provides a link between large earthquakes and the occurrence of self-sustained supershear rupture in the 'opposite' direction. One perhaps can contemplate the existence of a weak statistical preference for positively growing ruptures, since this link to large earthquakes is absent for '+GR' ruptures.

4.06.4.5 The Parkfield Earthquake Discussion in the Context of Experiments and of Recent Numerical Studies

The Parkfield earthquake sequence presents a very interesting case in the context of bimaterial rupture and the issue of the existence, or lack of, a preferable rupture direction. The slip on the San Andreas Fault is right-lateral, and the crust, near Parkfield, on the west side of the fault features faster wave speeds than the east side (Thurber *et al.*, 2003). The three most recent major Parkfield earthquakes ruptured the same section of the San Andreas Fault. The rupture directions of the 1934 and the 1966 events were southeastward (positive direction), whereas the 2004 earthquake ruptured in the opposite direction (negative direction). Following the discussion above, results from the early constant friction coefficient

studies in bimaterials would imply that the positive (southeastward) direction of the 1934 and the 1966 events is the preferable direction. According to this notion, the negative (northwestward) direction of the 2004 earthquake would not be favored. The experiments described above (Xia *et al.*, 2005), however, have clearly demonstrated that a rupture in the negative direction can indeed occur at least in model materials. Such a rupture, depending on tectonic loading conditions and on available rupture length, may grow at either sub-Rayleigh or at supershear rupture speeds. Motivated by the most recent 2004 Parkfield event and supported by the Xia *et al.* (2005), Harris and Day (2005) extended their earlier 1997 work to three dimensions and discussed their results in relation to extensive observational evidence collected from the past 70 years of earthquakes at the vicinity of Parkfield. They concluded that natural earthquake rupture propagation direction is unlikely to be predictable on the basis of bimaterial contrast.

The scientific discussion on the notion of rupture directionality is however far from being over. In a recent paper, Shi and Ben-Zion (2006) have presented an extensive parameter study for various bimaterial contrasts which clearly shows the effect of assumed frictional properties and nucleation conditions in promoting either unilateral (preferred) or bilateral (nonpreferred) rupture in the presence of bimaterial contrast. Their study utilized a classical linear slip-weakening model of friction of the same type used by Harris and Day (1997, 2005). It shows that for small differences $\mu^s - \mu^d \sim 0.1$ between the static and dynamic coefficient of friction and for nonbiased nucleation events, the bimaterial effect indeed induces ruptures which primarily grow in the positive direction at $+C_{GR}$. These cases are consistent with the constant coefficient friction models. However, when the difference between the two coefficients of friction is taken to be larger than or equal to 0.5 and for 20% shear wave speed contrast, bilateral rupture in both directions becomes possible with speeds equal to $+C_{GR}$ in the positive direction, and depending on the nucleation condition, with either subshear or supershear ($-P_{SLOW}$) speeds in the negative direction. These cases are consistent with the experiments and perhaps with the Parkfield observations. Their results also agree with the above-discussed explanation, originally offered by Xia *et al.* (2005), regarding the root cause of the differences between the experimental results and the prediction of a preferable direction in the early numerical studies. Indeed the difference of friction coefficient in the experiments is expected to be in the range of 0.4–0.5 (see Section 4.06.2.2.3 for the homogeneous case) which, within the slip weakening assumption, would promote bilateral ruptures of various types. Moreover, this recent study does not conclusively settle the issue of rupture direction preference in natural earthquakes where frictional characteristics are generally unknown. Nevertheless it clearly shows that the main arbitrator in whether bimaterial contrast is capable of inducing rupture directionality, through normal to shear coupling, is the frictional law. Clearly, much more work is needed to investigate this phenomenon. The inclusion of realistic frictional laws including strong velocity weakening or enhanced rate and state frictional descriptions which regularize the rupture problem could be a first step. Another step would necessitate the numerical and experimental study of a variety of bimaterial contrasts and of more naturally based nucleation conditions. Nucleation conditions are shown by Lu *et al.* (2005) (homogeneous case) and by Shi and Ben-Zion (2006) and Rubin and Ampuero (2006) (inhomogeneous case) to have a strong influence on subsequent rupture growth. We believe that transition lengths for example may be heavily dependent on nucleation conditions. The recent papers by Andrews and Harris (2005) and Ben-Zion (2006), as well as the work of Rubin and Ampuero (2006) are also very important recent references reflecting the on-going discussion on this subject.

4.06.4.6 Discussion of the Historic, North Anatolian Fault Earthquake Sequence in View of the Experimental Results

The 1999, M7.4, Izmit earthquake in Turkey is perhaps a prime example of a recent large earthquake event for which both modes of self-sustained rupture may have been simultaneously present, as is the case in our experiments. The event featured right-lateral slip and bilateral rupture of a rather straight strike-slip segment of the North Anatolian Fault shown in **Figure 18**. As reported by Bouchon *et al.* (2001), the westward propagating side of the rupture grew with a velocity close to the Rayleigh wave speed, while the eastward moving rupture grew at a supershear velocity that was slightly above the $\sqrt{2}C_S$ times the shear wave speed of crustal rock. The visible difference in the nature of the seismographs from stations ARC and SKR situated almost equidistantly to the hypocenter (star) support their conclusion (see bottom of **Figure 18**). Since the laboratory ruptures of the

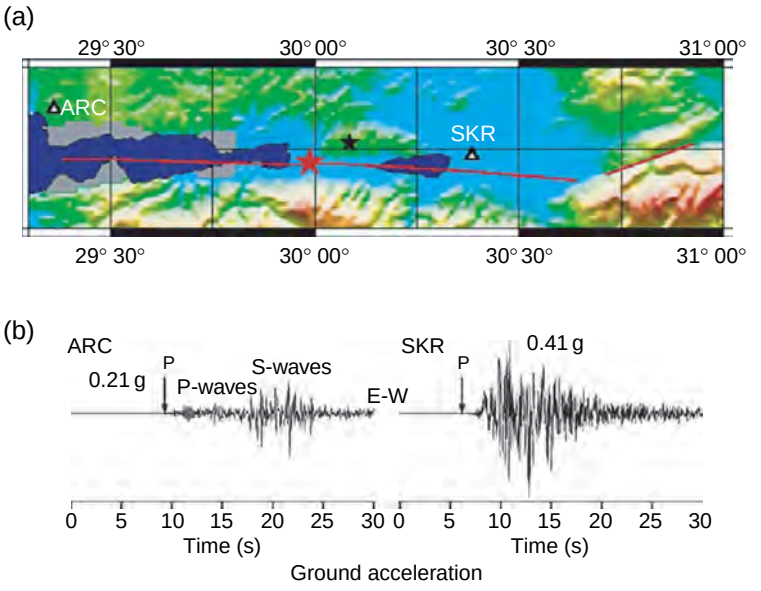


Figure 18 (a) Map view indicating the hypocenter of 1999 Izmit earthquake and two seismological stations ARC and SKR; (b) Seismograms obtained at ARC and SKR. Reproduced from Bouchon M, Bouin MP, Karabulut H, Toksoz MN, Dietrich M, and Rosakis AJ (2001) How fast is rupture during an earthquake? New insights from the 1999 Turkey earthquakes. *Geophysical Research Letters* 28(14); 2723–2726.

current paper are intentionally oriented similarly to the Izmit event (lower-wave-speed material to the south as indicated by Le Pichon *et al.*, 2003), a direct comparison with the case described in **Figure 14** becomes possible and it reveals some striking similarities between Bouchon *et al.*, (2001) interpretation of the earthquake and the experiment. In addition to featuring right lateral slip and asymmetric bilateral rupture, this experimental case (case-2) featured a subshear westward rupture propagating at $+C_{GR}$. To the east however, the rupture propagated at a velocity slightly lower than C_P^2, which also happens to be equal to 1.6 C_S^1 for the particular bimaterial contrast of the experiments. If one interprets the Izmit event as occurring in an inhomogeneous fault with the lower-wave-speed material being situated at the southern side of the fault (as is in the experiment), the field observations and the experimental measurements of both rupture directions and velocities are very consistent. Moreover, when the bimaterial contrast is low enough, the differences between C_{GR} and the average of the two Rayleigh wave speeds, $\left(C_R^1 + C_R^2\right)/2$, as well as the difference between an inferred rupture speed of 1.6 C_S^1 and $\sqrt{2}\left(C_S^1 + C_S^2\right)/2$ would be small enough not to be discriminated by the inversion process, even if the fault geology was completely known. In this respect, the agreement with experiment is as good as it can ever be expected. In addition, viewing the fault as inhomogeneous can explain the choice of direction for both the sub-Rayleigh and the supershear branches, respectively. This choice of rupture direction is consistent with both the present experiments and with the theories reviewed in the introduction.

The 1999 Düzce earthquake can also be interpreted through a similar line of argument used for Izmit. The Düzce rupture also featured right lateral slip (as did all events that occurred in the North Anatolian Fault between 1939 and 1999) and it extended the Izmit rupture zone 40 km eastward (negative direction) through asymmetric bilateral slip (Bouchon *et al.*, 2001). Thus, similar to the Izmit earthquake, numerical modeling by Bouchon *et al.* (2001) indicates both sub-Rayleigh westward and supershear eastward rupture fronts. As a result the direct comparison with case-2 described in **Figure 14** provides an explanation for the two rupture directions and respective velocities, similar to the one given for Izmit. This explanation is of course plausible only if one assumes, once again, that the material to the south of the North Anatolian Fault (at its western end) is the lower-wave-speed solid. Strong evidence supporting this assumption has recently been presented by Le Pichon *et al.* (2003).

By using similar arguments to the ones used for Izmit and Düzce, one can perhaps attempt to provide

a unified rationalization of the seemingly random rupture directions and rupture velocities of the interrelated series of earthquakes that occurred since 1939 along the North Anatolian Fault and ended in 1999 with the Izmit and Düzce events. The following argument requires the assumption that, in average and along its entire length, the North Anatolian Fault features the same type of bimaterial in-homogeneity as the one that has been summarized for Izmit and Düzce. However, rather limited evidence supporting such an assumption is currently available (Zor *et al.*, 2006). If in some average sense this is true, one would expect that the slight majority (60%) of the large ($M \geq 6.8$) earthquake events (i.e., (1939-M7.9), (1942-M6.9), (1944-M7.5), (1951-M6.8), (1957-M6.8), (1967-M7.0)) which featured westward growing ruptures, were probably of the '+GR' type. In other words, this assumption implies that they were classical sub-Rayleigh ruptures that moved with velocity equal to $+C_{GR}$ in the 'positive' direction. The remaining ruptures of the series were 'special' in the sense that they featured dominant eastward growth. As previously detailed, out of the remaining four ruptures of the series, the Izmit and Düzce events were bilateral with a western branch of the '+GR' type (consistent with the others) and an eastward, supershear branch of the '$-P_{SLOW}$' type. The 1943 and 1949 ruptures were purely unidirectional and eastward moving; however, their rupture velocities are not known. If these ruptures are to be consistent with the remaining events in the sequence then they could also have developed as the '$-P_{SLOW}$' type. This possibility is more likely for the 1943 event that featured over 250 km of growth length. As estimated in Section 4.06.2.2.3, this length is much larger than the critical length required for transition to supershear. By observing that the 1943 and the two 1999 (Izmit and Düzce) events were of a higher magnitude than most of the other events of the complete series, as reported by Stein, *et al.* (1997), further supports the assertion that at least three out of four 'special' events featured partial or total supershear growth along the 'opposite' (eastward) direction. The basic support for this assertion is provided by the experimentally established link between large earthquakes and supershear ruptures growing in the 'opposite' direction, and is consistent with the direct evidence of supershear from the two most recent 'special' events of 1999.

Finally, it should be noted that if earthquakes of lesser magnitude (in the range between M6.4–6.8) are also included in the discussion, the North Anatolian Fault series will feature a weak preference for western propagation. This is not very surprising given the above discussed link between large earthquakes and self-sustained supershear along the opposite direction, a link that does not exist for 'positive' (westward growing) '+GR' ruptures. Indeed, in addition to the actual number of ruptures that grew to the east or west, what is of importance here is the actual growth lengths. The results reported by Stein *et al.* (1997) show that the total length of westward growth is only slightly higher than that of eastward growth. This is again consistent with experiments which show that the self-sustained '+GR' mode is always and instantaneously present after nucleation. In contrast, the self-sustained '$-P_{SLOW}$' mode is often preceded by an unstable subshear phase. For smaller earthquakes this unstable phase may never transition to supershear and instead it may be arrested. This in turn would result in a total eastward rupture length, which is slightly shorter than the total western rupture length of the earthquake series.

4.06.5 Observing Crack-Like, Pulse-Like, Wrinkle-Like and Mixed Rupture Modes in the Laboratory

The duration of slip at a point on an interface (or fault) in comparison to the duration of the rupture of the entire fault is a central issue to the modeling of earthquake rupture. There are two widely accepted approaches to the description of dynamic sliding (Rice, 2001). The most classic approach uses elastodynamic shear rupture (or crack) models in which the surfaces behind the leading edge of the rupture (rupture tip) continuously slide and interact through contact and friction. More recently, pulse-like models in which sliding occurs over a relatively small region have been introduced (Heaton, 1990). In these models, the sliding is confined to a finite length which is propagating at the rupture speed and is followed by interfacial locking.

In faults separating identical materials (homogeneous systems) the crack-like mode of rupture has often (but not exclusively) been generated in many numerical simulations of spontaneous rupture when a rate-independent friction law was implemented (Madariaga, 1976; Andrews, 1976, 1985; Das and Aki, 1977; Day, 1982; Ruppert and Yomogida, 1992; Harris and Day, 1993). It has been pointed out, however, that inversions of seismic data for slip histories, from well-recorded events, indicate that the duration

of slip at a point on the fault is one order of magnitude shorter than the total event duration. An attempt to explain this phenomenon has given rise to the concept of a pulse-like rupture mode (Heaton, 1982, 1990; Hartzell and Heaton, 1983; Liu and Helmberger, 1983; Mendoza and Hartzell, 1988, 1989). Some eyewitness accounts have also reported short slip durations during some earthquakes (Wallace, 1983; Pelton *et al.*, 1984).

The concept of a pulse-like rupture went against a widely accepted view of how seismic rupture occurs. Its introduction was followed by various efforts to illuminate the physics leading to this process through analytical and numerical investigations. Different mechanisms for 'self-healing' pulse generation along faults in homogeneous systems have been proposed (Heaton, 1990). One postulation is that if the fault strength is low immediately behind the rupture front and is increased rapidly at a finite distance, then slip might be restricted to a short, narrow propagating area (Brune, 1976). Recent theoretical and numerical investigations show that a strong velocity-weakening friction law model could indeed allow for pulse-like behavior of rupture under certain conditions (Zheng and Rice, 1998). However, simulations utilizing velocity weakening have sometimes resulted in crack-like ruptures and sometimes in 'self-healing' pulses (e.g., Cochard and Madariaga, 1994, 1996; Perrin *et al.*, 1995; Beeler and Tullis, 1996; Ben-Zion and Rice, 1997; Lapusta *et al.*, 2000; Cochard and Rice, 2000; Nielsen *et al.*, 2000; Coker *et al.*, 2005). Friction laws along interfaces between two identical elastic solids have to include laboratory-based rate and state evolution features and they must not exhibit illposedness or paradoxical features (Cochard and Rice, 2000; Ranjith and Rice, 2001). It has been proved that generalized rate and state friction laws are appropriate candidates for homogeneous fault systems (Cochard and Rice, 2000; Ranjith and Rice, 2001; Rice *et al.*, 2001; Zheng and Rice, 1998). Within the frame of rate and state friction laws, the following three requirements have to be fulfilled for rupture to occur as a 'self-healing' pulse (Zheng and Rice, 1998; Rice, 2001). One requirement is that the friction law must include strengthening with time on slipped portions of the interface that are momentarily in stationary contact (Perrin *et al.*, 1995). Another is that the velocity weakening at high slip rates must be much greater than that associated with the weak logarithmic dependence observed in the laboratory during low-velocity sliding experiments. This is often termed 'enhanced' velocity weakening. Lastly, the third requirement is that the overall driving stress has to be lower than a certain value, but high enough to allow for self-sustained pulse propagation (Zheng and Rice, 1998).

While strong velocity weakening is one mechanism which explains the onset of short duration slip-pulses along interfaces (faults) which separate similar materials, it is important to note that other mechanisms exist as well. One involves geometric confinement of the rupture domain by unbreakable regions (barrier model). That is, sliding consists of a number of crack-like ruptures of short duration on a small rupture area that are separated by locked regions (Aki, 1979; Papageorgiou and Aki, 1983a, 1983b). In one implementation of this scenario, a pulse-like rupture behavior was found in a 3-D geometry when the rupture process was confined within a long but narrow region by unbreakable barriers (Day, 1982). It was observed that the rupture starts in a classic crack-like mode and it propagates in all directions. After some time, arresting waves arrive from the boundaries and they effectively relock the fault behind the rupture front, resulting in two slip pulses. Alternatively, a rupture nucleates and propagates bilaterally, but may arrest suddenly at a strong barrier at one end. Following its arrest, the reflected waves from the barrier spread back and heal the rupture surface. The combination of the still-propagating end of the rupture with the healing reflected wave forms a moving pulse-like configuration (Johnson, 1990).

In general, narrow slip pulses can be generated during dynamic sliding along interfaces by strongly velocity-weakening friction on a homogeneous system, by strong fault zone heterogeneities (Rice, 2001) or by variations in normal stress along the rupture interface. All of the above conditions (velocity weakening, heterogeneities, and local normal stress variation) can produce slip pulses with low dynamic stress at the active part of the slip. An extensive discussion of the subject is presented by Nielsen and Madariaga (2003). Most relevant to the discussion here is a set of recent finite element calculations which have been carried out by Coker *et al.* (2005) for a configuration which is very similar to the experimental setup described in the next section. These simulations have shown that it is possible to generate a great variety of dynamic rupture modes (cracks, pulses, and pulse trains) propagating along an interface characterized by a rate-enhanced, rate- and state-dependent frictional law. The choice of rupture speed and rupture mode clearly depends on the load

intensity and rate but it is also very sensitive to the type of constitutive law employed, as well as the values of its parameters.

4.06.5.1 Experimental Investigation of Dynamic Rupture Modes

As discussed in Section 4.06.2.1, the majority of the existing experimental studies on sliding concentrate on processes occurring over large timescales (large compared to wave transit within the specimen) and are concerned with developing relationships between time-averaged friction data and various system parameters. Many dynamic friction laws motivated by using various experimental configurations and apparati lack the reproducibility of friction data (survey by Ibrahim, 1994) to be of definite value to the theorists. The results of these experiments are multibranched friction versus slip velocity curves, which even for the same material and the same experimental configuration depend not only on the properties of the frictional interface but also on the dynamic properties of the apparatus, such as mass, stiffness, and damping. This suggests that the friction data obtained in the course of stick-slip motions do not purely reflect the intrinsic properties of the surfaces in contact but are also greatly affected by several of the dynamic variables involved in each particular experimental setup. Another shortcoming of the experiments described above, in relation to their relevance to real earthquakes, is the achievable range of sliding speeds. Typically, natural earthquakes involve sliding velocities on the order of 1–10 $m\,s^{-1}$ while these experiments involve sliding in the range of 1 $\mu m\,s^{-1}$ to 1 $mm\,s^{-1}$. An exception to this rule are the experiments of Prakash and Clifton (1993) and of Prakash (1998), who employed a plate-impact pressure-shear friction configuration to investigate the dynamic sliding response of an incoherent interface in the microsecond timescale (sliding speeds on the order of 10 $m\,s^{-1}$). The experimental results, deduced from the response to step changes imposed on the normal pressure at the frictional interface, reinforce the importance of including frictional memory in the development of the rate-dependent state variable friction models. However, this setup is essentially 1-D. As such it is not able to provide information about the detailed stress field developed along the length of the interface nor to shed light on the nature (e.g., uniformity or lack of) of the rupture process during dynamic sliding. As in many cases described in Section 4.06.2.1.1, the implicit assumption crucial to the interpretation of the data is that both stress and sliding processes remain spatially uniform. Another exception is the very recent work by Tullis and his co-workers (Di Toro *et al.*, 2004; Tullis and Goldsby, 2003) who have approached seismic slip rates in experiments involving crustal and quartz rock. They have investigated the phenomenon of flash heating as the main mechanism resulting in severe velocity weakening. They have shown that the phenomenon of flash heating (Rice, 1999) can be experimentally linked to the observed behavior of drastic velocity weakening observed in their experiments. Finally, the scarcity of fast sliding velocity data in the open literature has to be emphasized. We believe that this is an essential data set which could be proved as the ultimate arbitrator for the validation of various dynamic sliding models and friction laws.

Rubinstein *et al.* (2004) conducted experiments to investigate the onset of dynamic sliding. They considered two blocks of PMMA separated by a rough interface and subjected to vertical static compression and to a gradually increased horizontal driving force. During the sliding initiation, they recorded the relative change of the net local contact area as a function of time by measuring the changes of the light intensity transmitted across the interface. The transmitted light was imaged by a camera. Because of the low framing rates of their recording apparatus, their experimental configuration was particularly suited to the study of 'slow' phenomena, with respect to the characteristic wave speeds of the material. They discovered slow waves propagating at speeds between 40 $m\,s^{-1}$ and 80 $m\,s^{-1}$, which is one order of magnitude slower than the Rayleigh wave speed of PMMA. These detachment waves, which are perhaps related to Schalamach waves, because of their low propagation speeds (Schallamach, 1971), are very important since they give rise to a significant amount of slip. There are some doubts, however, to whether the characteristics of these waves are determined only by the properties of the surfaces in contact and by the loading. Because of their low rupture speeds, it is also likely that the dynamic properties of the entire specimen may play an important role in the formation of these waves. If the above is true, then the slow detachment waves, though very important to sliding of finite bodies, may be less relevant to natural earthquakes. Another very interesting question, which also remains to be answered, is whether these 'detachment' waves generate opening at the interface, thus justifying their name.

As it is clear from the experiments presented both in Section 4.06.2.1.1 and above, most of the experimental methods employed up to this point lack the spatial or/and the temporal resolution to address some important questions on dynamic sliding related to the sliding initiation process, to the duration of sliding (or sliding mode), and to the changes in the stress field distribution along the interface. This is particularly true in cases where sliding speeds on the order of $1\,\mathrm{m\,s^{-1}}$, or above, are involved. To remedy this situation, the physical plausibility of generating pulse-like and crack-like rupture modes along 'incoherent' (frictional) interfaces in homogeneous systems has been investigated by Lykotrafitis *et al.* (2006a) in a number of well-controlled experiments. The sliding speeds were on the order of $1\,\mathrm{m\,s^{-1}}$. This sliding speed is of the same order of magnitude as the one expected in most 'large' natural earthquake rupture events. Dynamic photoelasticity combined with a new laser interferometry-based technique has provided direct physical evidence of the rupture mode type, the exact point of rupture initiation, the sliding velocity history at a point on the interface, and the rupture propagation speed. A summary of these experiments and a discussion of the conditions leading to various modes of ruptures is given in the following sections.

4.06.5.1.1 Specimen configuration and loading

Two Homalite-100 plates, subjected to a uniform compressive stress, were frictionally held along the interface. The top plate was also subjected to an asymmetric impact shear loading (**Figure 19**). Each of the plates was 139.7 mm long, 76.2 mm wide, and 9.525 mm thick. The average roughness of the surfaces in contact was approximately $R_a = 400$ nm. All

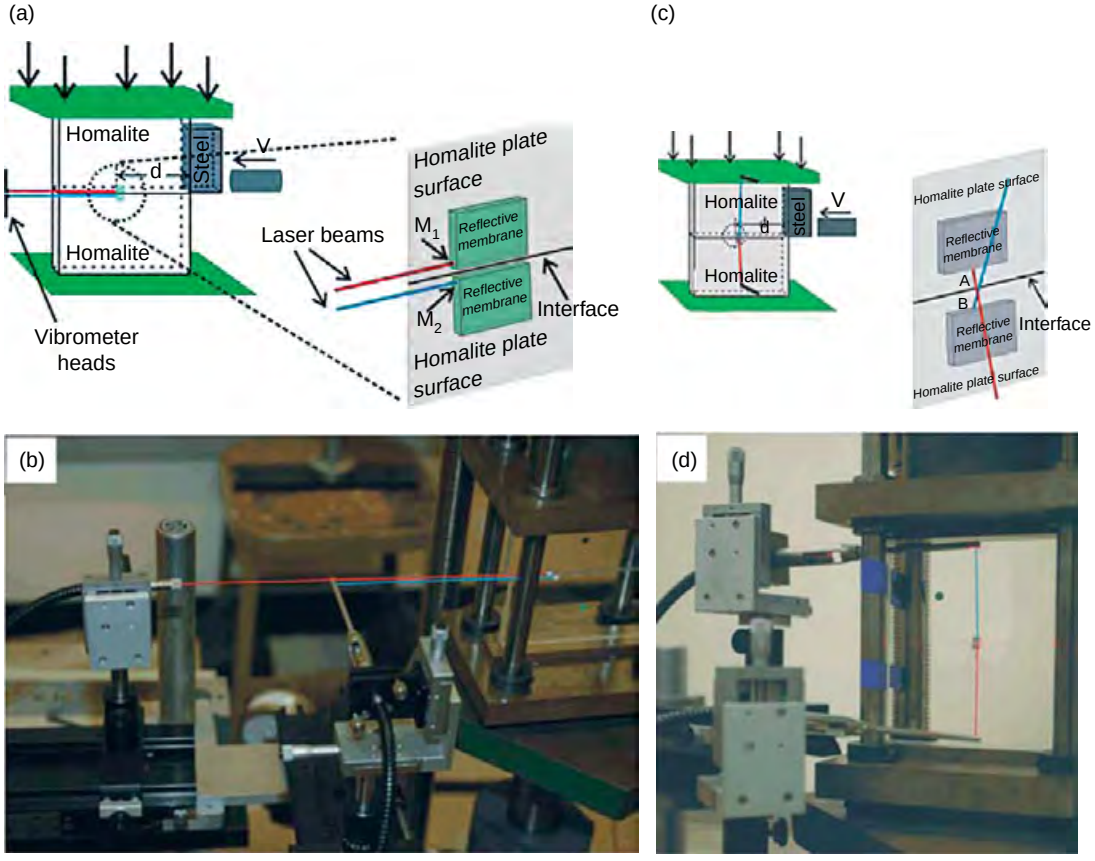


Figure 19 (a) Schematic illustration of the experimental configuration for the sliding velocity measurement. The area inside the dotted circle is shown magnified. Points M_1 and M_2 were at the same horizontal distance d from the impact side of the Homalite plate. (b) Photography of the actual setup displaying the arrangement of the velocimeters' heads for the measurement of the sliding velocity. (c) Schematic illustration of the experimental configuration for the relative vertical velocity measurement. The area inside the dotted line is shown magnified. Points M_1 and M_2 are at the same horizontal distance from the impact side of the Homalite plate. (d) Photography of the actual setup displaying the arrangement of the velocimeters' heads.

the experiments were executed at the same external confining stress of $P = 10$ MPa, applied by a calibrated press. The asymmetric impact loading was imposed via a cylindrical steel projectile of diameter 25.4 mm and length 50.8 mm, fired using a gas gun. The impact speeds ranged from 9 m s^{-1} to 72 m s^{-1}. A steel buffer 73 mm high, 25.4 mm long, and 9.525 mm thick was attached to the impact side of the upper plate to prevent shattering by direct impact and to induce a more or less planar loading wave.

4.06.5.1.2 Using particle velocimetry to measure slip and opening histories

The time evolution of the dynamic stress field in the bulk was recorded by high-speed digital photography in conjunction with dynamic photoelasticity (see Section 4.06.2.1). In addition, a new technique based on laser interferometry was adapted to locally measure the horizontal and vertical components of the in-plane particle velocity at various points above and below the sliding interface, thus allowing for the measurement of slip and opening velocity histories. The combination of the full-field technique of photoelasticity with the local technique of velocimetry is proved to be a very powerful tool in the study of dynamic sliding.

The configuration employed in the measurement of the slip velocity is as follows. A pair of independent fiber-optic velocimeters was used to measure the horizontal particle velocities at two adjacent points M_1 and M_2 across the interface. A schematic illustration of the experimental setup is shown in **Figure 19(a)**, whereas a photograph of the setup is shown in **Figure 19(b)**. The vertical distance of each point from the interface was less than 250 μm. Both points had the same horizontal distance from the impact side of the Homalite plates. By subtracting the velocity of the point below the interface (M_2) from that of the point above the interface (M_1), the relative horizontal velocity history was obtained. The velocimeter consists of a modified Mach–Zehnder interferometer and a velocity decoder. The decoder was set to a full range scale of ± 10 m s^{-1} with a maximum frequency of 1.5 MHz and a maximum acceleration of 10^7 g. The beam spot size was approximately 70 μm, whereas the error of the velocity measurements was 1%. The technique and the corresponding experimental setup are presented in detail in Lykotrafitis *et al.* (2006b).

Following the same strategy, a pair of independent velocimeters was employed to measure the vertical in-plane components of the velocities of two adjacent points M_1 and M_2, located across the interface. The relative vertical velocity was obtained by adding algebraically the corresponding velocities. The arrangement of the velocimeters is shown schematically in **Figure 19(c)**, whereas a photograph of the actual setup is shown in **Figure 19(d)**. These two types of measurements enabled them to record various modes of rupture propagating along the frictional interface.

4.06.5.2 Visualizing Pulse-Like and Crack-Like Ruptures in the Laboratory

The specimen was subjected to a uniform confining pressure of 10 MPa, whereas the impact speed was $V = 19$ m s^{-1}. An instantaneous isochromatic fringe pattern is shown in **Figure 20(a)**. An eye-like fringe structure is observed traveling, from right to left, behind the longitudinal wave front. The rupture tip A followed this fringe structure at a supershear speed of $1.36C_S$. Consequently, two Mach lines forming a shear Mach cone emanated from the rupture tip. The tip can be located by tracing the Mach lines to the interface. The rupture-tip speed was found to be constant.

The time evolution of the horizontal relative velocity of two adjacent points M_2 and M_1, belonging to the upper and lower plate respectively at a horizontal distance of 70 mm from the impact side of the Homalite plate, is displayed in **Figure 20(b)**. If the entire time history of photoelastic frames is taken into consideration in relation to the velocimeter time record, then this record can be superimposed on the photographs (see red line trace in **Figure 20(a)**) by converting time to spatial distance.

When the longitudinal wave front arrived at the measurement positions M_1 and M_2 of **Figure 20(a)**, where the pair of the interferometers was pointed, the velocities of both points started to increase. However, the relative horizontal velocity was zero for the next few microseconds and it remained very low for a time interval of approximately 13 μs. A numerical integration of the relative velocity with respect to time from 0 to 13 μs resulted in an accumulated relative horizontal displacement of 2 μm between points M_1 and M_2, which can easily be identified to elastic shear deformation prior to sliding. Indeed, the actual sliding started at approximately 13 μs when the rupture tip (point A as identified by the photoelastic image) arrived at the measurement position and the relative velocity increased sharply. The correlation between the two

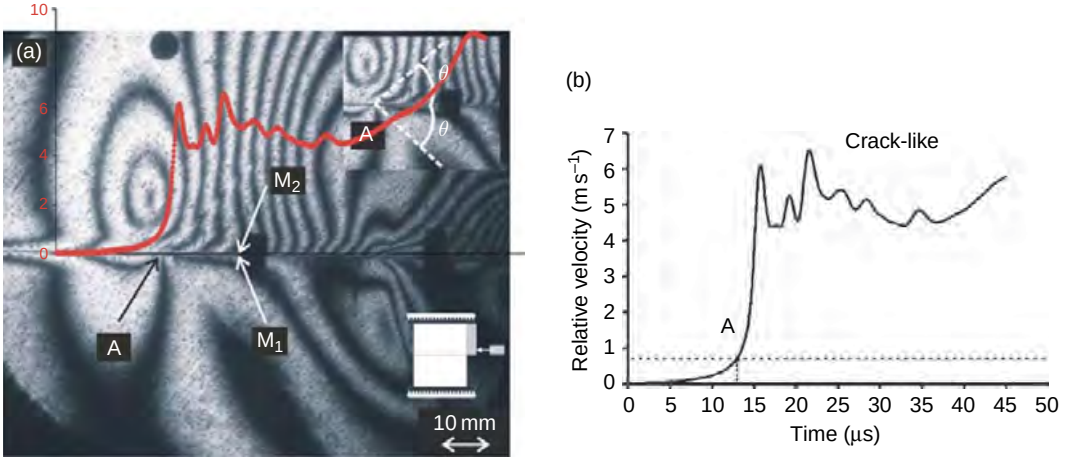


Figure 20 (a) Isochromatic fringe pattern generated during an experiment for which the impact speed was 19 m s^{-1} and the external compression was 10 MPa. The rupture tip is at the fringe concentration point A. The insets highlight the location of the Mach lines emanating from the rupture tip and the specimen configuration. The slip profile is superimposed (red line). (b) Relative velocity history of points M_1 and M_2 located at a distance of 70 mm from the impact side of the Homalite plates. The rupture commenced when the rupture tip A reached the velocity measurement position.

measurements has allowed us to thus establish an estimate of elastic displacement (~2 μm) preceding the sliding event for a confining pressure of 10 MPa. After it reached its maximum value of approximately 6 m s^{-1}, the relative velocity decreased and then it fluctuated, but never dropped below 4 m s^{-1} during the recording time. As is evident from **Figure 20(b)**, the sliding was continuous and thus we can safely say that rupture occurred in a classic crack-like mode.

As was already noted, the speed of the rupture tip was substantially higher than the shear wave speed of Homalite, and therefore the situation is similar to intersonic shear rupture propagation observed to occur along both 'coherent' (Rosakis *et al.*, 1999, 2000; Needleman, 1999; Coker and Rosakis, 2001; Samudrala *et al.*, 2002a, 2002b) and 'incoherent' (Xia *et al.*, 2004) interfaces separating identical monolithic solids. In the former experiments, the 'coherent' interfaces were bonded and they featured intrinsic strength and toughness in the absence of confining pressure. Unlike the present study, the resulting modes were always crack-like and the rupture speeds were not constant throughout the event.

In contrast to these early shear crack growth experiments, the present work involves incoherent or frictional interfaces and static far-field compressive loading. Here, the frictional resistance to sliding depends on the normal stress through the friction law. The normal stress, however, is a superposition of the static externally imposed pressure and a dynamic (inertial) compression generated by the impact loading as follows. The P-wave produced by the impact loading creates a primarily horizontal compressive stress in the upper plate close to the interface, and due to the Poisson effect it also creates compression in the direction vertical to the interface. As the sliding proceeds, the vertical stress to the rupture interface changes, and thus the friction changes as well. The change in friction, however, affects the evolution of sliding. Thus, we infer that sliding is dependent on impact loading not only through the horizontal compression, which is the driving force for sliding, but also through the vertical compression which affects the resistance to sliding. Because of that dependence, essential changes in the rupture process are to be expected as the impact speed is decreased.

In order to investigate the above line of reasoning, the impact speed was reduced to 15 m s^{-1}. **Figure 21(a)** shows an instantaneous isochromatic fringe pattern obtained at the above impact speed. The corresponding relative horizontal velocity history, measured at 30 mm from the impact side of the Homalite plate, is shown in **Figure 21(b)**. By combining the horizontal relative velocity measurement with the recorded photoelastic frames, we can identify two rupture tips A_1 and A_2, which are fringe concentration points and are propagating along the rupture interface at speeds of $1.09C_S$ and $0.98C_S$, respectively. The deformation at the velocity measurement position remains elastic until approximately 18 μs when the rupture tip A_1 arrived there

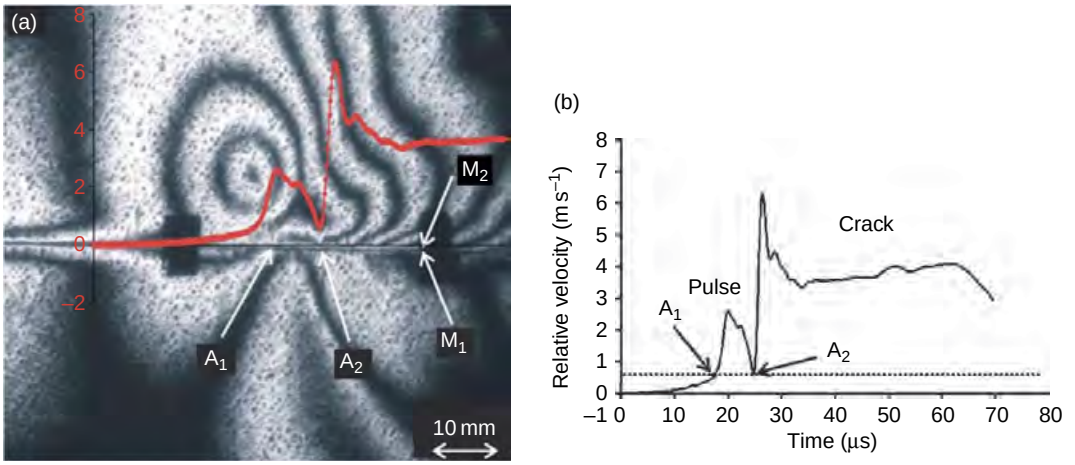


Figure 21 (a) Isochromatic fringe pattern generated during an experiment for which the impact speed was 15 m s^{-1} and the external compression was 10 MPa. The fringe concentration points A_1 and A_2 are the rupture tip and the rear edge respectively of the pulse-like rupture mode. The crack-like mode initiated at A_2 right after the pulse. The slip profile is superimposed (red line). (b) Relative velocity history of points M_1 and M_2 located at a distance of 30 mm from the impact side of the Homalite plates. The rupture commenced when the rupture tip A_1 reached the velocity measurement position and a pulse A_1A_2 was formed. The crack-like rupture mode initiated at A_2 right behind the second pulse.

and sliding commenced. As in the previous case, the commencement of slip corresponds to an accumulated relative horizontal displacement of 2 μm. Subsequently, the horizontal relative velocity increased rapidly from 0.6 m s^{-1} to a local maximum of 2.5 m s^{-1}. After 5 μs, the relative velocity decreased abruptly back to 0.6 m s^{-1} at point A_2 (**Figure 21(b)**). The slip ceased since the relative velocity was very low, allowing surface asperities to re-establish contact and be deformed elastically. The above observations show that the stable fringe structure (A_1A_2) represents a 'self-healing' slip pulse of approximately 7 μs in duration. Directly after the pulse, the relative velocity increased rapidly again to 6.4 m s^{-1} and retained its large value of approximately 4 m s^{-1} for a long period of time, of about 40 μs. This suggests that the initial rupture of the pulse-like mode was immediately followed by a second rupture of the crack-like mode. Thus, as it has been anticipated, the experimental results presented up to this point indicate that the rupture process is very sensitive to impact speed. Indeed, as the projectile speed was decreased while keeping the external confining stress constant, the rupture-tip speed was also decreased. Finally, the rupture mode changed from a crack-like mode to a mixed mode where a single 'self-healing' slip pulse was followed by a crack.

By further reducing the impact speed to 10 m s^{-1}, the rupture mode became purely pulse-like. The horizontal relative velocity was measured at a

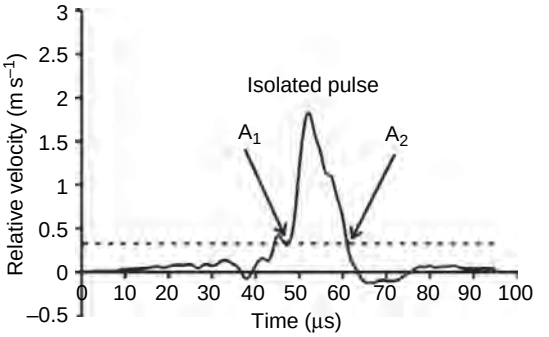


Figure 22 Relative velocity history of points M_1 and M_2 located at a distance of 70 mm from the impact side of the Homalite plates. The rupture commenced when the rupture tip A_1 reached the velocity measurement position and an isolated pulse A_1A_2 was formed.

distance of 70 mm from the impact side of the Homalite plate and its evolution over time is shown in **Figure 22**. The rupture started at A_1 and propagated at sub-Rayleigh speed of $0.76C_S$, whereas after 15 μs the sliding ceased at A_2. The duration of sliding was very short compared to the approximately 100 μs duration of the impact event, and thus we infer that an isolated 'self-healing' pulse was formed. Such a case clearly indicates that a pulse-like mode of rupture can definitely occur under appropriate conditions. Indeed this is the first time that such a dynamic pulse has been clearly seen in a controlled laboratory setting. The fact that 'self-healing' pulses

were obtained as the impact speed was decreased below certain point is consistent with the theoretical results of Zheng and Rice (1998). According to their predictions, and as discussed in Section 4.06.3.1, one of the necessary requirements of self-healing pulse formation is that the overall driving stress could be lower than a threshold value.

The combined use of classic dynamic photoelasticity with velocimetry allowed us to fully characterize the frictional sliding process. However, the maximum particle velocity which can be measured by the velocimeter is $10\,\mathrm{m\,s^{-1}}$. In order to comply with the above limit, low impact speeds were only used up to this point. For higher impact speeds, only dynamic photoelasticity in conjuction with high-speed photography was used as diagnostic tools.

In **Figure 23**, an instantaneous isochromatic fringe pattern is shown at a specific time for the same compressive load of 10 MPa as in the previous experiments and at much higher impact speed of $42\,\mathrm{m\,s^{-1}}$. The rupture propagated intersonically at approximately $1948\,\mathrm{m\,s^{-1}} = 1.56C_S = 0.75C_P$ giving rise to a shear Mach cone originating from the rupture tip B_1. In addition, a second Mach line which originated from point B_2 and was nonparallel to the first one was observed behind the rupture point (**Figure 23(a)**). The Mach line was at a shallower slope corresponding to a supershear (approximately sonic) propagation speed of $2514\,\mathrm{m\,s^{-1}} = 2.01C_S = 0.97C_P$. A more detailed view of the isochromatic fringe pattern in the neighbor of the rupture tip is shown in the inset of **Figure 23(a)**. The two shock waves are highlighted by dotted lines. Nonparallel shock lines imply a highly transient and unstable rupture process. Indeed, study of the entire set of the captured photoelastic pictures shows that the tip B_1 of the second Mach line gradually approached the end B_2 of the first Mach line. Finally, these two points merged as the second point caught up with the first point. The sliding continued at the lower speed and thus only one Mach line was observed in the next recorded frames (not shown here). **Figure 23(b)** shows the position histories of the first and second sliding tips for the case above. It is evident that the second sliding tip moved faster than the first sliding tip and at approximately 50 μs the pair of points coalesced. The existence of two Mach lines means that behind the onset of sliding (point B_1), there was a second point on the interface (point B_2), where the sliding speed again changed rapidly. Then, one could conjecture that the initial sliding which started at point B_1 stopped after a while and new sliding started at point B_2. In this way, an unstable sliding pulse was formed between the points B_1 and B_2 followed by a crack-like sliding which started at point B_2 in a similar manner to that shown in **Figure 21** though at higher sliding speeds. We finally note that behind the second Mach line a 'wrinkle' like pulse appeared at point C propagating at a speed of $0.92C_S$ which is close to the Rayleigh wave speed of Homalite. We extensively elaborate on the nature of this disturbance in Section 4.06.5.3.

An important comment on the frictional sliding experiments discussed above is that the rupture-tip

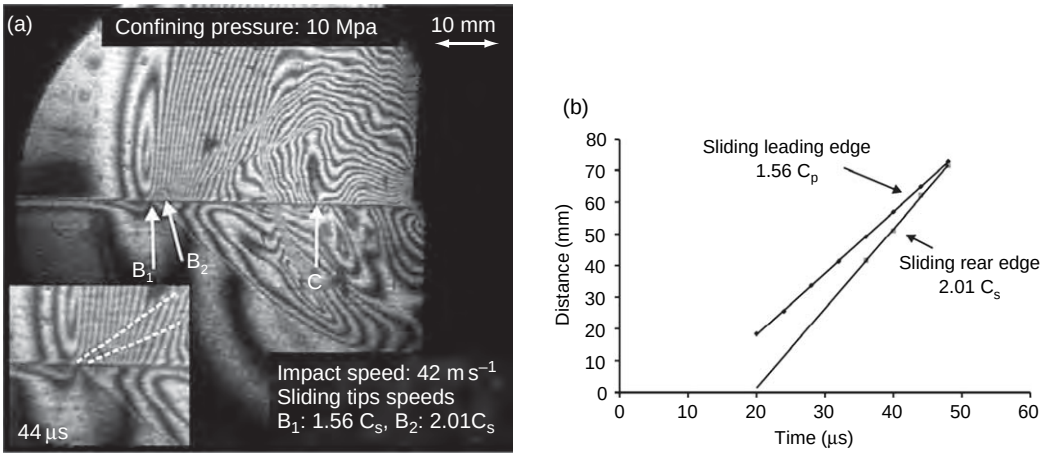


Figure 23 (a) Isochromatic fringes of sliding propagation at different time instances. Two sliding tips (B_1 and B_2) were propagating at different intersonic speeds. In the insets the dual and single Mach lines are marked. (b) Positions of the sliding leading edge and the second sliding tip as functions of time.

speeds were constants during the entire observation time. That is a general result and it holds true for all the experiments performed. This is a strong characteristic of frictional sliding and it agrees with theoretical results (Adams, 1998; Rice *et al.*, 2001; Ranjith and Rice, 2001) which predict constant and discrete propagation speeds for all the different disturbances and singularities along homogeneous interfaces subjected to uniform prestress.

In the last part of this section, the influence of the impact speed on the propagation speeds of the rupture tip is explored. **Figure 24** shows the variation of the propagation speed of the sliding front with the impact speed at a constant uniform confining stress of 10 MPa. The slowest achieved impact speed was $9\,m\,s^{-1}$. In this case and in other cases with impact speeds close to $10\,m\,s^{-1}$ the rupture-tip speed was sub-Rayleigh. At higher impact speeds the rupture-tip speed became supershear. It was observed that for impact speeds in the range of 20 to $40\,m\,s^{-1}$, the sliding speed was slightly above $\sqrt{2}C_S$. This is a special rupture speed (also discussed in Sections 4.06.3.1, 4.06.4.3, and 4.06.4.5) and it has been shown that it separates regions of unstable and stable intersonic shear crack growth (Samudrala *et al.*, 2002a, 2002b). When the impact speed increased, the rupture-tip speed increased toward the plane stress P-wave speed. It is worth mentioning that no sliding speed was observed in the interval between the Rayleigh wave speed and the shear wave speed of Homalite-100. This experimental observation agrees with theoretical predictions on steady-state shear crack propagation which exclude this speed interval based on energetic arguments (Freund, 1990; Broberg, 1999; Rosakis, 2002). It should also be noted that the experimental rupture-tip speed measurements presented here feature, as is proven below,

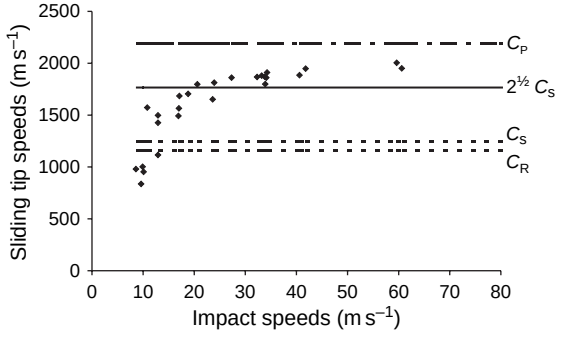


Figure 24 Variation of the sliding tip speed with the impact speed. The confining stress was 10 MPa.

high-enough resolution to obtain propagation speeds in the interval between C_R and C_S, if such speeds existed. It is finally noted that experiments performed at lower compression show that the speed of the rupture tip was influenced by the change in the static confining stress and it was supershear even at the lowest achieved impact speed.

4.06.5.3 Wrinkle-Like Opening Pulses along Interfaces in Homogeneous Systems

The possibility of generating wrinkle-like sliding pulses in incoherent frictionless contact between two dissimilar solids, when separation does not occur, was first investigated by Achenbach and Epstein (1967). These 'smooth contact Stonely waves' (also known as slip waves or generalized Rayleigh waves) are qualitatively similar to those of bonded contact (Stonely waves) and occur for a wider range of material combinations. Comninou and Dundurs (1977) found that self-sustained slip waves with interface separation (detachment waves or wrinkle-like opening slip pulses) can propagate along the interface of two similar or dissimilar solids which are pressed together. The constant propagation speed of these waves was found to be between the Rayleigh wave speed and the shear wave speed of the slowest wave speed material. For interfaces separating identical solids the propagation speed was between C_R and C_S. Weertman (1980) obtained a 2-D self-sustained wrinkle-like slip pulse propagating at the generalized Rayleigh wave speed along a bimaterial interface governed by Coulomb friction when the remote shear stress was less than the frictional strength of the interface. Finite-difference calculations of Andrews and Ben-Zion (1997) show the propagation of wrinkle-like opening pulses along a bimaterial interface governed by Coulomb friction. Particle displacement in a direction perpendicular to the fault is much greater in the slower material than in the faster material, resulting in a separation of the interface during the passage of the slip pulse. Anooshehpoor and Brune (1999) discovered such waves in rubber sliding experiments (using a bimaterial system consisting of two rubber blocks with different wave speeds). The above-mentioned detachment waves are radically different from the Schallamach waves (Schallamach, 1971) which propagate very slowly compared to the wave speeds of the solid.

Lykotrafitis and Rosakis (2006b) observed wrinkle-like opening pulses propagating along the interfaces of Homalite-steel bimaterial structures

subjected to impact shear loading. The propagation speeds of these wrinkle-like pulses of finite opening were constants and their values were always between the Rayleigh wave speed and the shear wave speed of Homalite, in accordance with the theoretical prediction of Comninou and Dundurs (1977). The wrinkle-like pulse in the bimaterial case generated a characteristic fringe pattern. Similar photoelastic fringe structures propagating behind the rupture tip along interfaces in homogeneous systems of Homalite were also observed by the same researchers (Lykotrafitis and Rosakis, 2006a). Prompted by this similarity, they decided to investigate whether wrinkle-like pulses were once more responsible for the generation of these fringe structures, in interfaces separating identical solids. Their results are summarized in this section. It is known that a system consisting of two identical half-planes and subjected to compression and to far-field shear loading cannot sustain a wrinkle-like pulse propagating along the interface. In the setup used in these experiments, however, the loading was not strictly shear and there is not any physical reason to exclude the possibility of generating wrinkle-like pulses.

In this section, photoelasticity and velocimetry is used to investigate the physical attainability of such wrinkle-like pulses along frictional interfaces separating identical solids. The equivalent issue for the case of Homalite-steel bimaterial system was investigated by Lykotrafitis and Rosakis (2006b) and Lykotrafitis *et al.* (2006b). **Figure 25** is used to illustrate the methodology used to identify the wrinkle pulses. The figure features an instantaneous isochromatic fringe pattern obtained at confining stress of 10 MPa and at an impact speed of 28 m s^{-1}. Point B is of particular interest. The rupture tip A is shown to propagate at a constant supershear speed of 1.49 C_S. A shear Mach cone is clearly visible in the photoelastic image. A fringe structure located at point B, in **Figure 25(a)**, propagates at a speed of 0.96 C_S. The time evolution of the relative vertical displacement of points M_1 and M_2, which were located 70 mm from the impact side of the Homalite plate, is displayed in **Figure 25(b)**. A simple 1-D calculation shows that the initial static compression of 10 MPa caused a negative relative displacement of approximately 1.3 μm. Negative relative displacement means that the two points approached each other under compression. The P-wave front arrives at about 2.5 μs after the triggering of the oscilloscope. Because of the Poisson effect, the horizontal dynamic (inertial) compression generated a vertical dynamic compression in addition to the static compression. A negative relative displacement was caused by this dynamic compression. The shear rupture point crossed the velocity measurement position at approximately 10 μs (marked A in **Figure 25(b)**) and it did not cause any visible change to the vertical components of the velocities of points M_1 and M_2. The relative vertical displacement becomes positive at approximately 31 μs. The photoelastic picture, captured at 30 μs, clearly shows that the fringe structure at B, which is the point under

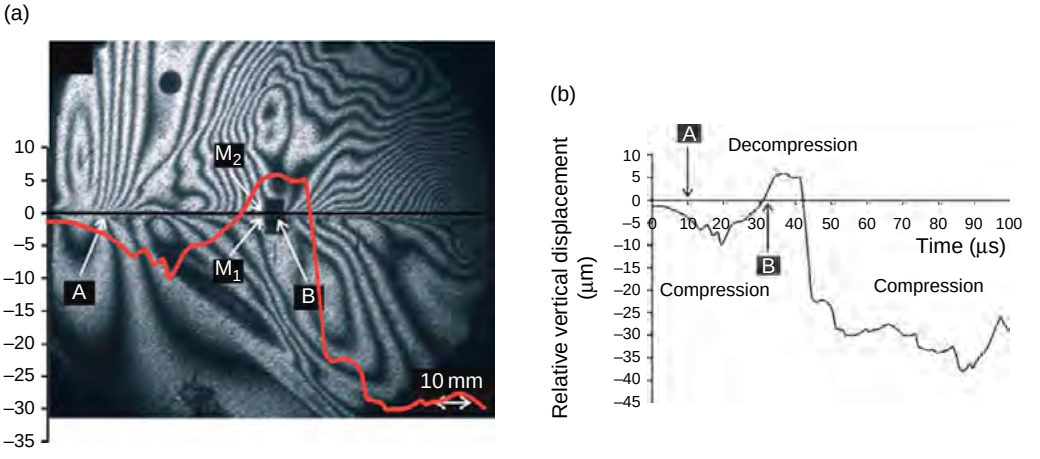


Figure 25 (a) Isochromatic fringe pattern generated during an experiment for which the impact speed was 28 m s^{-1} and the external compression was 10 MPa. The rupture tip is at the fringe concentration point A and the wrinkle-like pulse is at point B. The relative vertical displacement profile is superimposed (red line). (b) Relative vertical displacement history of points M_1 and M_2, located at a distance of 70 mm from the impact side of the Homalite plate. The rupture tip and the wrinkle-like disturbance crossed the velocity measurement position at approximately 10 μs and 32 μs after triggering, respectively.

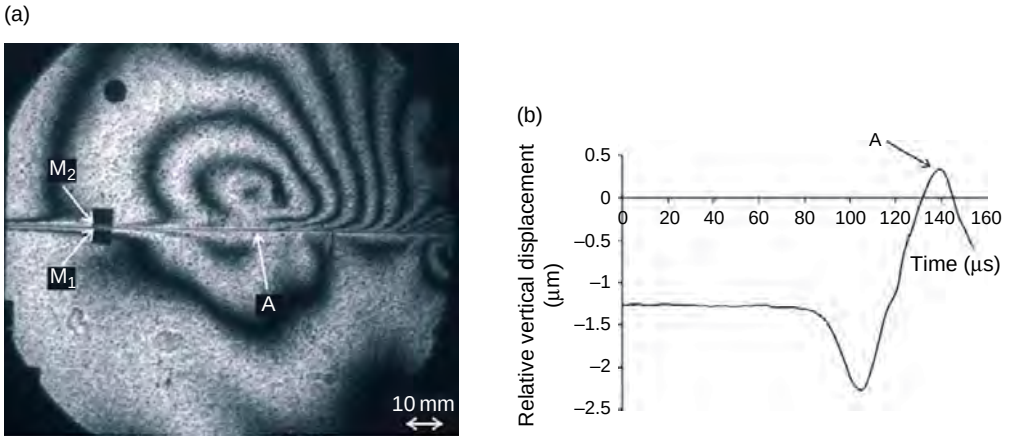


Figure 26 (a) Isochromatic fringe pattern generated during an experiment for which the impact speed was 9 m s^{-1} and the external compression was 10 MPa. The rupture tip is at the fringe concentration point A. No wrinkle-like pulse appeared. (b) Relative vertical displacement history of points M_1 and M_2 located at a distance of 100 mm from the impact side of the Homalite plate.

investigation, is very close to the measurement position. The study of the entire set of the 16 recorded photoelastic images in combination with the relative vertical displacement history sheds light on the distribution of the dynamic compression along the interface during sliding. The interface was locally under dynamic compression immediately after the arrival of the P-wave front until the arrival of the mentioned fringe formation at B. The entire area from the rupture point A to the location of the fringe structure at B was sliding under compression. The length of the area AB was estimated to be approximately 40 mm, whereas the entire length of the interface was 139.7 mm. The fringe structure at B caused a local opening (~5 μm) at the interface. At approximately 42 μs the relative vertical displacement became again negative and the compression increased abruptly. During the rest of the recording time the interface at the velocity measurement position was sliding under compression. The fringe structure at B can thus be clearly related to a wrinkle-like pulse propagating along the interface. As was stated earlier, its propagation speed was very close to 0.96 C_S which lies in the interval between the shear and the Rayleigh wave speeds of the material in agreement with the prediction of Comninou and Dundurs (1977).

Reviewing the results from the entire spectrum of experiments, it is verified that, at a confining stress of 10 MPa, the lowest impact speed which can generate a wrinkle-like pulse in the present setup is approximately 17 m s^{-1}. **Figure 26(a)** displays an isochromatic fringe pattern obtained at a confining stress of 10 MPa and at an impact speed of 9 m s^{-1}. It is clear that no fringe structure related to a wrinkle-like pulse appears. That is reflected by the relative vertical displacement history shown in **Figure 26(b)**, where the maximum positive value of the displacement was only about 0.3 μm. A comparison of the photoelastic image shown in **Figure 26(a)** with photoelastic images obtained at the same static compression of 10 MPa and at similar impact speeds during experiments where the horizontal particle velocities measurements were available, shows that the rupture started at point A.

The propagation speeds of wrinkle-like pulses at different impact speeds and at the same confining stress of 10 MPa are shown in **Figure 27**. The sliding speed was measured to be between the Rayleigh wave and the shear wave speed of Homalite-100. This result, in combination with the experimental results obtained via velocimetry, favors our conjecture that the interface disturbance was actually a wrinkle-like pulse. As has been already mentioned, the available theoretical and numerical analysis on the subject predicts the same speed range with this identified in **Figure 27**. It is also noted that both the prestress and the impact speed do not affect the propagation speed of the wrinkle-like pulses provided that such pulses could be generated.

4.06.5.4 Discussion

Finite element calculations of dynamic frictional sliding between deformable bodies have always been a very challenging task for the numerical analysts. As it was recently recognized (see Section

4.06.4.1), the main source of difficulty was the ill-posedness of the corresponding boundary and initial value problem, if a rich-enough friction law was not implemented. Coker *et al.* (2005) took advantage of the latest advances in the theory of the frictional sliding (Rice, 2001) and by using a rate-enhanced, rate and state friction law of Prakash–Clifton type, they were able to perform stable (grid-size independent) finite element calculations for a configuration very similar to the experiments described above. The use of a rate-enhanced Prakash–Clifton type of law was an essential element for the successful completion of the simulations because of the presence of fast changes of the local compression at the interface, caused by wave propagation, during dynamic sliding. As it is explained in Section 4.06.4.1, the Prakash–Clifton law is currently considered the only friction law which correctly describes the effect of fast changes in compression on the resulting frictional resistance. We note that for various impact speeds the numerical simulations exhibit a richer behavior than the experimental results and generate not only crack-like, pulse-like and mixed mode ruptures but also trains of pulses and pulses following a crack-like rupture. A representative result is shown in **Figure 28(a)**, where the distributions of the sliding speed ($\Delta\dot{u}_{slip}$) and the shear traction T_s along the interface at three times ($t = 32$, 38, 44 μs) are illustrated, for the case of 40 MPa static compression and of a 2 m s^{-1} impact speed. Contours of maximum shear stress, which correspond to isochromatic fringes in photoelasticity, at $t = 38$ μs are shown in **Figure 28(b)**. In this numerical experiment, a mixed mode of rupture, where a pulse was followed by a crack-like rupture, was obtained. This is reminiscent of the experimental case shown in **Figure 21**. We believe that the similarity between the numerical and the experimental results can be further improved by binding and eventually fine tuning the parameters of the friction law.

The experimental results presented in Section 4.06.5 elucidate the sliding process and provide conclusive evidence of the occurrence of various sliding rupture modes (crack-like, pulse-like, or mixed) propagating dynamically along incoherent interfaces. Of particular interest here is the experimental evidence of the formation of both supershear and sub-Rayleigh sliding pulses of the 'self-healing' type, leading the direct validation of predictions made on the basis of theoretical and numerical models of dynamic shear rupture. These pulses were found to propagate in the absence of interfacial opening. The experiments also provide hints of the dominant physical mechanisms

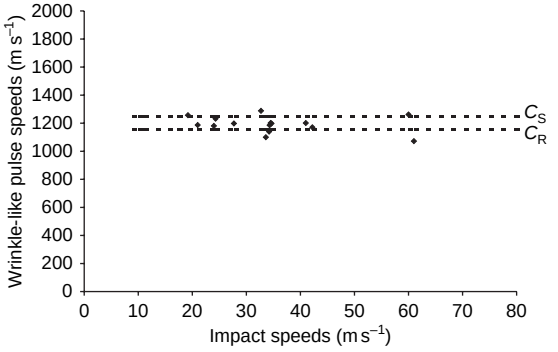


Figure 27 The wrinkle-like pulse speed remained between the Rayleigh wave speed and the shear wave speed of Homalite-100, independent of the impact speed.

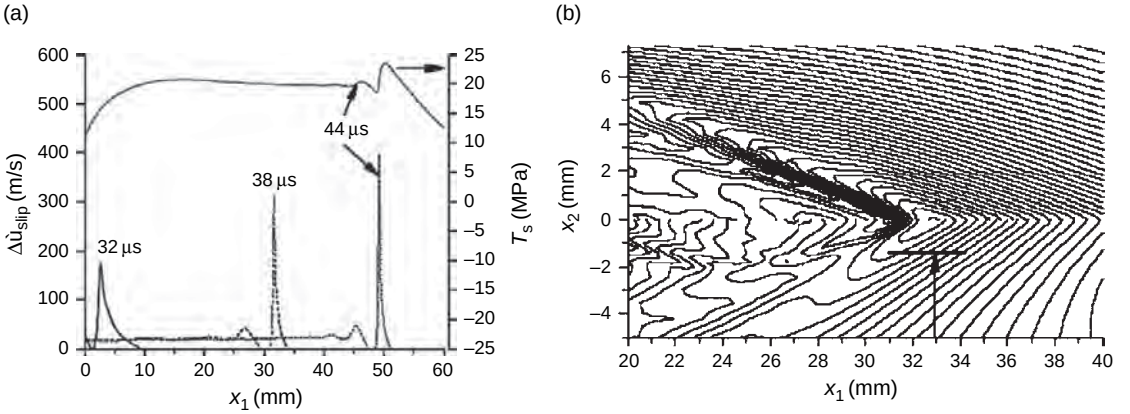


Figure 28 (a) Distributions of the sliding speed ($\Delta\dot{u}_{slip}$) at $t = 32$, 38, and 44 μs, and of the shear traction (T_S) at $t = 44$ μs along the interface for external compression of 40 MPa and impact speed of 2 m s^{-1}. (b) Contours of maximum shear stress (isochromatic fringe pattern) at $t = 38$ μs.

governing the choice of various rupture modes and their evolution. It is finally noted that 'wrinkle-like' pulses which feature finite opening in addition to sliding were discovered propagating along the interface at speeds between C_R and C_S for various loading conditions.

Acknowledgments

The authors gratefully acknowledge the support of NSF (Grant EAR 0207873), the US Department of Energy (Grant DE-FG52-06NA 26209) and the consistent support of the Office of Naval Research through (Grant N00014-03-1-0435) and MURI (Grant N000140610730) Dr. Y.D.S. Rajapakse, Program Manager.

References

Aagaard BT and Heaton TH (2004) Near-source ground motions from simulations of sustained intersonic and supersonic fault ruptures. *Bulletin of the Seismological Society of America* 94: 2064–2078.

Aagaard BT, Heaton TH, and Hall JF (2001) Dynamic earthquake ruptures in the presence of lithostatic normal stresses: Implications for friction models and heat production. *Bulletin of the Seismological Society of America* 91: 1765–1796.

Achenbach JD and Epstein HI (1967) Dynamic interaction of a layer and a half-space. *Journal of Engineering Mechanics* 5: 27–42.

Abraham FF and Gao HJ (2000) How fast can cracks propagate? *Physical Review Letters* 84: 3113–3116.

Adams GG (1995) Self-excited oscillations of two elastic half-spaces sliding with a constant coefficient of friction. *Journal of Applied Mechanics-Transactions of the ASME* 62(4): 867–872.

Adams GG (1998) Steady sliding of two elastic half-spaces with friction reduction due to interface stick-slip. *Journal of Applied Mechanics-Transactions of the ASME* 65(2): 470–475.

Adams GG (2001) An intersonic slip pulse at a frictional interface between dissimilar materials. *Journal of Applied Mechanics-Transactions of the ASME* 68(1): 81–86.

Aki K (1979) Characterization of barriers on an earthquake fault. *Journal of Geophysical Research* 84(6): 140–48.

Andrews DJ (1976) Rupture velocity of plane strain shear cracks. *Journal of Geophysical Research* 81(32): 5679–5687.

Andrews DJ (1985) Dynamic plane-strain shear rupture with a slip-weakening friction law calculated by a boundary integral method. *Bulletin of the Seismological Society of America* 75(1): 1–21.

Andrews DJ and Ben-Zion Y (1997) Wrinkle-like slip pulse on a fault between different materials. *Journal of Geophysical Research-Solid Earth* 102(B1): 553–571.

Andrews DJ and Harris RA (2005) The wrinkle-like slip pulse is not important in earthquake dynamics. *Geophysical Research Letters* 32(23): L23303 doi: 10.1029/2005GL023996.

Anooshehpoor A and Brune JN (1994) Frictional heat-generation and seismic radiation in a foam rubber model of earthquakes. *Pure and Applied Geophysics* 142(3–4): 735–747.

Anooshehpoor A and Brune JN (1999) Wrinkle-like Weertman pulse at the interface between two blocks of foam rubber with different velocities. *Geophysical Research Letters* 26(13): 2025–2028.

Archuleta RJ (1984) A faulting model for the 1979 Imperial-Valley Earthquake. *Journal of Geophysical Research* 89(NB6): 4559–4585.

Beeler NM and Tullis TE (1996) Self-healing slip pulses in dynamic rupture models due to velocity-dependent strength. *Bulletin of the Seismological Society of America* 86: 1130–1148.

Ben-Zion Y (2001) Dynamic ruptures in recent models of earthquake faults. *Journal of the Mechanics and Physics of Solids* 49(9): 2209–2244.

Ben-Zion Y (2006) Comment on ''Material contrast does not predict earthquake rupture propagation direction'' by R. A. Harris and S. M. Day. *Geophysical Research Letters* 33(13): L13310 doi: 10.1029/2005GL025652.

Ben-Zion Y and Huang YQ (2002) Dynamic rupture on an interface between a compliant fault zone layer and a stiffer surrounding solid. *Journal of Geophysical Research-Solid Earth* 107(B2): 2042 doi: 10.1029/2001JB000254.

Ben-Zion Y and Rice JR (1997) Dynamic simulations of slip on a smooth fault in an elastic solid. *Journal of Geophysical Research-Solid Earth* 102(B8): 17771–17784.

Bouchon M, Bouin MP, Karabulut H, Toksoz MN, Dietrich M, and Rosakis AJ (2001) How fast is rupture during an earthquake? New insights from the 1999 Turkey earthquakes. *Geophysical Research Letters* 28(14): 2723–2726.

Bouchon M and Vallee M (2003) Observation of long supershear rupture during the magnitude 8.1 Kunlunshan earthquake. *Science* 301(5634): 824–826.

Bowden FP and Tabor D (1986) *The friction and lubrication of solids*, 374 pp. Oxford: Clarenden Press.

Brace WF and Byerlee JD (1966) Stick-slip as a mechanism for earthquakes. *Science* 153(3739): 990–992.

Broberg KB (1999) *Cracks and fracture*, 725 pp. San Diego: Academic Press.

Brune JN (1973) Earthquake modeling by stick-slip along precut surfaces in stressed from rubber. *Bulletin of the Seismological Society of America* 63(6): 2105–2119.

Brune JN (1976) The physics of earthquake strong motion. In: Lomnitz C and Rosenbluth E (eds.) *Seismic Risk and Engineering Decisions*, pp. 141–171. New York: Elsevier.

Brune JN and Anooshehpoor A (1997) Frictional resistance of a fault zone with strong rotors. *Geophysical Research Letters* 24(16): 2071–2074.

Brune JN, Brown S, and Johnson PA (1993) Rupture mechanism and interface separation in foam rubber models of earthquakes - A possible solution to the heat-flow paradox and the paradox of large overthrusts. *Tectonophysics* 218(1–3): 59–67.

Burridge R (1973) Admissible speeds for plane-strain self-similar shear cracks with friction but lacking cohesion. *Geophysical Journal of the Royal Astronomical Society* 35(4): 439–455.

Burridge R, Conn G, and Freund LB (1979) Stability of a rapid mode-I shear crack with finite cohesive traction. *Journal of Geophysical Research* 84(NB5): 2210–2222.

Cochard A and Madariaga R (1994) Dynamic faulting under rate-dependent friction. *Pure and Applied Geophysics* 142(3–4): 419–445.

Cochard A and Madariaga R (1996) Complexity of seismicity due to highly rate-dependent friction. *Journal of Geophysical Research* 105: 25891–25907.

Cochard A and Rice JR (2000) Fault rupture between dissimilar materials: Ill-posedness, regularization, and slip-pulse response. *Journal of Geophysical Research-Solid Earth* 105(B11): 25891–25907.

Coker D, Lykotrafitis G, Needleman A, and Rosakis AJ (2005) Frictional sliding modes along an interface between identical elastic plates subject to shear impact loading. *Journal of the Mechanics and Physics of Solids* 53: 884–922.

Coker D and Rosakis AJ (2001) Experimental observations of intersonic crack growth in asymmetrically loaded unidirectional composite plates. *Philosophical Magazine A-Physics of Condensed Matter Structure Defects and Mechanical Properties* 81(3): 571–595.

Coker D, Rosakis AJ, and Needleman A (2003) Dynamic crack growth along a polymer composite-Homalite interface. *Journal of the Mechanics and Physics of Solids* 51(3): 425–460.

Comninou M and Dundurs J (1977) Elastic interface waves involving separation. *Journal of Applied Mechanics-Transactions of the ASME* 44(2): 222–226.

Dally JW and Riley WF (1991) *Experimental Stress Analysis*, 639 pp. New York: McGraw-Hill.

Das S and Aki K (1977) Numerical study of 2-dimensional spontaneous rupture propagation. *Geophysical Journal of the Royal Astronomical Society* 50(3): 643–668.

Day SM (1982) 3-dimensional simulation of spontaneous rupture - The effect of nonuniform prestress. *Bulletin of the Seismological Society of America* 72(6): 1881–1902.

Dieterich JH (1972) Time-dependent friction in rocks. *Journal of Geophysical Research* 77(20): 3690–3697.

Dieterich JH (1979) Modeling of rock friction. 1. Experimental results and constitutive equations. *Journal of Geophysical Research* 84(NB5): 2161–2168.

Dieterich JH (1992) Earthquake nucleation on faults with rate-dependent and state-dependent strength. *Tectonophysics* 211(1–4): 115–134.

Dieterich JH and Kilgore BD (1994) Direct observation of frictional contacts - New insights for state-dependent properties. *Pure and Applied Geophysics* 143(1–3): 283–302.

Di Toro G, Goldsby DL, and Tullis TE (2004) Friction falls toward zero in quartz rock as slip velocity approaching seismic rates. *Nature* 427: 436–439.

Dunham EM, Favreau P, and Carlson JM (2003) A supershear transition mechanism for cracks. *Science* 299(5612): 1557–1559.

Eberhart-Phillips D and Michael AJ (1998) Seismotectonics of the Loma Prieta, California, region determined from three-dimensional V-p, V-p/V-s, and seismicity. *Journal of Geophysical Research-Solid Earth* 103(B9): 21099–21120.

Ellsworth WL, Çelebi M, Evans JR, Jensen EG, Nyman DJ, and Spudich P (2004) Processing and Modeling of the Pump Station 10 Record from the November 3, 2002, Denali Fault, Alaska Earthquake. In: *Proceedings of the Eleventh International Conference of Soil Dynamics and Earthquake Engineering*. Berkeley, CA, pp. 563 Jan. 7–9.

Freund LB (1990) *Dynamic Fracture Mechanics*. Cambridge: Cambridge University Press.

Fukuyama E and Madariaga R (1998) Rupture dynamics of a planar fault in a 3D elastic medium: Rate- and slip-weakening friction. *Bulletin of the Seismological Society of America* 88(1): 1–17.

Gao HJ, Huang YG, and Abraham FF (2001) Continuum and atomistic studies of intersonic crack propagation. *Journal of the Mechanics and Physics of Solids* 49(9): 2113–2132.

Geubelle PH and Kubair DV (2001) Intersonic crack propagation in homogeneous media under shear- dominated loading: Numerical analysis. *Journal of the Mechanics and Physics of Solids* 49(3): 571–587.

Gu YJ and Wong TF (1994) Development of shear localization in simulated quartz gouge - Effect of cumulative slip and gouge particle-size. *Pure and Applied Geophysics* 143(1–3): 387–423.

Harris RA and Day SM (1993) Dynamics of fault interaction - Parallel strike-slip faults. *Journal of Geophysical Research-Solid Earth* 98(B3): 4461–4472.

Harris RA and Day SM (1997) Effects of a low-velocity zone on a dynamic rupture. *Bulletin of the Seismological Society of America* 87(5): 1267–1280.

Harris RA and Day SM (2005) Material contrast does not predict earthquake rupture propagation direction. *Geophysical Research Letters* 32(23): L23301 doi:10.1029/2005GL023941.

Hartzell SH and Heaton TH (1983) Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America* 73: 1553–1583.

Heaton TH (1982) The 1971 San Fernando earthquake; a double event?. *Bulletin of the Seismological Society of America* 72: 2037–2062.

Heaton TH (1990) Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors* 64(1): 1–20.

Ibrahim RA (1994) Friction-induced vibration, chatter, squeal, and chaos, Part I: Mechanics of contact and friction. *Applied Mechanics Review ASME* 47: 209–226.

Ida Y (1972) Cohesive force across tip of a longitudinal-shear crack and Griffiths specific surface-energy. *Journal of Geophysical Research* 77(20): 3796–3805.

Ide S and Takeo M (1997) Determination of constitutive relations of fault slip based on seismic wave analysis. *Journal of Geophysical Research-Solid Earth* 102(B12): 27379–27391.

Johnson E (1990) On the initiation of unidirectional slip. *Geophysical Journal International* 101: 125–132.

Johnson TL and Scholz CH (1976) Dynamic properties of stick-slip friction of rock. *Journal of Geophysical Research* 81(5): 881–888.

Kanamori H (1994) Mechanics of earthquakes. *Annual Review of Earth and Planetary Sciences* 22: 207–237.

Kubair DV, Geubelle PH, and Huang YGY (2003) Analysis of a rate-dependent cohesive model for dynamic crack propagation. *Engineering Fracture Mechanics* 70(5): 685–704.

Kubair DV, Geubelle PH, and Huang YY (2002) Intersonic crack propagation in homogeneous media under shear- dominated loading: theoretical analysis. *Journal of the Mechanics and Physics of Solids* 50(8): 1547–1564.

Lambros J and Rosakis AJ (1995) Dynamic decohesion of bimaterials - experimental-observations and failure criteria. *International Journal of solids and Structures* 32(17–18): 2677–2702.

Lapusta N, Rice JR, Ben-Zion Y, and Zheng G (2000) Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate- and state-dependent friction. *Journal of Geophysical Research* 105: 23,765–23,789.

Lapusta N and Rice JR (2003) Nucleation and early seismic propagation of small and large events in a crustal earthquake model. *Journal of Geophysical Research-Solid Earth* 108(B4): 2205 (doi:10.1029/2001JB000793).

Le Pichon X, Chamot-Rooke N, Rangin C, and Sengor AMC (2003) The North Anatolian fault in the Sea of Marmara. *Journal of Geophysical Research-Solid Earth* 108(B4) 2170–2179.

Lin AM, Fu BH, Guo JM, *et al.* (2002) Co-seismic strike-stip and rupture length produced by the 2001 M-s 8.1 Central Kunlun earthquake. *Science* 296(5575): 2015–2017.

Liu HL and Helmberger DV (1983) The near-source ground motion of the 6 August 1979 Coyote Lake, California, earthquake. *Bulletin of the Seismological Society of America* 73: 201–218.

Lu X, Lapusta N, and Rosakis AJ (2005) Testing friction laws by comparing simulation results with experiments of spontaneous dynamic rupture. *EOS Transactions of the American Geophysical Union* Fall Meet. Suppl., Abstract S32B–07.

Lykotrafitis G and Rosakis AJ (2006a) Sliding along frictionally held incoherent interfaces in homogeneous systems subjected to dynamic shear loading: A photoelastic study. *International Journal of Fracture* 140: 213–233.

Lykotrafitis G and Rosakis AJ (2006b) Dynamic sliding of frictionally held bimaterial interfaces subjected to impact shear loading. *Proceedings of the Royal Society of London*, *Series A* 462: 2997–3026.

Lykotrafitis G, Rosakis AJ and Ravichandran G (2006a) Self-healing pulse-like shear ruptures in the laboratory. *Science* 313: 1765–1768.

Lykotrafitis G, Rosakis AJ, and Ravichandran G (2006b) Particle velocimetry and photoelasticity applied to the study of dynamic sliding along frictionally-held bimaterial interfaces: Techniques and feasibility. *Experimental Mechanics* 46: 205–216.

Madariaga R (1976) Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America* 66: 639–666.

Madariaga R and Olsen KB (2000) Criticality of rupture dynamics in 3-D. *Pure and Applied Geophysics* 157(11–12): 1981–2001.

Magistrale H and Sanders C (1995) P-wave image of the peninsular ranges Batholith, Southern California. *Geophysical Research Letters* 22(18): 2549–2552.

Mendoza C and Hartzell SH (1988) Inversion for slip distribution using GDSN P waves: North Palm Springs, Borah peak, and Michoakan earthquakes. *Bulletin of the Seismological Society of America* 78: 1092–1111.

Mendoza C and Hartzell SH (1989) Slip distribution of 19 September 1985 Michoacan, Mexico, earthquake: Near-source and teleseismic constraints. *Bulletin of the Seismological Society of America* 79: 655–669.

McGuire JJ, Zhao L, and Jordan TH (2002) Predominance of unilateral rupture for a global catalog of large earthquakes. *Bulletin of the Seismological Society of America* 92(8): 3309–3317.

Needleman A (1999) An analysis of intersonic crack growth under shear loading. *ASME, Journal of Applied Mechanics* 66: 847–857.

Nielsen S and Madariaga R (2003) On the self-healing fracture mode. *Bulletin of the Seismological Society of America* 93: 2375–2388.

Nielsen SB, Carlson JM, and Olsen KB (2000) Influence of friction and fault geometry on earthquake rupture. *Journal of Geophysical Research-Solid Earth* 105(B3): 6069–6088.

Ohnaka M (2003) A constitutive scaling law and a unified comprehension for frictional slip failure, shear fracture of intact rock, and earthquake rupture. *Journal of Geophysical Research-Solid Earth* 108(B2): 2080 doi:10.1029/2000JB000123.

Ohnaka M and Shen LF (1999) Scaling of the shear rupture process from nucleation to dynamic propagation: Implications of geometric irregularity of the rupturing surfaces. *Journal of Geophysical Research-Solid Earth* 104(B1): 817–844.

Okubo PG and Dieterich JH (1984) Effects of Physical Fault Properties on Frictional Instabilities Produced on Simulated Faults. *Journal of Geophysical Research* 89(NB7): 5817–5827.

Olsen KB, Madariaga R, and Archuleta RJ (1997) Three-dimensional dynamic simulation of the 1992 Landers earthquake. *Science* 278(5339): 834–838.

Palmer AC and Rice JR (1973) Growth of slip surfaces in progressive failure of over-consolidated clay. *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* 332(1591): 527–548.

Papageorgiou A and Aki K (1983a) A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. I. Description of the model. *Bulletin of the Seismological Society of America* 73: 693–722.

Papageorgiou A and Aki K (1983b) A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. II. Applications of the model. *Bulletin of the Seismological Society of America* 73: 953–978.

Parsons T, Toda S, Stein RS, Barka A, and Dieterich JH (2000) Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation. *Science* 288(5466): 661–665.

Pelton JR, Meissner CW, and Smith KD (1984) Eyewitness account of normal surface faulting. *Bulletin of the Seismological Society of America* 74: 1083–1089.

Peltzer G, Crampe F, and King G (1999) Evidence of nonlinear elasticity of the crust from the Mw7.6 Manyi (Tibet) earthquake. *Science* 286(5438): 272–276.

Perrin G, Rice JR, and Zheng G (1995) Self-healing slip pulse on a frictional surface. *Journal of the Mechanics and Physics of Solids* 43: 1461–1495.

Prakash V and Clifton RJ (1993) Time resolved dynamic friction measurements in pressure-shear. In: Ramesh KT (ed.) *Experiemental Techniques in the Dynamics of Deformable Solids* Applied mechanics division, pp. 33–48. New York: ASME.

Prakash V (1998) Frictional response of sliding interfaces subjected to time varying normal pressures. *Journal of Tribology, ASME* 120: 97–102.

Ranjith K and Rice JR (1999) Stability of quasi-static slip in a single degree of freedom elastic system with rate and state dependent friction. *Journal of the Mechanics and Physics of Solids* 47(6): 1207–1218.

Ranjith K and Rice JR (2001) Slip dynamics at an interface between dissimilar materials. *Journal of the Mechanics and Physics of Solids* 49(2): 341–361.

Renardy M (1992) Ill-posedness at the boundary for elastic solids sliding under Coulomb-friction. *Journal of Elasticity* 27(3): 281–287.

Rice JR (1983) Constitutive relations for fault slip and earthquake instabilities. *Pure and Applied Geophysics* 121(3): 443–475.

Rice JR (1999) Flash heating at asperity contacts and rate-dependent friction. *EOS Transactions of the American Geophysical Union* 80: F681.

Rice JR (2001) New perspectives in crack and fault dynamics. In: Aref H and Phillips JW (eds.) *Mechanics for a New Millennium* (Proceedings of the 20th International Congress of Theoretical and Applied Mechanics, 22 Aug - Sept 2000), pp. 1–23. Chicago: Kluwer.

Rice JR, Lapusta N, and Ranjith K (2001) Rate and state dependent friction and the stability of sliding between elastically deformable solids. *Journal of the Mechanics and Physics of Solids* 49(9): 1865–1898.

Rosakis AJ (2002) Intersonic shear cracks and fault ruptures. *Advances in Physics* 51(4): 1189–1257.

Rosakis AJ, Samudrala O, and Coker D (1999) Cracks faster than the shear wave speed. *Science* 284(5418): 1337–1340.

Rosakis AJ, Samudrala O, and Coker D (2000) Intersonic shear crack growth along weak planes. *Materials Research Innovations* 3: 236–243.

Rosakis AJ, Samudrala O, Singh RP, and Shukla A (1998) Intersonic crack propagation in bimaterial systems. *Journal of the Mechanics and Physics of Solids* 46(10): 1789–1813.

Rubin AM and Gillard D (2000) Aftershock asymmetry/rupture directivity among central San Andreas fault

microearthquakes. *Journal of Geophysical Research-Solid Earth* 105(B8): 19095–19109.
Rubin AM and Ampuero J (2006) Aftershock asymmetry on a bimaterial interface. *Journal of Geophysical Research* (to appear).
Rubinstein S, Cohen G, and Fineberg J (2004) Detachment fronts and the onset of dynamic friction. *Nature* 430: 1005–1009.
Ruina A (1983) Slip instability and state variable friction laws. *Journal of Geophysical Research* 88(NB12): 359–370.
Ruppert SD and Yomogida KA (1992) Crack-like rupture model for the 19 September 1985 Michoacan, Mexico, earthquake. *Pure and Applied Geophysics* 138: 407–2.
Samudrala O, Huang Y, and Rosakis AJ (2002a) Subsonic and intersonic mode II crack propagation with a rate- dependent cohesive zone. *Journal of the Mechanics and Physics of Solids* 50(6): 1231–1268.
Samudrala O, Huang Y, and Rosakis AJ (2002b) Subsonic and intersonic shear rupture of weak planes with a velocity weakening cohesive zone. *Journal of Geophysical Research-Solid Earth* 107(B8): 10129/2001JB000460.
Schallamach A (1971) How does rubber slide? *Wear* 17: 301–312.
Scholz CH (2002) *The Mechanics of Earthquakes and Faulting*, 471 pp. Cambridge: Cambridge University Press.
Scholz C, Molnar P, and Johnson T (1972) Detailed studies of frictional sliding of granite and implications for earthquake mechanism. *Journal of Geophysical Research* 77(32): 6392–6406.
Shi ZQ and Ben-Zion Y (2006) Dynamic rupture on a bimaterial interface governed by slip weakening friction. *Geophysical Journal International* 165(2): 469–484.
Spudich P and Cranswick E (1984) Direct observation of rupture propagation during the 1979 Imperial Valley Earthquake using a short baseline accelerometer array. *Bulletin of the Seismological Society of America* 74(6): 2083–2114.
Stein RS, Barka AA, and Dieterich JH (1997) Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International* 128(3): 594–604.
Tullis TE and Goldsby DL (2003) Flash melting of crustal rocks at almost seismic rates. *EOS. Transactions of the American Geophysical Union* 84(46): Fall Meeting Suppl., Abstract S51B-05.
Thurber, Roecker S, Roberts K, Gold M, Powell L, and Rittger K (2003) Earthquake locations and three-dimensional fault zone structure along the creeping section of the San Andreas fault near Parkfield, CA: Preparing for SAFOD. *Geophysical Research Letters* 30(3): 1112 doi: 10.1029/2002GL016004.
Uenishi K and Rice JR (2003) Universal nucleation length for slip-weakening rupture instability under nonuniform fault loading. *Journal of Geophysical Research-Solid Earth* 108(B1): 2042 doi: 10.1029/2001JB001681.
Uenishi K, Rossmanith HP, and Scheidegger AE (1999) Rayleigh pulse-dynamic triggering of fault slip. *Bulletin of the Seismological Society of America* 89(5): 1296–1312.
Wallace RE (1983) Eyewitness account of surface fault during the earthquake of 28 October 1983 Borah peak, Idaho. *Bulletin of the Seismological Society of America* 74: 1091–1094.
Weertman J (1980) Unstable slippage across a fault that separates elastic media of different elastic-constants. *Journal of Geophysical Research* 85(NB3): 1455–1461.
Wu FT, Thomson KC, and Kuenzler H (1972) Stick-slip propagation velocity and seismic source mechanism. *Bulletin of the Seismological Society of America* 62(6): 1621–1628.
Xia KW, Chalivendra VB, and Rosakis AJ (2006) Observing ideal 'Self-similar' crack growth in experiments. *Engineering Fracture Mechanics* 73: 2748–2755.
Xia KW, Rosakis AJ, and Kanamori H (2004) Laboratory earthquakes: The sub-Rayleigh-to-supershear rupture transition. *Science* 303 (5665): 1859–1861.
Xia KW, Rosakis AJ, Kanamori H, and Rice JR (2005) Laboratory earthquakes along inhomogeneous faults: Directionality and supershear. *Science* 308(5722): 681–684.
Zheng G and Rice JR (1998) Conditions under which velocity-weakening friction allows a self-healing versus a crack-like mode of rupture. *Bulletin of the Seismological Society of America* 88: 1466–1483.
Zor E, Sandvol E, Gurbuz C, Turkelli N, Seber D, and Barazangi M (2006) The crustal structure of the East Anatolian plateau (Turkey) from receiver functions. *Geophysical Research Letters* 30(24): 8044, doi:10.1029/2003GL018192.

4.07 Slip Inversion

S. Ide, University of Tokyo, Tokyo, Japan

Nomenclature

$\mathbf{d}^0$	data vector
$\mathbf{d}^e$	model estimate vector
$\mathbf{e}$	constraint vector
$\mathbf{e}_j$	jth constrant vector
$\mathbf{m}$	model parameter vector
$p(\mathbf{d}^0\|\mathbf{m})$	liklihood function for given data vector and unknown parameter vector
$\mathbf{C}$	constraint matrix
$\mathbf{C}_j$	jth constrant matrix
D_c	slip-weakening distance
$\mathbf{E}_d$	data covariance matrix
$\mathbf{E}_m$	parameter covariance matrix
E_s	seismic energy (J)
$\mathbf{G}$	kernel matrix
G_c	fracture surface energy (J m^{-2})
$G_{ip}(\mathbf{x}, t; \xi, \tau)$	Green's function
$L(\sigma^2, \sigma_1^2)$	marginal likelihood, function of hyperparameters, $L(\sigma^2, \sigma_1^2) = \int p(\mathbf{d}^o\|\mathbf{m}) p(\mathbf{m}\|\sigma_1^2) d\mathbf{m}$
M_0	seismic moment (N m)
M_w	moment magnitude
N_c'	rank of constraint matrix $\mathbf{C}$
N_d	length of data vector
N_m	length of parameter vector
α_1^2	hyperparameter $\alpha_1^2 = \sigma^2/\sigma_1^2$
$\gamma(\mathbf{x})$	free surface energy, energy to make a new surface
σ^2	data variance
$\sigma_{ij}(\mathbf{x}, t)$	ij conponent of stress tensor at $(\mathbf{x}, t)$
σ_j^2	variance of jth constraint
$\Delta u_j(\xi, \tau)$	slip distribution in ith direction at (ξ, τ)

4.07.1 Introduction

Earthquake source is dynamic shear ruptures including fracture and frictional slip on fault planes. To further study the physics of earthquake rupture, we have to know what occured during an earthquake by resolving the spatial and temporal behavior of the rupture. The information of detailed rupture process is also useful for realistic simulations of strong ground motion from a complex source. The rupture process is usually represented by fault slip distribution using a parametric model, which is called a slip model, a heterogeneous slip model, or a finite fault model. Since slip models only describe the rupture history without specific reference to underlying rupture physics, they are also referred to as kinematic models. The method to objectively constrain a slip model by data, usually seismic and/or geodetic data, is slip inversion, which is the central topic of this chapter. The data are usually contaminated with background noise and our knowledge about earthquake locations and underground structure are not sufficient. To overcome these difficulties, various techniques of slip inversion have been developed with many applications during these decades. We review the history, typical formulation, application examples, and extension of slip inversion, paying attention to the limitation of the method.

First we review the history of the development of slip inversion. Modern seismology began in the 1960s after the establishment of a basic picture of seismic sources as shear slips on fault planes (*see* Chapter 4.02). Revealing the spatial distribution of the slip soon became a major topic of earthquake seismology. The first earthquake whose spatial extent is quantitatively discussed based on seismograms was the 1960 Chilean earthquake (Mw 9.5), the largest event in the twentieth century. The studies of surface waves using long-period seismograms suggest that this earthquake propagated 750–1000 km to the south at a rupture velocity a little slower than the S-wave velocity (Benioff *et al.*, 1961; Press *et al.*, 1961).

In the 1960s macroscopic seismic source models described by handful parameters were well studied. The most famous macroscopic model that includes simple rupture propagation is the Haskell's model (Haskell, 1964, 1969; Aki and Richards, 2002). In this model, a line dislocation of a constant amount and a constant duration (rise time) propagates on a rectangular plane unidirectionally at a constant velocity. The 1966 Parkfield earthquake (M 6.0) was an example that was successfully explained by a propagating dislocation (Aki, 1968). The Haskell's model and similar macroscopic models are used to explain observed records from many large earthquakes during the 1960s and 1970s. Based on the results for earthquakes larger than magnitude 6, Kanamori and Anderson (1975) derived scaling relations for macroscopic earthquake ruptures. In these scaling relations, macroscopic parameters such as the fault length, the fault width, and the average slip satisfy a geometrical similarity, indicating that the average stress drop is almost constant. This is a standard macroscopic image of earthquake source established in the 1970s.

However, from the very beginning of the seismic observation history, it has also been known that an earthquake is not a simple rupture but consists of several distinct shocks. Quantitative evidence of such complexity was clearly given from the observation of far-field body waves. Far-field records of some large earthquakes contain multiple pulses that are not explained by the effect of the underground structure. These events are called multiple shocks. One example is the 1976 Guatemala earthquake (Mw 7.5) studied by Kanamori and Stewart (1978) and Kikuchi and Kanamori (1982, 1991). This earthquake consists of 10 discrete ruptures that radiated pulse-like body waves. Although not all earthquakes are that complex, some degree of complexity is visible in any large earthquake.

From the late 1970s, seismologists started developing systematic procedures to analyze the complexity of earthquake source. Trifunac (1974) constructed and solved the first slip inversion problem that objectively determines seismic slip on an assumed fault plane for the 1971 San Fernando earthquake. He divided the assumed fault plane into a number of rectangular subfaults and determined the amount of slip on each of the subfaults using the least-squares method. Following the development of the inversion theory in other fields, such as statistics and information engineering, various systematic procedures have been developed to reveal heterogeneity of seismic source. In 1982, two important papers for inversion methods were published. One is the subevent deconvolution method proposed by Kikuchi and Kanamori (1982). They consider seismic sources as a sequence of spatially distributed point sources and identified the time and moment of these sources iteratively using waveform correlations of far-field body waves. This method and its extended version were applied for many earthquakes (Kikuchi and Kanamori, 1986, 1991, 1994). The other inversion method proposed by Olson and Apsel

(1982) determined spatial and temporal slip distribution on a fault plane using successive time windows propagating at a constant velocity. Hartzell and Heaton (1983) also proposed a similar method. These methods are referred to as the multi-time-window method and have been frequently applied for many earthquakes since the 1979 Imperial Valley earthquake (Mw 6.4). We will review this method more thoroughly in Section 4.07.2.

In the history of slip inversion, a small number of large earthquakes played significant roles to provide new kinds of data, to enable development of special treatments, and to increase our knowledge of earthquake physics. Very near-field records of the 1966 Parkfield earthquake (M 6.0), at 80 m from the surface fault trace, confirmed the image of seismic source as a propagating dislocation (Aki, 1968). The 1971 San Fernando earthquake (Mw 6.7) provided more than 250 near-field strong-motion records and enabled the first slip inversion from the data at five stations within the source area as described above (*see* Chapter 4.18; Trifunac, 1974). The 1979 Imperial Valley earthquake (Mw 6.4) is the first event whose slip models were published by different groups (Olson and Apsel, 1982; Hartzell and Heaton, 1983; Archuleta, 1984). In the 1990s, geodetic measurements using satellites became popular and these data are introduced into joint inversions as constraints of final slip. The 1992 Landers earthquake (Mw 7.2), the 1994 Northridge earthquake (Mw 6.7), and the 1995 Hyogo-Ken Nanbu (Kobe) earthquake (Mw 6.9) are well-studied events using both seismic and geodetic data. Serious damage by the Kobe earthquake stimulated the improvement of nationwide observational systems in Japan, the dense networks of high-sensitivity and strong-motion seismometers and global positioning system (GPS), which enabled detailed studies of smaller earthquakes. The 1999 Chi-Chi, Taiwan, earthquake (Mw 7.6) is so far the best recorded and best resolved earthquake owing to the dense strong-motion network in Taiwan. The first M9 earthquake in the digital seismogram era, the 2004 Sumatra earthquake (Mw 9.3) was also analyzed for slip distribution. A dense far-field seismometer network can reveal the whole rupture process of such a great earthquake even without near-field strong-motion records (Krüger and Ohrnberger, 2005; Ishii *et al.*, 2005).

In recent years, slip inversion has become a kind of routine analysis. Rapid solutions are published via the Internet from several groups immediately after large earthquakes, usually within one or a few days. Since there are some distributed computer programs of slip inversion, there is a possibility that it is applied without careful investigation. Sometimes, assumptions to get solutions are not explicity written in research papers. If there are enough data, the relative significance of the assumptions is low, but the data are often limited and the assumptions can be essential for the details of the solution.

In the following sections, we review general treatments and assumptions in slip inversion decomposing the problem into basic elements: preparation of data, model parametrization, and calculation of synthetic data from the model. All of these are essential to construct even a forward problem that just compares data and model prediction. In Section 4.07.2 we introduce various examples for each element. Section 4.07.2.2 reviews the availability of seismic waveform data and several geodetic measurements. Most popular seismic waveform data are local strong-motion records and global broadband records. In Section 4.07.2.3 we summarize the representation of slip distribution by a finite number of model parameters, which are the models to be determined by slip inversion and generally classified into linear or nonlinear types. Synthetic data are calculated using Green's functions, which are the displacements at stations due to an impulsive force and connect data and model parameters. As we will discuss in Section 4.07.2.4, they are theoretically calculated based on the knowledge of underground structure, or empirically modified from the records of small events. Section 4.07.3 explains how to solve the problem based on inversion theory. After a short review of an ordinary solution for the least-squares problem in Section 4.07.3.1, Bayesian modeling for underdetermined problems is reviewed with expressions in Section 4.07.3.2. While most slip inversion problems are solved in the time domain, sometimes information in the frequency domain is also important, although there have been only a few studies in the frequency domain. We will discuss these studies in Section 4.07.3.3. Section 4.07.4 shows examples of slip inversion, by comparing slip models of the Chi-Chi earthquake published by different groups. Section 4.07.5 reviews various derivative studies based on slip models, finding characteristics, dynamic implication and modeling, and scaling problem. Finally, Section 4.07.6 summarizes this chapter discussing the current problems and the future prospects of slip inversion.

There have been good review articles about slip inversion. The study of the 1986 North Palm Springs

earthquakes by Hartzell (1989) discussed merits and demerits for different settings, comparing teleseismic and strong motion data, linear and nonlinear models, and empirical and theoretical Green's functions. The developments till the 1980s are summarized by Kikuchi (1991) and Iwata (1991). Yoshida (1995) reviewed the problem paying special attention to classify the problems from a point source to a spatio-temporal slip distribution. Beroza (1995) described characteristics of analysis with far-field and near-field data. The review of strong-motion seismology by Anderson (2003) included a list of previous slip inversion studies using strong-motion records. From the viewpoint of earthquake physics, Kanamori (2004) mentioned characteristics of some results of slip inversion. There are also good textbooks and review articles for general inversion problems in geophysics (e.g., Menke, 1989; Matsu'ura, 1991; Tarantola, 2005).

4.07.2 Construction of Slip Inversion Problem

4.07.2.1 Outline

First, we introduce the mathematical formulation of slip inversion problems using seismic and/or geodetic data. An earthquake rupture is the spatial and temporal distribution of displacement discontinuity $\Delta\mathbf{u}(\mathbf{x}, t)$ across fault planes Σ. We consider only shear slip whose direction is perpendicular to the local normal vector of fault plane, $\boldsymbol{\nu}(\mathbf{x})$, so that $\Delta\mathbf{u}(\mathbf{x}, t)\cdot\boldsymbol{\nu}(\mathbf{x}) = 0$ everywhere. In the elastic medium, displacement at $(\mathbf{x}, t)$ due to a slip distribution $\Delta\mathbf{u}(\boldsymbol{\xi}, \tau)$ is written as (equation 10.1 of Aki and Richards, 2002)

$$u_i(\mathbf{x}, t) = \int_{-\infty}^{\infty} \mathrm{d}\tau \int\int_{\Sigma} \Delta u_j(\boldsymbol{\xi}, \tau) \times C_{jkpq} G_{ip,q}(\mathbf{x}, t; \boldsymbol{\xi}, \tau) \nu_k(\boldsymbol{\xi}) \mathrm{d}\sum(\boldsymbol{\xi}) \quad [1]$$

where C_{jkpq} is the elastic constant. G_{ip} is a Green tensor function that is ith component of displacement at $(\mathbf{x}, t)$ due to an impulsive force at $(\boldsymbol{\xi}, \tau)$ in pth direction, q represents the derivative in ξ_q direction. The summation convention is used for index j, k, p, and q. Equation [1] is the expression for seismic waves, but it is also applicable to geodetic measurements if we take $t \to \infty$. The constituents of this problem are data u_i, slip model Δu_j, and Green's functions G_{ip}, which we will discuss independently in the following subsections (*see* Chapter 4.02).

4.07.2.2 Data Preparation

Among various data available for slip inversion, seismic waveform data are essential to resolve temporal change of fault slip. Broadband seismograms recorded by global seismic networks are useful to resolve slip distribution for any earthquake larger than M7. Far-field P- and S-waves in the range of hypocentral angular distances from 30° to 100° are separated from other large phases and interpreted as a source time function with minor modification for the reflection and refraction due to local structure. Seismic waveform data from regional networks are useful to improve the resolution of the slip distribution for large earthquakes and enable slip inversion even for earthquakes of moderate size. Especially, strong-motion records are most important to get a fine image of rupture propagation for large earthquakes. When on-scale records are available at around surface fault traces, the near-field components can strongly constrain the timing of local rupture propagation (*see* Chapters 4.15 and 4.18).

Many slip models have been determined using strong-motion records. The resolution and reliability of each model depend mainly on the quantity and quality of strong-motion data. Digital data generally have wide dynamic range than digitized analog data and the accuracy of absolute timing is often critical for the analysis. Strong-motion data are usually limited and only small number of earthquakes have been observed at many near-field stations. **Table 1** summarizes the slip models determined by relatively large amount of data, together with the models in early studies for the San Fernando earthquake and the Imperial Valley earthquake. The number of strong-motion stations and types of data increased with time and the number of model parameters also increased until recently. While one research group uses strong-motion data as velocity data after integration, another group uses them as displacement data after double integration. The choice actually controls the roughness of the slip model as we will discuss in Section 4.07.3.3.

Recently we have been able to easily access several data sets of seismic waves. Global data had been available to world researchers before the first slip inversion study. World Wide Standard Seismograph Network (WWSSN; Oliver and Murphy, 1971) constructed in the 1960s was also useful for the study of seismic sources. This network has been gradually replaced by the network of broadband seismometers with digital recording system at more than 120

Table 1 Well-studied earthquakes and slip models

Earthquake	Date (year/month/day)	M_w	Reference	Near-field data		Other data	Model
				Type[1)]	Freq. (Hz)	Type[2)]	Type[3)]
San Fernando, USA	1971/02/09	6.7	Trifunac (1974)	D (5)	<0.11	—	L (22)
Imperial Valley, USA	1979/10/15	6.4	Olson & Apsel (1982)	A (26)	0.1–0.33	—	L (320)
			Hartzell & Heaton (1983)	V (12)	0.1–1.0	FD (19)	L (336)
			Archuleta (1984)	D (18)	<0.5	—	N (840)
Loma Prieta, USA	1989/10/18	6.9	Beroza (1991)	D (20)	0.3–2.0	—	N (1680)
			Steidl *et al*. (1991)	V (38)	0.05–1.0	—	N (456)
			Wald *et al*. (1991)	V (16)	0.1–1.0	FV (16)	L (288)
			Liu and Archuleta (2004)	V (16)	0.05–1.0	—	N (480)
Landers, USA	1992/06/28	7.2	Cohee & Beroza (1994)	D (18)	0.018–0.25	—	L (612)
			Wald & Heaton (1994)	D (16)	0.077–0.5	FD (11), G	L (1116)
			Cotton & Campillo (1995)	D (11)	0.05–0.5	—	N (144)
Northridge, USA	1994/01/17	6.7	Hartzell *et al*. (1996)	V (35)	0.1–1.0	—	N (784)
			Wald *et al*. (1996)	V (38)	0.1–1.0	FV (13), G	L (1176)
Hyogoken-Nanbu (Kobe), Japan	1995/01/17	6.9	Horikawa *et al*. (1996)	V (18)	0.05–0.5	G	L (480)
			Sekiguchi *et al*. (1996)	D (19)	0.1–1.0	—	L (2240)
			Wald (1996)	V (19)	0.05–0.5	FV (13), G	L (1728)
			Yoshida *et al*. (1996)	D (18)	<1.0	FD (17), G	L (360)
			Ide & Takeo (1997)	D (18)	0.025–0.5	—	L (3344)
Chi-Chi, Taiwan	1999/09/21	7.6	Chi *et al*. (2001)	V (21)	0.02–0.5	—	L (8320)
			Ma *et al*. (2001)	D (21)	no filtering	FD(22), G	L (4032)
			Wu *et al*. (2001)	V (47)	0.016–0.5	G	L (4840)
			Zeng *et al*. (2001)	V (15)	0.05–0.25	G	N (960)
			Ji *et al*. (2003)	V (36)	<1.6	—	N (1620)
Tottori, Japan	2000/10/06	6.7	Iwata and Sekiguchi (2002)	V (12)	0.1–1.0	G	L (1632)
			Semmane *et al*. (2005)	D (23)	0.1–1.0	G	N (480)

1) A, acceleration; V, velocity; D, displacement; the number of stations in parenthesis
2) FD, far-field displacement; FV, far-field velocity; G, geodetic data; the number of stations in parenthesis
3) L, linear presentation; N, nonlinear presentation; the number of parameters in parenthesis

worldwide stations (Lee, 2002). Today, we can obtain these data from Data Manager Center (DMC) of Incorporated Research Institutions for Seismology (IRIS) from a few hours after the occurrence of earthquakes. Recently, near-field strong-motion stations have been organized. Websites are maintained by some large regional networks of strong-motion seismometers such as California Strong Motion Instrumentation Program (CSMIP), in USA, Kyoshin network (K-NET), and KiK-net in Japan. The Consortium of Organizations for Strong-Motion Observation Systems (COSMOS) operated since 2001 provides a virtual server (search engine) for the strong-motion data from stations worldwide.

While new earthquakes provide new data sets for slip inversion, old analog seismograms were digitized and used for slip inversion of old significant earthquakes. It is a time-consuming task to digitize the seismograms from smoked papers and microfilms and apply appropriate corrections to them. Sometimes the information necessary for the corrections, such as the pendulum period, magnification, and pen arc length, is missing. Nevertheless, owing to recent progresses in image processing technique, the study of such old earthquakes is getting popular. For example, the slip models are determined for the 1906 San Francisco earthquake (Mw 7.7; Wald *et al*., 1993; Song *et al*., 2005), the 1923 Kanto, Japan, earthquake (Mw 7.9; Wald and Somerville, 1995), the 1944 Tonankai, Japan, earthquake (Mw 7.9; Kikuchi *et al*., 2003; Ichinose *et al*., 2003), and the 1948 Fukui earthquake (Mw 6.8; Kikuchi *et al*., 1999; Ichinose *et al*., 2005).

Although the resolution and reliability are not comparable to the state-of-the-art results, analyses of these events are important for the study of not only earthquake physics, but also for the regional tectonics and the assessment of long-term earthquake potential.

Only the final slip distribution (without its temporal evolution) can be estimated using geodetic data alone (e.g., Ward and Barrientos, 1986; Yabuki and Matsu'ura, 1992). Geodetic data can be combined with seismic data to better constrain the amount of total slip, which is often poorly constrained from seismic data alone. The examples of geodetic data are triangulation (e.g., Yoshida and Koketsu, 1990), leveling (e.g., Horikawa *et al.*, 1996; Yoshida *et al.*, 1996), GPS (e.g., Wald and Heaton, 1994; Wald *et al.*, 1996; Horikawa *et al.*, 1996; Yoshida *et al.*, 1996), and synthetic aperture radar interferometry (InSAR; Hernandez *et al.*, 1999; Delouis *et al.*, 2002; Salichon *et al.*, 2003, 2004). Geodetic data can be used to determine the slip history of extraordinary slow slip events, such as volcanic deformation (e.g., Aoki *et al.*, 1999) and interplate silent earthquakes (Yagi *et al.*, 2001; Miyazaki *et al.*, 2003, 2006). Tidal gage records of tsunamis excited by earthquakes, which provide information about low-frequency behavior of fault slip, are also used to constrain slip distributions (e.g., Satake, 1989, 1993; Tanioka *et al.*, 1996). Since typical tsunamigenic earthquake occurs outside of regional seismic and geodetic networks, tidal gage data are useful to independently constrain the slip distribution.

Before solving an inversion problem, all these data must be properly preprocessed and arranged into a data vector $\mathbf{d}^0$. Analog data require digitization and digital data often need decimation. The seismometer response should be removed by deconvolution to get displacement or velocity data. In this preprocessing, we can estimate the level of noise, apply a bandpass filter to reduce them, and remove characteristic noises if they exist. Problems in the treatment of digital data, such as deconvolution and filtering, are discussed in detail in the textbook of Scherbaum (1996). As seen in **Table 1**, the type of data (displacement, velocity, or acceleration) is different among studies, which controls the complexity of the slip model.

When we use different types of data or data from different stations simultaneously, we must determine the relative weights of these data. The weight should be determined according to the magnitude of noise in each data set. In addition to natural background noises and instrumental noises, we should consider model errors due to incompleteness of our knowledge to calculate synthetic data using assumed models. In the case of seismic data, the largest source of uncertainty comes from the inaccuracy of Green's functions due to uncertain underground structure. The magnitude of such errors is roughly proportional to the amplitude of calculated synthetic data. Therefore, to equalize the weight of all data, we may normalize data by their maximum value for each station. Furthermore, when the spatial distribution of stations is not homogeneous, we should increase the weights for the stations in more sparsely covered areas to regularize the problem (*see* Chapters 3.11 and 3.12).

4.07.2.3 Setting Fault Models

4.07.2.3.1 Parametrization of slip distribution

When we model an earthquake rupture process, we usually assume one or a few rectangular fault planes in the source region. Each rectangular plane has seven parameters: the location of one corner (latitude, longitude, and depth), the size (length and width), and the orientation (strike and dip angles). With two parameters concerning the slip, the slip amount, and the rake angle, these nine parameters represent the macroscopic image of the static fault slip as defined in Aki and Richards (2002). The determination of these macroscopic parameters had been a major research topic in earthquake seismology till the 1980s and numerous macroscopic static fault models have been determined for many earthquakes.

To discuss the spatial slip distribution, we usually divide the rectangular fault planes into a number of subfaults. If we emphasize the objectivity of slip models, the macroscopic fault parameters, seven for each plane, should be determined in slip inversion. However, the problem that solves heterogeneous slip distribution and fault geometry simultaneously is nonlinear and intractable. Instead, we usually assume these seven parameters based on the results of macroscopic analysis with low resolution or supplemental information such as the aftershock distribution, the surface fault trace, and the focal mechanism solutions including the centroid moment tensor (CMT)

solution that represents force system acting at the source (Dziewonski and Woodhouse, 1983).

Figure 1 shows the variety of assumed fault models for the Kobe earthquake from the catalog of Hashimoto (2005). All the assumed planes are in the clouds of aftershocks. The difference of location is as large as 5 km in the southern part because there are fewer stations compared to the urbanized northern area. The final slip distributions of all models have common characteristics such as large shallow slip in the southwestern part and rather deep small slip in the northeastern part. Cho and Nakanishi (2000) used 3-D moment tensor distribution instead of 2-D planes and also found similar characteristics. The fact that their 3-D distribution of seismic moment coincides with the short-term aftershock distribution implies that the data have enough resolution to determine the location of fault planes. When there are stations near the surface offsets, the geometry of fault at surface needs to be determined in detail, as complex curves. For example, researchers assumed a set of complex fault planes for the Chi-Chi earthquake since it is recorded at many stations near fault traces. However, it is not always clear whether the same geometry continuously extends downward and the assumption of deep complex fault plane may have unwanted effects on the result.

Once fault planes are assumed, the spatial and temporal slip distribution in the ith direction on each of the fault planes is spatially expanded by $\mathcal{J}$ basis functions as

$$\Delta\dot{u}_i(\mathbf{x},\ t) = \sum_{j=1}^{\mathcal{J}} p^j \hat{u}_i{}^j(t)\phi^j(\mathbf{x})f^j(\mathbf{x},\ t) \qquad [2]$$

where $\hat{u}_i{}^j(t)$ is a unit vector representing the slip direction. The direction of slip is always parallel to the fault plane and generally a time-dependent function. $\phi^j(\mathbf{x})$ is a spatial basis function and $f^j(\mathbf{x}, t)$ is a slip time function. Both $\phi^j(\mathbf{x})$ and $f^j(\mathbf{x}, t)$ are normalized so that the integral of $\phi^j(\mathbf{x})\, f^j(\mathbf{x}, t)$ for space and time is unity. Therefore, the physical dimension of p^j is equal to a volume. A popular rectangular subfault division corresponds to assuming $\phi^j(\mathbf{x})$ as boxcar functions each of which does not overlap with another. In this case, p^j represents the seismic

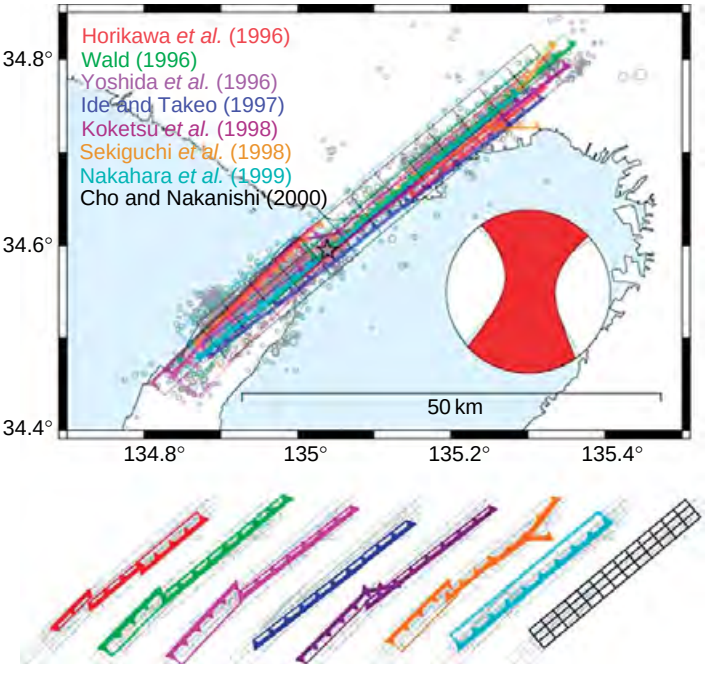


Figure 1 Fault models assumed in the studies of slip inversion by different research groups for the 1995 Kobe earthquake. Except for the model of Cho and Nakanishi (2000), colored bold lines with triangles indicating the direction of dip show uppermost edges and colored thin lines show fault planes projected to the surface. Cho and Nakanishi (2000) used 3-D blocks instead to the fault planes. Star shows the epicenter location determined by Japan Meteorological Agency (JMA). Black circles are aftershock epicenters determined by JMA within one week of the mainshock. Harvard CMT solution is shown by a red beach ball. Each of the bottom figures shows one of the fault models by a color and the others by gray lines.

potency of the subfault, which is the product of the subfault area and the slip amount. The seismic moment of the subfault is the potency p^j multiplied by the rigidity. Linear or cubic spline functions are often adopted as $\phi^j(\mathbf{x})$. The point subevent representation of Kikuchi and Kanamori (1982, 1991) is essentially equivalent to the use of delta functions as $\phi^j(\mathbf{x})$.

Integrating eqn [2] to $t=\infty$ gives the final slip distribution as

$$\Delta u_i(\mathbf{x}) = \sum_{j=1}^{\mathcal{J}} p^j \hat{u_i}^j \phi^j(\mathbf{x}) \quad [3]$$

where $\hat{u_i}^j$ represents the slip direction of the final slip. We can express

$$\hat{u_i}^j = v_i^1 \cos\hat{\theta^j} + v_i^2 \sin\hat{\theta^j} \quad [4]$$

using two perpendicular unit vectors on the fault plane, $\mathbf{v}^1$ and $\mathbf{v}^2$. $\hat{\theta^j}$ is the angle of $\hat{\mathbf{u}}^j$ measured from $\mathbf{v}^1$ to $\mathbf{v}^2$. The final slip distribution is determined by potency parameters p^j with a fixed $\hat{\mathbf{u}}^j$, or potency parameters $p^{j1}=p^j\cos\hat{\theta^j}$ and $p^{j2}=p^j\sin\hat{\theta^j}$ for two unit vectors. Various geodetic data can constrain these static parameters. Unless we consider $\hat{\mathbf{u}}^j$ or $\hat{\theta^j}$ as unknown parameters, the slip distribution is expressed as a linear function of unknown parameters in eqn [3] and we can apply a linear inversion theory that will be described in the next subsection.

To discuss slip history, we have to parametrize the time function of each spatial basis (subfault), $\hat{u_i}^j(t)f^j(\mathbf{x},t)$. Two major ways of parametrization are often used. One is the multi-time-window method in which only slip (potency) amounts are unknowns. The other uses timings of rupture of each subfault as unknowns. Typically, the former is a linear problem with a large number of parameters, while the latter is a nonlinear problem with a small number of parameters. In the following, we explain both methods and discuss the strength and weakness of each method.

4.07.2.3.2 Example of linear expression: multi-time-window method

As reviewed in the introduction the multi-time-window inversion originates from Olson and Apsel (1982) and Hartzell and Heaton (1983), and it is currently the most popular inversion method of seismic waveform data to estimate a slip model. **Figure 2(a)** illustrates the multi-time-window method. We expand the time function $\hat{u_i}^j(t)f^j(\mathbf{x},t)$ in eqn [2] using K expansion coefficients q^k and a function $f(t)$ as

$$\hat{u_i}^j(t)f^j(\mathbf{x},t) = \sum_{k=1}^{K} q^k v_i^{jk} f(t-(k-1)\Delta t - t_0(\mathbf{x})) \quad [5]$$

where the unit vector v_i^{jk} indicates the direction of slip vector for the spatially jth and temporally kth discrete unit in the spatiotemporal source volume. Before $t=0$, $f(t)=0$ and integration of $f(t)$ up to infinity is unity. Practically, a boxcar function and a linear spline (triangle) function are most frequently used as $f(t)$. Δt is the difference between successive time functions (time windows). We force the expansion coefficient q^k to satisfy

$$\sum_{k=1}^{K} q^k = 1 \quad [6]$$

so that the meaning of p^j, potency, is unchanged. $t_0(\mathbf{x})$ is the first time when the fault can slip at location $\mathbf{x}$ and usually assumed that $t_0(\mathbf{x})=|\mathbf{x}-\mathbf{x}_0|/V_r$, where V_r is the propagation velocity of a hypothetical rupture front from the hypocenter $\mathbf{x}_0$. Therefore, all multiple time windows defined by [5] propagate at this velocity. It should be noted that V_r is different from the true rupture propagation velocity. Rather it is the maximum value of possible rupture velocity from the hypocenter to any point of the fault plane. If the rupture does not start at $t_0(\mathbf{x})$, the actual rupture speed is slower than V_r. When $V_r\to\infty$ the model has no assumption about the hypocenter location, which is useful to check the dependence of the solution on this assumption (e.g., Yoshida, 1992).

Substituting [5] into [2], and decomposing the slip vector into two directions like [4], we have

$$\Delta\dot{u}_i(\mathbf{x},t) = \sum_{j=1}^{\mathcal{J}}\sum_{k=1}^{K}(p^{jk1}v_i^1 + p^{jk2}v_i^2)\phi^j(\mathbf{x})f(t-(k-1)\Delta t - t_0(\mathbf{x})) \quad [7]$$

where $p^{jkl}=p^j\ q^k\ \cos\theta^{jk}$ and $p^{jk2}=p^j\ q^k\ \sin\theta^{jk}$. Namely p^{jkl} is the seismic potency of the kth slip direction on the spatially jth and temporally kth discrete unit. Equation [7] is the general form of source expression in multi-time-window inversion, where the number of unknowns is $2\mathcal{J}K$. The first slip inversion by Trifunac (1974) is categorized as the special form of [7] with $K=1$.

Substitution of [7] into [1] and integration for space and time yield a predictive data vector, $\mathbf{d}^e$ as

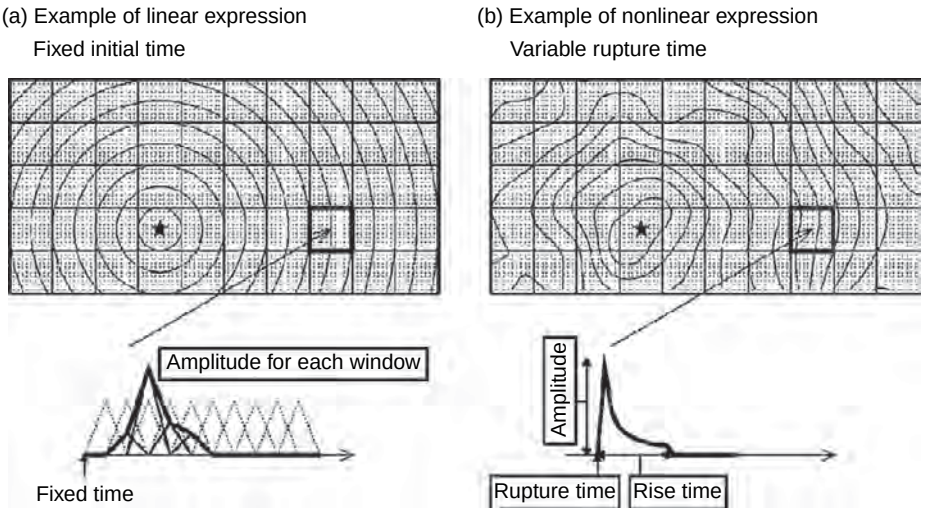


Figure 2 Schematic illustrations showing how to parameterize spatiotemporal slip distribution. The initial time of time function is shown by contours in each of the top figures. (a) Typical linear expression of multi-time-window analysis. Unknown parameters are amplitude for each window. (b) An example of nonlinear expression, in which rupture time is also an unknown parameter.

$$\mathbf{d}^e = \mathbf{Gm} \quad [8]$$

where $\mathbf{m}$ is a model vector that consists of potency parameters p^{jkl}, and $\mathbf{G}$ includes theoretical waveforms calculated for every discrete unit of spatiotemporal source volume. This linear equation is solved using standard procedures of the linear least-squares inversion (e.g., Menke, 1989) on the assumption that the difference between observation and model prediction follows the normal distribution with zero means. However, a flexible model requires a large number of model parameters. In recent analysis the number of parameters is sometimes larger than 3000–4000 (**Table 1**). Usually the number of data points is much larger than the number of parameters, but these data points are almost surely dependent and the problem tends to be underdetermined. The effective rank of matrix $\mathbf{G}$ is less than the number of parameters, and the best estimates of model parameters are, even if they were obtained, unstable with quite large standard deviations. Such problems require regularization with various constraints, which will be discussed in Section 4.07.3.2. Another deficiency of the multi-time-window method is that it limits the timing of slip. When the duration of the temporal basis functions is too short and real slip occurs longer than that, unexpected errors occur in the solution. This problem was demonstrated by Hartzell and Langer (1993) using the models having different number of time windows.

4.07.2.3.3 Example of nonlinear expression

If the timings of slip are treated as unknown parameters in Equation [2], we can reduce the number of parameters. An example is

$$f^j(\mathbf{x}, t) = f((t - t^j - t_0(\mathbf{x}))/r^j) \quad [9]$$

where $f(t)$ is a function which is zero before $t = 0$ ($f^j(\mathbf{x}, t)$ is 0 before $t^j + t_0(\mathbf{x})$) and normalized to be unity when integrated with time (**Figure 2(b)**). The parameter r^j is a scaling parameter in time. When $f(t)$ is a boxcar function from $t = 0$ to 1, r^j is called the rise time. A boxca function (e.g., Yoshida, 1986) and a triangle function (e.g., Fukuyama and Irikura, 1986; Takeo, 1987) are frequently used. A merit of nonlinear expression is that we can use a dynamically plausible function as $f(t)$. Beroza and Spudich (1988) assumed

$$f(t) \propto t^{-1/2}[H(t) - H(1-t)] \quad [10]$$

where H(t) is a step (Heaviside) function. This is a truncated version of the slip-rate function $t^{-1/2}$ in the analytic solution of dynamic crack propagation. We can change the direction of slip vector using unknown parameters θ^j as

$$\Delta\dot{u}_i(\mathbf{x}, t) = \sum_{j-1}^{\mathcal{J}} p^j v_i(\theta^j)\phi^j(\mathbf{x}) f((t - t^j - t_0(\mathbf{x}))/r^j) \quad [11]$$

Substituting [11] into [1] and integrations for space and time yields a predictive data vector from

the model parameter vector $\mathbf{m}$ consisting of p^j, t^j, r^j, and θ^j:

$$\mathbf{d}^e = \mathbf{G}(\mathbf{m}) \quad [12]$$

If we assume Gaussian errors of model prediction, linear inversion can provide a unique solution, while nonlinear inversion does not eliminate the problem of nonuniqueness. Instead, the model with nonlinear expression can be extended to include more general features of fault slip.

4.07.2.4 Calculation of Synthetic Data

To carry out integration of eqn [1] and obtain predictive data vectors [8] or [12], Green's function must be properly prepared for each point on the fault plane. Unless the spatial basis function $\phi^j(\mathbf{x})$ in eqn [2] is a delta function, we have to calculate Green's functions from every point source located at a small interval to account for rupture directivity effect in each discrete unit (subfault) as shown in **Figure 2**.

Theoretical formulation of elastic wave propagations from a point source in layered structure had been developed mainly during the 1970s and 1980s. For global studies, the method using a propagator matrix (Bouchon, 1976) and generalized ray theory (Langston and Helmberger, 1975) are frequently used to calculate far-field P- and S-waves with important contributions from local source and site structures, such as depth phases, sP, sP, and sS. For calculation of near-field full waveforms, the method that uses frequency–wave number integration is effective. The reflectivity method (e.g., Fuchs and Müller, 1971; Kind, 1976) and the Kennett–Bouchon method (e.g., Kennett and Kerry, 1979; Bouchon, 1981; Yao and Harkrider, 1983) are well-known methods and several numerical codes based on these methods are developed and distributed (e.g., Kohketsu, 1985; Takeo, 1985; Saikia, 1994). A general textbook about seismic wave propagation has been published by Kennett (2001). The combination of ray theory and isochrone integration (Bernard and Madariaga, 1984; Spudich and Frazer, 1984) is also effective in high frequency for near-field stations. For static deformations, the most popular study is Okada (1985) that gives analytic solutions for the surface displacement due to a buried dislocation on a rectangular fault in half space, with a concise summary of previous works (*see* Chapters 1.04 and 1.05).

The seismic velocity structure profiles determined by explosion surveys are helpful to construct a layered structure. Most studies assume a 1-D structure from the complex result of 2-D or 3-D surveys with minor improvements by trial and error. When the structure difference is large among stations, several 1-D structures are assumed in one study. Waveforms of small events are useful to empirically improve 1-D structure as shown by Ichinose *et al.* (2003) who compared synthetic waveforms with aftershock records to estimate average structures between the source and the stations.

Though 1-D structure is thought to be a fair approximation in many cases, it is obvious that there are many problems for which 1-D structure is no longer valid. For example, body waves from submarine earthquakes are followed by large reverberations from water layers and dip angle of sea floor is critical to simulate these waves (Wiens, 1987, 1989; Okamoto and Miyatake, 1989; Yoshida, 1992). Another example is the existence of thick sediment layers recently revealed by structural surveys in urbanized areas. Using a finite difference method for 3-D structure, Graves and Wald (2001) and Wald and Graves (2001) demonstrated that the use of 3-D Green's functions can increase the resolution of slip inversion. They also emphasized that this improvement is only possible with the accurate underground structure and careful examination of its validity is required for slip inversion. As demonstrated by Tsuboi *et al.* (2003) for the 2002 Denali Fault earthquake (Mw 7.9), one of the current fastest computers can calculate seismic waves up to 5 s in a laterally heterogeneous global Earth model. Introduction of more accurate 3-D Green's functions into slip inversion is promising.

Accurate 3-D structures have been determined only for small areas on the Earth. Moreover, small earthquakes radiate high-frequency waves, which cannot be modeled theoretically. An alternative of theoretical Green's function is empirical Green's function (eGf) which is usually the record of a small earthquake. This concept is originally proposed by Hartzell (1978) for strong-motion simulations. Note that the displacement of small earthquake is not a Green's function in the strict sense in eqn [1], where Green's function is defined using an impulsive single force while an earthquake has a double couple force system with a finite duration. Nevertheless, as long as the source duration of the small event is negligible, the difference of the force system is practically insignificant for slip inversion. Fukuyama and Irikura (1986) applied eGf to the 1983 Japan Sea earthquake (Mw 7.7). The large source area of this event requires two aftershocks each of which represents the north and the south subfaults, since an eGf is valid only for the limited space. For a large earthquake which

breaks shallow faults, the difference of distance from the free surface is hard to neglect and a number of aftershocks are required to obtain a complete set of empirical Green's functions, which is almost impossible. Since it is also known that aftershocks are rare in the area where the mainshock slip was large, it is hard to find proper small earthquakes for eGf method. Unfortunately, the waves from this most important part of the fault plane are most difficult to simulate using eGf. Nevertheless, in the study with relatively low resolution especially for small earthquakes, eGf is an effective tool as shown by many studies (e.g., Mori and Hartzell, 1990; Courboulex *et al.*, 1996; Fletcher and Spudich, 1998; Hellweg and Boatwright, 1999; Ide, 2001; Okada *et al.*, 2001). In these studies seismic waves at a station can resolve slip distribution mainly in the direction of the ray from the source to the station. To better resolve the slip distribution, a set of various ray directions is required. **Figure 3** is an example of the comparison between waveforms from the large event (the 2003 Miyagi-Oki, Japan, earthquake; Mw 7.1) and its possible eGf event of the similar mechanism. The waveforms of eGf events have short durations and site-specific complexity mainly due to local shallow structure. The mainshock records have some distinct phases and the arrivals of the second phases suggest the northward rupture propagation, which is consistent with the slip models presented by Okada and Hasegawa (2003) and Wu and Takeo (2004).

Misalignment between predicted and observed data can be a significant source of errors though it is not always carefully discussed in research papers. Usually the arrival times calculated for an assumed structure have some errors and sometimes data do not have accurate information of absolute timing. Therefore, we have to arrange predicted and observed data to minimize the difference between the corresponding reference phases. Usually S-waves are dominant in near-field displacement records but exact determination of S-wave arrival times is difficult because it is masked by P coda waves. It is plausible to estimate the S-wave arrival times using the records of small events in an empirical manner. Sometimes, poorly determined timings in near-field records lead to considerably different results of slip inversion. This might be one reason of large discrepancy between slip models for the 1999 İzmit earthquake (Mw 7.4; Yagi and Kikuchi,

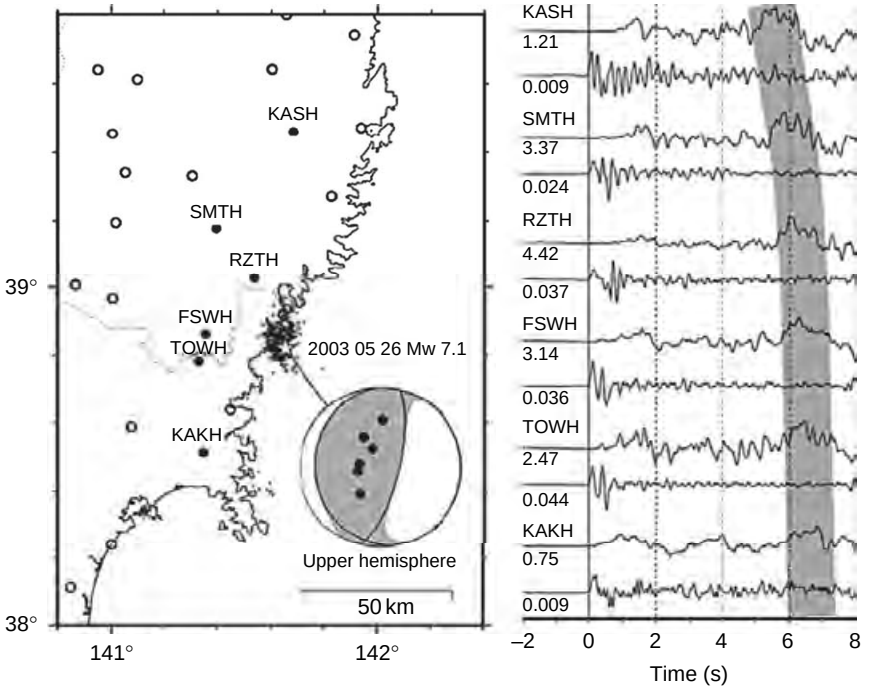


Figure 3 Comparison of waveforms of a large earthquake and its aftershocks, the 2003 Miyagi-Oki, Japan, earthquake. From all available Hi-Net stations (circles), we showed records at 6 stations (black circles) located in the map and on the focal sphere. For each station, upper and lower traces show the displacement (cm) of the mainshock and the aftershock, respectively. While the records of aftershock are impulsive, the waves of the mainshock last longer. In the shaded area, we can identify the second large phases that indicate rupture propagation northward.

2000; Bouchon *et al.*, 2002; Delouis *et al.*, 2002; Gülen *et al.*, 2002; Sekiguchi and Iwata, 2002) as pointed out by Beresnev (2003) and Ide *et al.* (2005).

4.07.3 Solving Inverse Problem

4.07.3.1 Best Estimate

From an observed data vector $\mathbf{d}^0$ and a corresponding prediction vector $\mathbf{d}^e$ including a model parameter vector $\mathbf{m}$, we solve the inverse problem to minimize the difference between these vectors, $\mathbf{d}^0 - \mathbf{d}^e$, and find the best estimate of the model parameters and their estimation errors. There are good textbooks for such inversion or optimization problems (e.g., Menke, 1989; Tarantola, 2005). Here we review only basic expressions using the maximum likelihood methods as an introduction to the next subsection.

The first step of inversion problem is to estimate the errors of the data vector. Usually the errors are assumed to have the Gaussian distribution with zero mean and a covariance matrix. In the next subsection we review Bayesian method in which the absolute value of covariance matrix is also estimated from the observed data through marginal likelihood. It is also necessary to estimate relative errors between independent data sets that can have different physical dimensions. For example, we estimate relative magnitude of errors between displacements of two stations, taking into account background noise levels and inaccuracy of the assumed model, that is, the assumption of fault planes and velocity structure. We can assess the effect of model inaccuracy by calculation using slightly different model settings. The relative magnitude of errors between data of different types, for example, seismic waveforms and static displacements, can be estimated similarly.

The maximum likelihood method gives the best estimate of model parameter vector by maximizing a likelihood function $p(\mathbf{d}^0|\mathbf{m})$. It is formally identical to the probability distribution function of the data $p(\mathbf{d}|\mathbf{m})$ but it is thought as a function of $\mathbf{m}$ for the observed data $\mathbf{d}^0$. With the assumption of the Gaussian error, it is

$$p(\mathbf{d}^0|\mathbf{m}) = (2\pi\sigma^2)^{-N_d/2}|\mathbf{E}_d|^{-1/2} \times \exp\left(-\frac{1}{2\sigma^2}(\mathbf{d}^0 - \mathbf{d}^e(\mathbf{m}))^t\mathbf{E}_d^{-1}(\mathbf{d}^0 - \mathbf{d}^e(\mathbf{m}))\right) \quad [13]$$

where we assume that the mean values and covariance matrix of the data as $\mathbf{d}^0$ and $\sigma^2\mathbf{E}_d$, respectively. Superscript t denotes the transpose of matrix. N_d is the number of data points. The maximum value of the likelihood function corresponds to the minimum of the $L2$ norm in the exponent, $(\mathbf{d}^0 - \mathbf{d}^e(\mathbf{m}))^t\mathbf{E}_d^{-1}(\mathbf{d}^0 - \mathbf{d}^e(\mathbf{m}))$. Therefore, this is the well-known weighted least-squares problem and the best estimate (maximum likelihood values) of the model parameter, $\mathbf{m}^0$, is given by the least-squares solution.

When the model is linear as eqn [8], that is, $\mathbf{d}^e(\mathbf{m}) = \mathbf{G}\mathbf{m}$, and the matrix $\mathbf{G}^t\mathbf{E}_d^{-1}\mathbf{G}$ is full rank, the solution is

$$\mathbf{m}^0 = (\mathbf{G}^t\mathbf{E}_d^{-1}\mathbf{G})^{-1}\mathbf{G}^t\mathbf{E}_d^{-1}\mathbf{d}^0 \quad [14]$$

The covariance matrix of the model parameter is written as

$$\mathbf{E}_m = \sigma^2(\mathbf{G}^t\mathbf{E}_d^{-1}\mathbf{G})^{-1} \quad [15]$$

However, the solution in the form [14] and [15] is rarely used. One reason is that usually the matrix $\mathbf{G}^t\mathbf{E}_d^{-1}\mathbf{G}$ is not full rank. The regularization to increase the rank will be discussed in the next subsection. Another reason is that [14] often gives the slips in the negative direction. Although it is still debated whether slip can occur in the negative direction, it is hard to imagine that slip reverses its direction under high shear stress environment in the crust, and so we usually assume that the slip is unidirectional. To prohibit slip in the negative direction, the non-negative least squares (NNLS; Lawson and Hanson, 1974) is useful. If the solution of NNLS exists, the uniqueness of the solution is verified by the Kuhn–Tucker theorem. A similar positive constraint is also possible with a nonlinear error function that rapidly increases near zero when we solve a nonlinear problem, for which the uniqueness is hard to prove. For example, Yoshida and Koketsu (1990) adopted an error function, $1/\sqrt{X}$, which rapidly increases as slip rate $X \to 0$.

When the model expression is nonlinear (eqn [12]), it is generally difficult to find the best estimates that minimize $(\mathbf{d}^0 - \mathbf{d}^e(\mathbf{m}))^t\mathbf{E}_d^{-1}(\mathbf{d}^0 - \mathbf{d}^e(\mathbf{m}))$. We can linerize the observation equation and apply an iterative inversion method such as the Gauss–Newton method and the Levenberg–Marquart method, in which the solution is derived through successive iterations of linear inversion. The estimate of the parameter vector after ith iteration $\mathbf{m}_i$ gives a prediction vector $\mathbf{d}^e(\mathbf{m}_i)$. Then the difference between the data vector and the prediction vector is approximated by Taylor series to the first order as

$$\mathbf{d}^0 - \mathbf{d}^e \approx \left(\frac{\partial \mathbf{d}^e}{\partial \mathbf{m}}\right)_{\mathbf{m}_i} \delta\mathbf{m}_i = \mathbf{G}'\delta\mathbf{m}_i \qquad [16]$$

where $\mathbf{G}'$ is a Jacobian matrix. In the Gauss–Newton method, the solution is derived similarly to eqn [14] and the updated model parameter vector is represented as

$$\mathbf{m}_{i+1} = \mathbf{m}_i + \delta\mathbf{m}_i = \mathrm{m}_i + \left[\mathbf{G}'^{\mathrm{t}}\mathbf{E}_d^{-1}\mathbf{G}'\right]^{-1}\mathbf{G}'^{\mathrm{t}}\left(\mathbf{d}^0 - \mathbf{d}_i^e\right) \qquad [17]$$

This procedure is iterated till the solution converges. However, the convergence does not guarantee that the solution is the best estimate because the solution depends on the initial value and can be just a local least-squares solution. In the case of slip inversion with many unknown parameters, many local least-squares solutions may reduce $L2$ norm similarly, and it is almost impossible to find a unique solution that globally minimizes the $L2$ norm. Jackson and Matsu'ura (1985) showed that prior information for model parameters, as discussed in the following section, can reduce the nonuniqueness.

Rapid improvement of computers has enabled seismologists to solve nonlinear problems by methods including random search iterations such as the Monte Carlo method. As examples to slip inversion in the early period, Hartzell and Liu (1995) and Hartzell *et al.* (1996) analyzed the Landers earthquake using simulated annealing (SA) and Zeng and Anderson (1996) used Genetic algorithm (GA) for the Northridge earthquake. Recently, Ji *et al.* (2002a) and Liu and Archuleta (2004) recommended the heat-bath algorithm (Rothman, 1986) that is a class of SA method and efficient for slip inversion.

Despite the frequent assumption of the Gaussian errors, which is supported by the central limit theorem, it is also known that real data include many 'outliers' that are far from the prediction and the errors cannot be explained by the Gaussian distribution. One remedy is just to remove these outliers from the data set, but it is also effective to use a long-tailed probability function instead of the Gaussian distribution, which gives more robust solution relatively independent of the outliers. If we assume that the probability function is double exponential (Laplace) distribution, a typical long-tailed function, the likelihood is written as

$$p(\mathbf{d}^0|\mathbf{m}) = 2^{-N_\mathrm{d}} \prod_{i=1}^{N_\mathrm{d}} \sigma_i \exp\left(-\sum_{i=1}^{N_\mathrm{d}} \left|\frac{d_i^0 - d_i^e(\mathbf{m})}{\sigma_i}\right|\right) \qquad [18]$$

The maximum likelihood solution minimizes the $L1$ norm,

$$\sum_{i=1}^{N_\mathrm{d}} \left|\frac{d_i^0 - d_i^e(\mathbf{m})}{\sigma_i}\right| \qquad [19]$$

This is called the least absolute values problem, which is solved using an iterative method. There is no dependence on the initial value like general nonlinear problem and the solution is obtained as a point or a finite hyperplane in the model space (e.g., Tarantola, 2005). The examples of slip inversion using $L1$ norm are found in Das and Kostrov (1990) and Hartzell *et al.* (1991) who solved the problem by the linear programming method, and in Ihmlé (1998) who used SA. The comparison of two solutions based on $L1$ and $L2$ norms is useful to find universal features of the solution that are independent of the norm assumption (Hartzell *et al.*, 1991).

4.07.3.2 Constraints and Regularization

Since the parametrization of slip distribution depends on the specific setup of the model, such as the size and locations of subfaults and the shape of time function, it is a good idea to use a large number of model parameters to make a flexible model. However, with a large number of model parameters, the solution in the form [14] and [15] cannot be obtained or become unstable with large parameter errors. This is because the effective rank of matrix $\mathbf{G}^{\mathrm{t}}\mathbf{E}_\mathrm{d}^{-1}\mathbf{G}$ is not full. Although the number of data N_d usually exceeds the number of model parameters N_m, the effective rank becomes smaller because the data are not independent or because a part of the model cannot be resolved by the data. Such a problem requires regularization with some constraints to model parameters. One of the simplest constraints is minimizing the $L2$ norm of the parameter vector. Another popular assumption in slip inversion is that the slip distribution is smooth temporally and spatially in some degree, which is expressed by constraints minimizing the time derivatives and the Laplacian of the slip distribution, respectively. Alternatively, we can fix the total amount of slip or seismic moment at a value suggested from other information (Das and Kostrov, 1990; Ji *et al.*, 2002b), for example the seismic moment of CMT solution determined using global long-period seismic waves (Dziewonski and Woodhouse, 1983).

The introduction of constraints raises another problem: how strong should these constraints be? A strong smoothing constraint makes the solution smooth, and vice versa. While early studies

introduced *ad hoc* criteria, Bayesian modeling and the information theory provide a way to find an objective solution with a proper weight. In the theory of Bayesian statistics, these constraints are thought to be prior information that provides a prior probability function (pdf) of model parameters with hyperparameters (Jackson, 1979; Jackson and Matsu'ura, 1985). When there is no prior information (constraints), we find the best estimate of model parameter vector maximizing the likelihood function as shown in Section 4.07.3.1. Similarly, with constraints as prior information, we find the solution maximizing a posterior pdf, which is proportional to the product of the likelihood function and the prior pdf. The proper weights of constraints, the best estimates of hyperparameters, are determined by maximizing the expectation of the likelihood for model parameters, a marginal likelihood. The condition is rewritten in a compact form of the minimum Akaike's Bayesian information criterion (ABIC; Akaike, 1980). ABIC was first applied by Yabuki and Matsu'ura (1992) and many applications of ABIC to slip inversion are found in, for example, Yoshida and Koketsu (1990), Ide and Takeo (1997), and Fukahata *et al.* (2004). We review the usage following Akaike (1980) and Yabuki and Matsu'ura (1992).

We may use several constraints of different types using constraint matrices $\mathbf{C}_j$ and constraint value vectors $\mathbf{e}_j$. A constraint is written with errors of Gaussian distribution with zero mean and a covariance matrix $\sigma_j^2\mathbf{I}$, $\mathbf{C}_j\mathbf{m} = \mathbf{e}_j + N\left(0, \sigma_j^2\mathbf{I}\right)$. To minimize the *L2* norm of the parameters, we use a unit matrix for $\mathbf{C}_j$ and a zero vector for $\mathbf{e}_j$. A typical smoothing constraint that minimizes the difference between two successive parameters is

$$\boldsymbol{C}_j = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ -1 & 1 & 0 & \\ 0 & -1 & 1 & \\ \vdots & & \ddots & \ddots \end{pmatrix} \qquad [20]$$

and $\mathbf{e}_j = 0$. When the parameters are constrained to some known values $\mathbf{e}_j$, $\mathbf{C}_j$ is a unit matrix. Each constraint defines a pdf of model parameters (Jackson and Matsu'ura, 1985):

$$p\left(\mathbf{m}|\sigma_j^2\right) = \left(2\pi\sigma_j^2\right)^{-N_j/2}\left|\mathbf{C}_j^{\mathrm{t}}\mathbf{C}_j\right|_D^{1/2} \times \exp\left(-\frac{1}{2\sigma_j^2}\left(\mathbf{e}_j - \mathbf{C}_j\mathbf{m}\right)^{\mathrm{t}}\left(\mathbf{e}_j - \mathbf{C}_j\mathbf{m}\right)\right) \qquad [21]$$

where N_j is the number of independent constraint equations, namely the rank of $\mathbf{C}_j$, and $||_D$ represents the product of all nonzero eigen values. When there are N_c constraint matrices, the joint pdf is given by a product of [21] for $j = 1, \cdots, N_c$ as

$$p\left(\mathbf{m}|\sigma_1^2, \ldots, \sigma_{N_c}^2\right) = A\exp\left(-\sum_{j=1}^{N_c}\frac{1}{2\sigma_j^2}\left(\mathbf{e}_j - \mathbf{C}_j\mathbf{m}\right)^{\mathrm{t}}\left(\mathbf{e}_j - \mathbf{C}_j\mathbf{m}\right)\right) \qquad [22]$$

where A is a normalization factor that keeps the integration of [22] for $\mathbf{m}$ to unity.

Hereafter we consider only one constraint matrix $\mathbf{C}_1$, whose rank is equal to N_m, for simplicity. Following a Bayes' theorem, a posterior pdf is obtained by multiplying eqn [21] and the likelihood function eqn [13]:

$$p\left(\mathbf{m}, \sigma^2, \sigma_1^2|\mathbf{d}^0\right) \propto p\left(\mathbf{d}^0|\mathbf{m}\right)p\left(\mathbf{m}|\sigma_1^2\right) \times \propto\left(\sigma^2\right)^{-N_d/2}\left(\sigma_1^2\right)^{-N_m/2}\exp\left(-\frac{\mathrm{Res}(\mathbf{m})}{2\sigma^2}\right) \qquad [23]$$

$$\mathrm{Res}(\mathbf{m}) = \left(\mathbf{d}^0 - \mathbf{G}\mathbf{m}\right)^{\mathrm{t}}\mathbf{E}_d^{-1}\left(\mathbf{d}^0 - \mathbf{G}\mathbf{m}\right) + \alpha_1^2\left(\mathbf{e}_1 - \mathbf{C}_1\mathbf{m}\right)^{\mathrm{t}}\left(\mathbf{e}_1 - \mathbf{C}_1\mathbf{m}\right) \qquad [24]$$

where, $\alpha_1^2 = \sigma^2/\sigma_1^2$. For given σ^2 and σ_1^2, the problem is formally identical to the maximum likelihood problem in eqns [13]–[15], and the solution is similarly obtained by maximizing the posterior pdf eqn [23], in other words minimizing the least square residual Res($\mathbf{m}$). The best estimates and the variance matrix of parameters are

$$\mathbf{m}_0 = \left(\mathbf{G}^{\mathrm{t}}\mathbf{E}_{\mathrm{d}}^{-1}\mathbf{G} + \alpha_1^2\mathbf{C}_1^{\mathrm{t}}\mathbf{C}_1\right)^{-1}\left(\mathbf{G}^{\mathrm{t}}\mathbf{E}_{\mathrm{d}}^{-1}\mathbf{d}^0 + \alpha_1^2\mathbf{C}_1^{\mathrm{t}}\mathbf{e}_1\right) \qquad [25]$$

$$\mathbf{E}_{\mathrm{m}} = \left(\mathbf{G}^{\mathrm{t}}\mathbf{E}_{\mathrm{d}}^{-1}\mathbf{G} + \alpha_1^2\mathbf{C}_1^{\mathrm{t}}\mathbf{C}_1\right)^{-1}\sigma^2 \qquad [26]$$

To find the best estimates of σ^2 and σ_1^2, we integrate [23] for the model parameters, and obtain the marginal likelihood as

$$L(\sigma^2, \sigma_1^2) = \int p(\mathbf{d}^0|\mathbf{m})p(\mathbf{m}|\sigma_1^2)\mathrm{d}\mathbf{m} \times \propto(\sigma^2)^{-N_d/2}\left(\alpha_1^2\right)^{N_m/2}\left|\mathbf{G}^{\mathrm{t}}\mathbf{E}_{\mathrm{d}}^{-1}\mathbf{G} + \alpha_1^2\mathbf{C}_1^{\mathrm{t}}\mathbf{C}_1\right|^{-1/2} \times \exp\left(-\frac{1}{2\sigma^2}\mathrm{Res}(\mathbf{m}_0)\right) \qquad [27]$$

where $||$ represents the determinant. At the maximum of the marginal likelihood, the variance of data error σ^2 is calculated as (Yabuki and Matsu'ura, 1992)

$$\sigma^2 = \mathrm{Res}(\mathbf{m}_0)/N_\mathrm{d} \quad [28]$$

Following the definition of well-known Akaike's information criterion (AIC), Akaike (1980) introduced ABIC as

$$\begin{aligned}\mathrm{ABIC} = &-2(\text{log marginal likelihood}) \\ &+2(\text{number of hyperparameters})\end{aligned} \quad [29]$$

For the current problem, the expression of ABIC is given by [27] and [28] as

$$\begin{aligned}\mathrm{ABIC}(\alpha_1) = &N_\mathrm{d} \ln \mathrm{Res}(\mathbf{m}_0) - N_\mathrm{m} \ln \alpha_1^2 \\ &+ \ln\left|\mathbf{G}^\mathrm{t}\mathbf{E}_\mathrm{d}^{-1}\mathbf{G} + \alpha_1^2\mathbf{C}_1^\mathrm{t}\mathbf{C}_1\right| + \mathrm{const}\end{aligned} \quad [30]$$

ABIC is a function of the hyperparameter α_1^2, and its minimum is found numerically.

When the constraint is too strong, that is, α_1^2 is very large, and the second and the third terms of [30] almost cancel, but the least-squares residual $\mathrm{Res}(\mathbf{m}_0)$ in the first term is also large and so is ABIC. When the constraints are too weak, α_1^2 is very small, the residual is small, but the second term becomes large to make a large ABIC. Between these two limits, we can find a minimum of ABIC. In [30] $1/\alpha_1^2$ defines the condition number of the matrix in the parenthesis of the third term. The tradeoff between the residual and the condition number of the matrix is essential for ABIC. Cohee and Beroza (1994) compared the residual and the condition number to find an appropriate value of weight of constraint. Equation [30] gives a statistical background for such a comparison. In other words, ABIC solves the well-known problem of tradeoff between the resolution and accuracy of the solution. Practically, finding the minimum of ABIC is not easy for general inversion problems including nonlinearity and non-Gaussian error distribution. However, at least in principle, we need not worry about this tradeoff by using Bayesian modeling and ABIC.

We can use a combination of different constraints using [22] instead of [21]. ABIC with two constraints is explained in Fukahata *et al.* (2004). **Figure 4** is a typical example of application of ABIC for slip inversion, with prior information that slip distribution is smooth in space and time. The synthetic waveforms from an input slip model (**Figure 4(a)**) contaminated with artificial noises are inverted for slip models (**Figures 4(b)–4(f)**) using different values of two hyperparameters in space and time, $\alpha_{\mathrm{space}}^2$ and α_{time}^2, respectively. The hyperparameters (weights of constraints) change the roughness of the slip distribution and source time functions. If the input model is consistent with the prior information and smooth to some extent, the slip model based on the minimum ABIC (**Figure 4(b)**) is similar to the input model.

4.07.3.3 Comparison in the Frequency Domain

The final goal of slip inversion is to obtain a model that can explain seismic waves over a broad frequency band. Accurate strong-motion prediction requires such a broadband seismic source. However, most previous slip models are determined by mainly low-frequency (long-period) seismic waves. Not many studies discussed validity of their model in the frequency domain. There have been a few studies that solved inversion problems in the frequency domain, such as Olson and Anderson (1988) and Cotton and Campillo (1995). The study of Ji *et al.* (2002a) using wavelet transform instead of Fourier transform is a mixed one in the time and frequency domains (wavelet domain). In this subsection, we discuss the nature of seismic waves at different frequencies and review studies that focused on the frequency contents of seismic waves.

A good reference model for seismic wave spectra is the omega square model (Aki, 1967; Brune, 1970), in which the displacement spectrum is proportional to

$$\frac{1}{1 + (\omega/\omega_0)^2} \quad [31]$$

where ω_0 is a corner angular frequency. High-frequency waves attenuate as squared frequency, and lower-frequency waves relative to the corner frequency dominate in displacement records. Since waveform inversion in the time domain gives a solution that explains waves of dominant frequencies, the inversion of displacement data tends to explain these low-frequency waves. Since velocity data have large amplitude around the corner frequency, the model is determined by waves around this frequency. Similarly, acceleration data have implicit weights above the corner frequency. Naturally, records of seismometers with different instrumental responses (e.g., Mendoza and Hartzell, 1988a), or records with different bandpass filters (e.g., Ide, 1999), result in the apparently different slip distributions. This is one reason of difference between models for the same earthquake, because different studies use different type (displacement, velocity, or acceleration) of data as shown in **Table 1**. To obtain a slip model that is valid over a broad frequency band, we must apply some

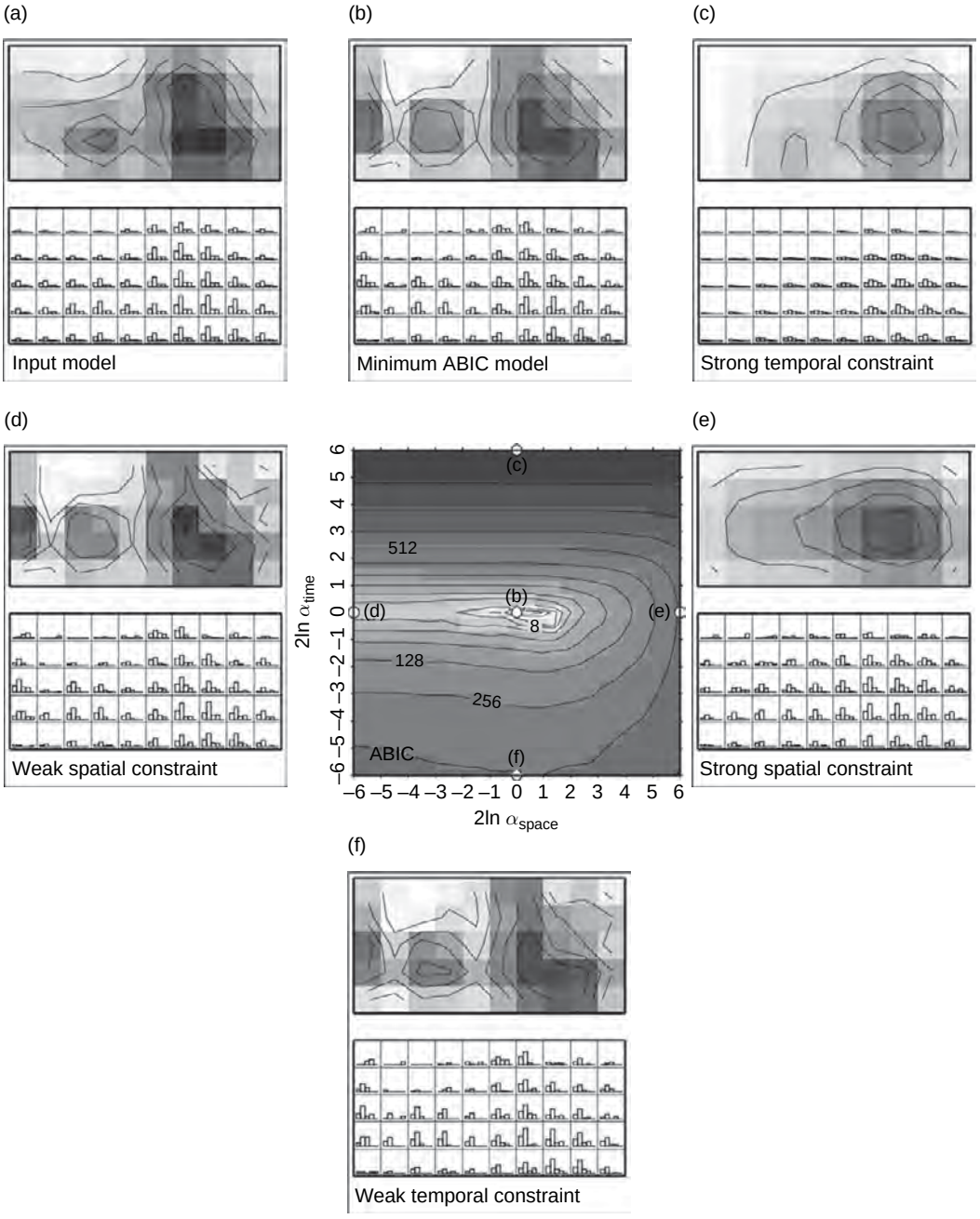


Figure 4 Simulated example of application of ABIC to slip inversion. Synthetic displacement at several stations surrounding the input source model are calculated and contaminated with noise of fractional Gaussian type. Slip models are determined for the different set of spatial and temporal hyperparameters, α^2_{space} and α^2_{time}, and the corresponding values of ABIC are shown in the central panel as a function of α^2_{space} and α^2_{time}. Since the absolute values are meaningless in this simulation, we set the origin of hyperparameters at the minimum of ABIC. (a) Input model: the slip distribution (upper panels) and time functions (lower panels). There are 10×5 subfaults and each subfault has five time windows consisting 250 parameters. (b) The model that provides the minimum ABIC. (c)–(f) Different models determined by stronger or weaker constraints.

weighting schemes that weaken this omega square frequency dependence of seismic waves, which is difficult in time domain.

Olson and Anderson (1988) proposed the first inversion in the frequency domain. They transformed time domain waves to frequency domain

spectra and also converted Green's functions and slip distribution to the frequency–wave number domain. They compared data and predictions at each frequency, which greatly simplifies the problem especially at low frequencies and small wave numbers, preserving its linearity. For high frequencies and large wave numbers, the number of parameters increases and some regularization is required. They proposed the minimum length regularization, but did not apply it to the real data. Mendez and Anderson (1991) applied essentially the same method to the 1985 Michoacan earthquake (Mw 7.9) with a different regularization that minimizes an error function based on the discretized version of a wave equation, which is minimized by a propagating rupture at a constant velocity. One shortcoming of the frequency–wave number domain transform of slip distribution is that positivity constraint is difficult to apply because the slip at a point is represented by the summation of slip functions at various frequencies and wavelengths.

Cotton and Campillo (1995) compared Fourier coefficients of data and model prediction in the frequency domain but they modeled slip distribution in the time domain, in similar form to the nonlinear expression in **Figure 2(b)**. They applied this method to the real data of the Landers earthquake and obtained a slip model which is consistent with other studies. Actually, their method is essentially the same as an ordinary time domain inversion that compares the time domain data $\mathbf{d}^0$ and corresponding model predictions $\mathbf{d}^e$ and minimizes the least-squares residual, $(\mathbf{d}^0 - \mathbf{d}^e)^t(\mathbf{d}^0 - \mathbf{d}^e)$. This value is an invariant for any orthonormal transformation $\mathbf{T}$.

$$(\mathbf{T}\mathbf{d}^0 - \mathbf{T}\mathbf{d}^e)^t(\mathbf{T}\mathbf{d}^0 - \mathbf{T}\mathbf{d}^e) = (\mathbf{d}^0 - \mathbf{d}^e)^t(\mathbf{d}^0 - \mathbf{d}^e) \qquad [32]$$

Since Fourier transform is a kind of orthonormal transformation, the comparison of Fourier coefficients in the frequency domain gives an identical residual to the one in the time domain. This property is also known as Parseval's theorem. Nevertheless, the first evaluation of validity of slip model in the frequency domain was a significant result.

Although Fourier transform is popular and its properties are well known, it has an aspect that is not easy to deal with. Fourier transform of a boxcar function in the time domain extends over infinite frequency band, and vice versa. In contrast, orthonormal wavelet transform is a transform between a finite time domain signal and a finite frequency domain spectrum. Ji *et al.* (2002a) adopted Meyer-Yamada wavelet transform to manifest contribution of finite frequency signal in slip inversion. Their method enables easy reweighting in each frequency band in the wavelet domain. They applied this method to the Hector Mine earthquake (Mw, 7.1; Ji *et al.*, 2002b).

While constructive wave interference occurs at low frequencies, high-frequency seismic waves are incoherent and the wave energies, squared amplitudes, are summed up at observation point. Special methods have been proposed to invert high-frequency incoherent waves and find the origin of these waves. A class of slip inversion developed for high-frequency waves is the envelope inversion. Synthetic envelopes from a source model are calculated using isochrone approximation of Spudich and Frazer (1984) (Zeng *et al.*, 1993), using empirical Green's function (Hartzell *et al.*, 1996; Kakehi and Irikura, 1996, 1997; Kakehi *et al.*, 1996), and based on the scattering theory of Sato *et al.* (1997) (Nakahara *et al.*, 1998, 1999). In many cases, the sources of high-frequency waves seem to be located near the periphery of large slip, which suggests high-frequency wave generation due to rapid accelerations and decelerations of the rupture front (Madariaga, 1977). However, many exceptions exist for this observation and there is no general relation between low-frequency slip distribution and high-frequency energy radiation. Apparently this is a field that requires much future investigation.

4.07.4 Example of Slip Model: The 1999 Chi-Chi Earthquake

The Chi-Chi, Taiwan, earthquake is an interplate thrust earthquake between the Eurasian and the Philippine Sea plates which occurred on 21 September 1999 (local time). Surface offsets appeared along an 85 km segment of the Chelungpu Fault. The maximum offset of about 8 m is located near the northern end where fault trace becomes complex and changes the strike direction to the east (**Figure 5**). Aftershocks are distributed to the east of this surface trace, reflecting the east dipping fault plane. Digital strong-motion records are available from 441 stations, 60 of which are within 20 km from the surface trace. These stations are maintained by the Taiwan Strong-Motion Instrumentation Program (Lee *et al.*, 2001). Furthermore, coseismic static displacements at about 100 sites are available from GPS survey (Yu *et al.*, 2001). They comprised

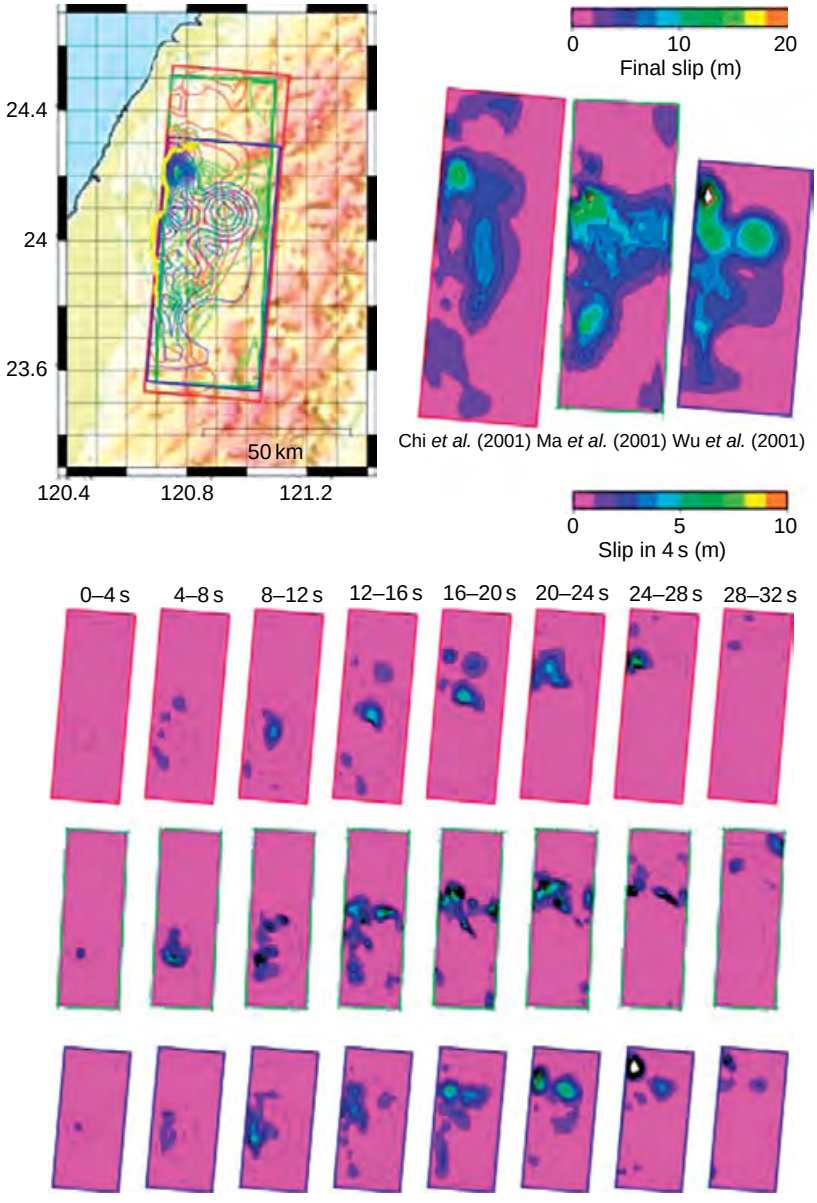


Figure 5 Comparison of three slip models by Chi *et al*. (2001), Ma *et al*. (2001), and Wu *et al*. (2001). All distribution is shown as the projection to the free surface. (Top left) Location in the map. Surface offsets along the Chelungpu Fault is shown by a yellow line. (Top right) Final slip distributions. (Bottom) Slip distributions during each time period.

the best data set in quality and quantity among all the earthquakes before 2005.

Several groups determined a spatiotemporal slip model using this data set. Chi *et al.* (2001) assumed one fault plane and applied a multi-time-window method to invert strong-motion velocities and surface offsets. Wu *et al.* (2001) used spatiotemporal basis functions and Bayesian modeling to jointly invert strong-motion velocities and GPS displacements. They assumed from one to three fault planes to explain the northern fault bend. Ma *et al.* (2001) used a multi-time-window method for strong-motion displacements, far-field displacements, and GPS data. They compared two analyses with and without a northern small fault striking to the northeast. The northern small subfaults assumed by Wu *et al.* (2001) and Ma *et al.* (2001) were necessary to explain the records at some stations close to the surface offsets, but they are almost independent of the rest of the data. All three studies used linear parameterization and

solved using NNLS. The large number of parameters from 4000 to 8000 requires regularization to get reasonable solutions. On the other hand, Zeng and Chen (2001) inverted strong-motion and GPS data to determine the distribution of slip, rupture time, and rise time in a nonlinear expression using a genetic algorithm. Their fault plane has a 3-D geometry, an extension of surface trace to the eastward down. Ji *et al.* (2003) applied their wavelet transform inversion method with simulated annealing to resolve slip distribution on a 3-D complex fault plane.

Figure 5 shows the exact locations of the assumed fault planes, final slip distributions, and propagation behaviors of rupture area of Chi *et al.* (2001), Ma *et al.* (2001), and Wu *et al.* (2001) (models C, M, and W respectively). Models M and W are the single fault models in these papers and the slip data of models C and M are taken from the supplemental CD of the original papers. Although the values of seismic moment are written as $(2.7–4.7) \times 10^{20}$ Nm in the original paper, all the seismic potencies calculated from the model parameters are within $(1.0–1.15) \times 10^{10}$ m^3, which is reasonable because the colors of final slip in **Figure 5** are apparently similar.

The common features visible in these models are: (1) total potency and rupture area, (2) small slip around the hypocenter, (3) large and deep slip in the northern part, (4) small and shallow slip in the southern part, and (5) the maximum slip of about 20 m. The good agreement between models M and W is probably due to addition of GPS data. In model C, the slip area extending over the northern end of the surface trace is hard to explain. The above features from (1) to (5) are also found in the models of Zeng and Chen (2001) and Ji *et al.* (2003).

The most prominent similarity among the slip histories of these models is the large slip at 40 km north of the hypocenter and 20–24 s after the hypocentral time. The rupture propagation velocity is less than 2 km s^{-1}, which is slower than the average shear wave velocity in the source area of 2.5–3.0 km s^{-1}. This is also shown by Zeng and Chen (2001) and Ji *et al.* (2003). However, there are considerable discrepancies such as the timing of slip propagating to the southward. Slip amount within each period does not match well. In **Figure 5**, the origin time of model M is shifted 2 s to match other models better. This difference is not small but it is not unrealistic because the strong-motion records used in this study do not have accurate timing information and researcher independently corrected them based on other information.

When large amounts of data are available as the Chi-Chi earthquake, we observe similarities among models, which are certainly constrained by data regardless of slightly different model assumptions. A dense GPS network is especially useful to constrain spatial distribution of slip. The amount and timing of shallow slip cannot be very different when the measurements of surface offsets and very near-fault strong-motion records are available. While very near-fault records are only sensitive to the local slip, a good coverage of seismic data is essential to reveal spatio-temporal rupture propagation. All these data are available for the Chi-Chi earthquake. Similarly, we can find common timing and spatial extent of large slip in the models of well-studied events with sufficient data, such as the Landers, Northridge, Kobe, and Tottori earthquakes (**Table 1**). Nevertheless, we should also acknowledge that considerable differences exist in the details of slip model that is affected by slight difference of the way of analysis even with such good data set. The behavior of small slips, often including those in the initial stage of rupture, is difficult to resolve.

4.07.5 Extended Studies Based on Slip Models

4.07.5.1 Characteristics of Slip Models

There are qualitative features frequently pointed out in many slip models, for example, pulse-like slip propagation, few aftershocks in the area with large slip, and the tendency of rupture initiation from the bottom of the fault plane. In this section, we review the discussion and physical interpretation of these features.

In many slip models, slip propagates like a pulse with a narrow width compared to the size of whole fault plane. Summarizing seven slip models, Heaton (1990) concluded that the slip pulse is a general feature of seismic rupture process and not an artifact due to the setting of time windows or limited frequency contents of inverted data. One good example among later studies is the Landers earthquake that ruptured 65 km faults, but the width of slip area at any moment is less than 15 km (e.g., Wald and Heaton, 1994). Heaton (1990) proposed that the velocity-weakening nature of rock friction is responsible for pulse-like behavior. Cochard and Madariaga (1994) confirmed the generation of slip pulse using velocity-weakening friction law in numerical simulation of dynamic crack propagation. However, this is

not the only mechanism of slip pulse. Alternatively, Beroza and Mikumo (1996) and Peyrat *et al.* (2001) demonstrated that spatial heterogeneity of stress can produce slip pulse without a velocity-weakening friction law.

A reliable gross feature of slip inversion is the location and time of the largest slip. The location and time of rupture initiation (hypocenter) are far more reliable. Therefore, comparing these numbers, we can estimate overall rupture velocity. In the 1970s, rupture velocity was discussed using macroscopic earthquake models. Geller (1976) compiled rupture velocity of many macroscopic models and gave a typical rupture velocity as 72% of the S-wave velocity, with a caution that the value depends on the model assumption. The result of slip inversion is useful to measure model-independent rupture velocity. The rupture velocity of $2.2\,\mathrm{km\,s^{-1}}$ in the Chi-Chi earthquake (Section 4.07.4) is a well-determined example. It is easy to measure a rupture velocity when there is a slip pulse in the model. In the model of Wald and Heaton (1994), the Landers earthquake propagated at $2.5–3.0\,\mathrm{km\,s^{-1}}$, with deceleration near fault steps. The rupture velocity of the Sumatra earthquake is also fixed well as $2.8\,\mathrm{km\,s^{-1}}$ from the spatiotemporal energy radiation distribution determined by Ishii *et al.* (2005). In all models, average rupture velocity is less than S-wave velocity, but a little faster than 70% suggested from macroscopic source models.

Unlike the average rupture velocity, local rupture velocity can exceeds S-wave velocity, although it is not robust because of smoothing introduced in slip inversion. In the slip model of the Imperial Valley earthquake based on trial and error, Archuleta (1984) showed that local rupture velocity is faster than S-wave velocity. Frankel (2004) reported that the local rupture velocity is as fast as $5\,\mathrm{km\,s^{-1}}$ while the average is about $3\,\mathrm{km\,s^{-1}}$ in the case of the 2002 Denali fault, Alaska, earthquake (Mw 7.9). Such local super shear rupture is expected when stress is heterogeneous (Day, 1982a) or fracture energy is heterogeneous with homogeneous stress (Ide and Aochi, 2005). However, the observational evidence is not strong enough to conclude that such super shear rupture propagation is a general phenomenon of earthquake.

Estimation of rupture velocity for small earthquakes is difficult. Ide (2001) determined slip models of 18 moderate (M4–5) earthquakes that occured in the Hida Mountains, Japan, but the estimation of rupture velocity has large uncertainties because of poor resolution. In the smallest slip model ever determined for earthquakes, the model for the micro (M0.8–1.4) earthquake in South Africa gold mine, Yamada *et al.* (2005) found that the rupture velocity is not less than 65% of *S* wave velocity. Although it is not the results based on slip models, Imanishi *et al.* (2004) found high rupture velocity for small (M1.3–2.7) earthquakes in western Nagano, Japan, using the method utilizing stopping phases (Imanishi and Takeo, 2002). These studies suggests that the rupture velocity is similar for earthquakes from M1 to M9, but each estimation was subject to large errors and obviously further studies are necessary.

Slip models are frequently compared with aftershock distributions. It is often claimed that aftershocks are rare near or in the areas on the fault where the slip was large. Mendoza and Hartzell (1988b) investigated seven slip models of earthquakes in the western United States that are determined using near-field records and hence relatively reliable. Although it was qualitative, they showed that aftershocks occur mostly outside of or near the edges of the source areas indicated by the patterns of mainshock slip. Das and Henry (2003) investigated various slip models for many events including those in subduction zones and ridges. They found that few, and usually the smaller, aftershocks occur in the high-slip regions of the fault, with only one exception. Wiemer and Katsumata (1999) discussed slip models and b value of Gutenberg–Richter frequency-magnitude relation (Gutenberg and Richter, 1944). For the 1984 Morgan Hill (Mw 6.2), Landers, Northridge, and Kobe earthquakes, they found that large slip regions seem to correspond to large b values. However, it should be noted that large slip region has few aftershocks and the estimations of b value in most of these regions were impossible. Although the mechanism of aftershock genesis is not fully understood, the existence of many aftershocks surrounding the areas of large slip indicates the importance of the stress redistribution by the mainshock.

Another often reported behavior is rupture initiation from the bottom of the fault (e.g., Das and Scholz, 1983). One may easily compare aftershock distribution and the location of mainshock hypocenter, but the comparison between slip distribution and the initial rupture location is more informative. From the large collection of published slip models, Mai *et al.* (2005) compared the location of initial rupture and slip distribution and concluded that the rupture started from near the bottom of the assumed fault

plane in strike-slip and crustal dip-slip earthquakes. Although Das and Henry (2003) reported that the rupture tends to nucleate in the region of low slip or at the edge of high-slip regions, Mai *et al.* (2005) did not find such tendency.

Rupture direction of large earthquakes tends to be unidirectional as shown by McGuire *et al.* (2001) using a 2nd moment inversion method for far-field records. We can estimate the 2nd moment for slip models and determine overall rupture direction. McGuire *et al.* (2002) investigated slip models for earthquakes larger than M7 and confirmed his earlier conclusion based on far-field observation.

4.07.5.2 Implication of Slip Models for Fault Dynamics

Spatiotemporal slip distribution is directly connected with spatiotemporal stress distribution. Once spatiotemporal slip distribution is obtained, it is used as a boundary condition for calculation of stress change during the rupture propagation (Ide and Takeo, 1997; Bouchon, 1997). This yields a slip–stress relation at each point on the fault plane. **Figure 6** is the first result of such calculation for the Kobe earthquake (Ide and Takeo, 1997). Slip–stress relation is basically slip-weakening, similar to the theoretical assumption of Ida (1972) and Andrews (1976). The slip-weakening distance D_c is estimated as 50 cm to 1 m. However, it was also shown by Ide and Takeo (1997) that D_c less than 50 cm is not detectable for the Kobe earthquake because of the limited resolution. Olsen *et al.* (1997) also estimated D_c of 0.8 m for the Landers earthquake using dynamic modeling. However, this estimation might suffer the effect of the limited resolution, too. Guatteri and Spudich (2000) demonstrated that similar seismic waves are radiated from dynamic models of different D_c distribution, as far as they have similar distribution of fracture energy G_c, which scales with the product of D_c and breakdown stress drop (=difference between yield stress and dynamic friction level). The estimation of G_c is robust compared to that of D_c because G_c is controlled by rupture velocity and less affected by the limited resolution. Therefore, the most stable dynamic constant derived from slip inversion is not D_c, but G_c.

Beroza and Spudich (1988) first estimated G_c as 2 MJ m^{-2} from the slip model of the Morgan Hill earthquake. Guatteri *et al.* (2001) applied a rate- and state-dependent friction law for the model of Ide and Takeo (1997) and obtained a value of G_c as 1.5 MJ

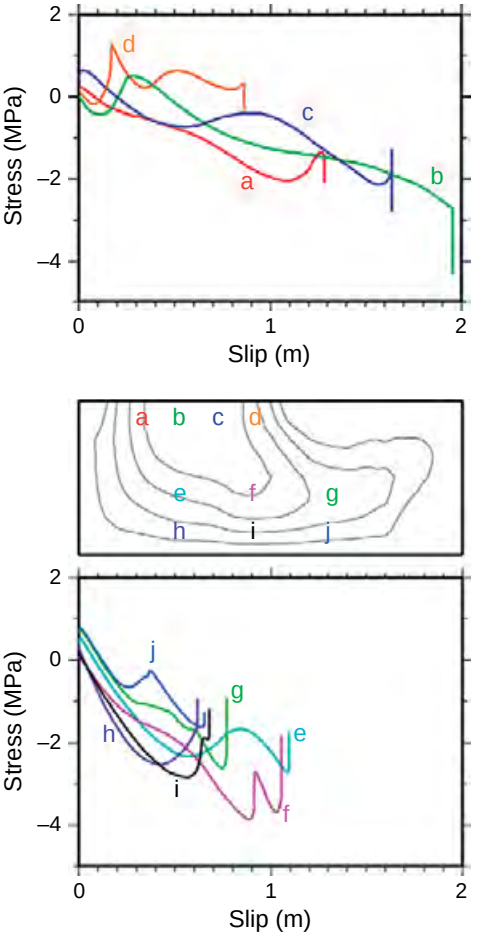


Figure 6 The relation between slip and stress on the fault plane, determined for the 1995 Kobe earthquake (Ide and Takeo, 1997). Each trace is the function calculated at the corresponding location in the middle figure that shows final slip distribution by contours.

m^{-2}. In addition to the fracture energy G_c measured from slip-stress relation, radiated energy from earthquake E_s is also calculated using the expression of Kostrov (1974) as

$$E_s = \frac{1}{2}\int_{\Sigma}(\sigma_{ij}(\mathbf{x},0)-\sigma_{ij}(\mathbf{x},\infty))\Delta u_i(\mathbf{x},\infty)\nu_j\,\mathrm{d}S - \int_{\Sigma}\gamma(\mathbf{x})\mathrm{d}S + \int_0^{\infty}\mathrm{d}t\int_{\Sigma(t)}\dot{\sigma}_{ij}(\mathbf{x},t)\Delta u_i(\mathbf{x},t)\nu_j\,\mathrm{d}S \quad [33]$$

where $\sigma_{ij}(\mathbf{x}, t)$ is the stress on the fault plane, and $\gamma(\mathbf{x})$ is the free surface energy to make an infinitesimally small slip on the fault plane. The third term contains the fracture energy derived from slip–stress relation. Summation convention is used. Ide (2002, 2003) applied this equation to slip models of several

earthquakes to estimate seismic energy and fracture energy. While the fracture energy is in the order of MJ m^{-2} for the Kobe earthquake, total radiated energy calculated by this method is smaller by a factor of 3 than the values determined using far-field body waves (Boatwright and Choy, 1986). This suggests that the slip model does not account for the total energy radiation because it is determined by seismic waves of relatively lower frequencies with smoothing constraints.

Another method to estimate the slip-weakening distance D_c was proposed by Mikumo *et al.* (2003) and Fukuyama *et al.* (2003). Instead of usual definition of D_c in a slip–stress diagram, they measure the slip amount when slip rate reaches the maximum as D_c'. Applying this method to the 2000 Tottori and the Kobe earthquake, they concluded that D_c' scales with the final slip. However, Spudich and Guatteri (2004) suggested that it can be an artifact due to low pass filters. In fact, when the frequency contents of the data have an upper limit, that limit affects the maximum value of D_c'. Therefore, as Spudich and Guatteri (2004) pointed out, it is important to compare data and model prediction in the frequency domain as reviewed in Section 4.07.3.3.

If fault plane is flat, absolute stress level is usually irrelevant to seismic wave radiation. However, when the rake of fault slip rotates during the rupture process and frictional traction is collinear with the instantaneous slip motion, there is a unique absolute stress consistent with the motion (Spudich, 1992). Some slip models allow rake angles to vary during slip, which can be used to estimate the absolute stress level (Spudich, 1992; Guatteri and Spudich, 1998). Since the details of slip distribution from slip inversion depend on researcher's assumptions, we should be careful about judging whether the observed rake rotation is constrained directly by the data, and not an artifact due to model assumptions.

4.07.5.3 Dynamic Modeling and Slip Models

Most slip models are kinematic models and some of the features are not physically realistic. For example, stress diverges to infinity around subfault boundaries with a constant slip and dynamically feasible slip function is usually asymmetric unlike the frequently used triangle function (**Figure 2**(**a**)). In the previous subsection, we reviewed dynamic implication directly deduced from such kinematic slip models. On the other hand, dynamic modeling is performed to construct an elastodynamically reasonable source model using only some aspects of slip inversion, for example, final slip distribution or rupture timing. A dynamic source model is constructed based on fracture mechanics, with the assumption of stress state and friction/fracture law. This is one practical way to discuss source behavior and seismic wave radiation at high frequencies where kinematic slip model is no longer reliable.

The problem of crack propagation in elastic medium has been solved numerically since the late 1970s (Andrews, 1976; Das and Aki, 1977; Day, 1982b). While early studies solved some ideal cases, the first application to the complex rupture process of real earthquake is Quin (1990) who carried out dynamic modeling of the Imperial earthquake based on the slip model of Archuleta (1984). He calculated the stress drop from the slip distribution of the slip model, assumed the distribution of strength, and simulated spontaneous crack propagation at prescribed rupture velocity of the original model satisfying the fracture criterion of the maximum stress (strength). Miyatake (1992) applied a similar method to four inland earthquakes in Japan. Since these procedures refer only final slip distribution and rupture timing of the original slip model, seismic waves radiated from the dynamic model cannot explain the observation. Fukuyama and Mikumo (1993) and Ide and Takeo (1996) developed dynamic modeling methods that iteratively improve dynamic models to explain observed waveforms. The above-mentioned studies of dynamic modeling assume the maximum stress fracture criterion (Das and Aki, 1977), which depends on the grid size. Rather, natural fault slips must obey a friction law. Olsen *et al.* (1997) carried out the first dynamic modeling to simulate the Landers earthquake, in which slip occurs following a slip-weakening friction law. This dynamic model is later improved by Peyrat *et al.* (2001) to determine heterogeneous stress distribution on a flat fault plane.

One of the recent major topics in dynamic modeling is the geometry of fault planes. Unlike a strike slip fault, a shallow dipping fault does not have symmetry because of the existence of the free surface. Oglesby *et al.* (1998) carried out dynamic modeling of a 2-D shallow thrust fault and showed that the change of fault normal stress has an important effect on the fracture criterion at the rupture front. This result suggests that a complex slip distribution that implies stress heterogeneity can be due to complex fault

geometry. Actually, Oglesby and Day (2001) showed that homogeneous stress and dipping fault plane can explain the gross feature of observed near-source ground motion for the Chi-Chi earthquake, for example, the larger movement in the hanging wall than the footwall. Dynamic rupture propagation along more complex fault system has become feasible. Aochi and Fukuyama (2002) showed that homogeneous background stress and change of strike angles in a fault system are essential to explain spontaneous rupture transfer between three major faults during the Landers earthquake. Aochi *et al.* (2003) also showed that the stress distribution of Peyrat *et al.* (2001) improves the details of dynamic slip distribution.

4.07.5.4 Scaling of Earthquake Heterogeneity

Most scaling relations between macroscopic parameters of earthquake source are derived from a geometrical similarity, that is, constant stress drop scaling (e.g., Kanamori and Anderson, 1975). However, the breakdown of the scaling relations is not well understood. There are still debatable issues, such as the effect of finite width of seismic layer on the scaling between fault length and slip amount (e.g., Matsu'ura and Sato, 1997; Scholz, 2002), and size dependence of the ratio between seismic energy and seismic moment, especially for small earthquakes (e.g., Ide and Beroza, 2001; Kanamori and Brodsky, 2001).

On the other hand, studies have recently started for the scaling of the complex earthquake source. It is not straightforward to extract meaningful parameters from various slip models. Using many published slip models, Somerville *et al.* (1999) determined the size of fault plane excluding the regions of little slip near the edges. Furthermore, they defined an asperity as "a region on the fault rupture surface that has large slip relative to the average slip on the fault," measured the total area of rectangular asperities in each model, and concluded that the asperities occupy 17.5% of the fault plane. The fault size of slip model is also measured by Mai and Beroza (2000) using autocorrelation function. They obtained a nonlinear relation between slip and fault length for large strike slip events. Mai and Beroza (2002) calculated autocorrelation of slip distribution and showed that they are explained by von Karman autocorrelation function or functions with power-law dependence at high frequencies. Such scaling about complexity of earthquake is useful to constrain input source models for strong-motion prediction. However, it should be noted that slip inversion includes many assumptions and smoothing constraints for regularization that greatly affect high-frequency components of seismic waves. Moreover, although previous studies dealt with only final slip distribution, time history of slip will be also important for such purpose in future studies.

Slip models have been obtained for moderate earthquakes of M3–4 (Courboulex *et al.*, 1996; Ide, 2001). Although the resolution is worse than that of large events like the events summarized in **Table 1**, these models can be used to compare slip area and aftershock distribution and to distinguish a simple rupture process from a sequence of discrete subevents. So far the slip models of the smallest size were determined for five microearthquakes in South African gold mine (Yamada *et al.*, 2005). **Figure 7** compares one of these models (Mw 1.4) with the slip models of the Kobe earthquake (Ide and Takeo, 1997). It is interesting that all scales differ about 1000 times, which implies a geometrical similarity roughly consistent to the 5.4 unit difference in moment magnitude scale. Although the resolution is poor, this small event certainly consists of subevents that are also identified directly in observed waves. Some small earthquakes do have complex rupture processes comparable to large events. However, it is still unclear how general such complexity is (*see* Chapter 4.18).

Seismic waves from an earthquake generally have so-called initial phases before the large main phase (e.g., Umeda, 1990; Iio, 1992, 1995; Abercrombie and Mori, 1994; Ellsworth and Beroza, 1995). Sometimes initial phases are quite different from the main phase. Nakayama and Takeo (1997) determined a slip model for the 1994 Sanriku-Haruka-Oki earthquake (Mw 7.7). This earthquake began with slow propagation at about 2.0 km s^{-1} with small slip during the first 20 s and later it was accelerated up to 3.0 km s^{-1} and radiated large seismic waves. The construction of slip models only for initial phases is possible. Shibazaki *et al.* (2002) determined a slip model for the initial 0.6 s of the Kobe earthquake whose total duration is about 10 s. Comparing with another slip model of an M4 aftershock in similar scale, they found that the rupture propagation of the mainshock is slightly slower than that of the aftershock. Such comparison is still rare, but it is an interesting future topic.

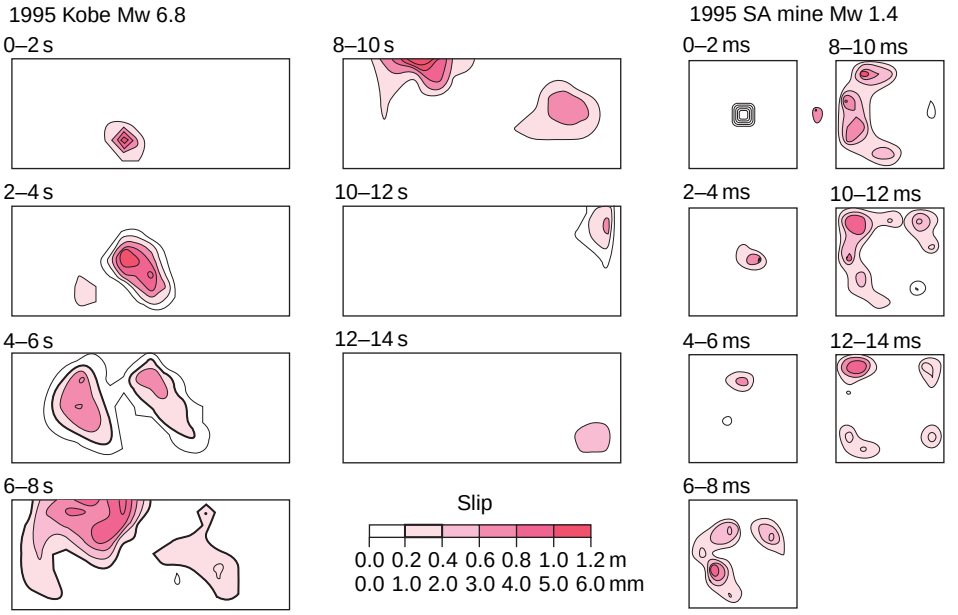


Figure 7 Comparison between two slip models: the 1995 Kobe earthquake and a small earthquake in South Africa gold mine. Slip distributions during each period are shown. The maximum slips are about 1.2 m and 6.0 mm for the Kobe and South African earthquakes, respectively.

4.07.6 Discussion and Conclusion

This chapter reviews the history, formulation, application, and extension of slip inversion using seismic waveform and/or geodetic data. We introduce some popular forms of slip inversion, dividing the problem into basic elements of data, model parameters, and synthetic data. The history of slip inversion started from the early 1980s, after the pioneering study of Trifunac (1974). Although the speed of development is not rapid since the establishment of the basic forms at around 1990, there are some recent technical improvements, such as application of various nonlinear minimization methods, objective weighting for regularization constraints, and addition of new kinds of data. Slip inversion in the frequency domain is still being developed (Miyake *et al.*, 2006). Except for these improvements, slip inversion is a mature analysis method of seismic sources. Some computer programs are distributed via the Internet (e.g., Kikuchi and Kanamori, 2004). Determination of slip model for relatively large earthquakes using far-field body waves and distributed programs is a kind of routine work or a good exercise for students. Routine reports of such slip models are usually published for significant earthquakes in electronic formats within one or a few days.

It is often overlooked that slip models have only limited reliability. Usually only gross features such as the location and timing of large slip, and average direction and velocity of rupture propagation are well resolved. Even such features were not identified for some events. An example is the 1999 İzmit earthquake, for which the strong-motion records are not sufficient and their timing information was not correct (Beresnev, 2003; Ide *et al.*, 2005). Each paper presenting slip model does not always mention the assumptions and limits of the inversion method, because these problems have been usually investigated in the previous papers. This ambiguity may cause artifacts when we evaluate characteristics and scaling of slip models or derive physical implications from slip models. There is no established way to qualitatively evaluate the reliability of slip models and we have to assess the limitation of models case by case.

The increasing power of computers has contributed to the improvement of slip inversion, increasing the number of data and model parameters. However, the number of model parameters is not increasing during this decade. Rather, computer power is recently used to solve nonlinear problems with a large number of unknown parameters using Monte Carlo-like iterative methods. There are large rooms to improve optimization methods for various

problems of slip inversion. The degree of freedom is high and the review of this field in this chapter (Section 4.07.3.1) is not sufficient. Another field that requires powerful computer is calculation of Green's function for heterogeneous Earth structure. One landmark is given by Tsuboi *et al.* (2003) in the forward simulation of the Denali Fault earthquake using Earth simulator. Such calculation will increase resolution of slip inversion not only for global problem, but also for regional earthquakes.

Needless to say, the quality of slip inversion fully depends on the data. Geodetic data will be important for slip inversion to fix the lower-frequency image of the source. InSAR provides unprecedented spatial resolution and the frequency band of GPS partly overlaps with that of broadband seismometers. Nevertheless, seismic data, especially near-field data is essential to resolve temporal change in detail. Although the inland earthquakes can be resolved quite well, they are just a small fraction of all earthquakes in the world. Most events occur beneath the ocean, where little data are available for slip inversion. The first large subduction earthquake whose near-field strong-motion seismograms were recorded is the 2003 Tokachi-Oki, Japan, earthquake (Mw 8.3; Hirata *et al.*, 2003). The records of cabled ocean bottom stations are quite expensive, but to increase such instruments is an important task for the study of seismic source.

There have been many slip models for more than 100 earthquakes, which tell us the image of fault slip and provide information about the physics of earthquakes behind them. However, there are only a small number of earthquakes whose rupture processes were resolved well enough for the discussion of fault dynamics. The Chi-Chi, Landers, Northridge, Kobe, Loma Prieta, Imperial Valley, and Tottori (**Table 1**) are those events. We need more case studies together with quality controls of the models. It is also important to develop analysis methods that enable broadband slip inversion that can use higher-frequency waves in complex structure.

Acknowledgments

The author thanks Prof. M. Hashimoto for the fault data of **Figure 1**, Prof. M. Matsu'ura and Dr. H. Aochi for useful suggestions. The fault models compiled by Drs. M. Mai and D. Wald were quite helpful for comparison between models. The catalog of slip models for Japanese earthquakes compiled by Profs. R. Sato and K. Koketsu was also useful. Some figures are prepared using generic mapping tool (Wessel and Smith, 1991). This work is partly supported by Grant-in-Aid for Scientific Research and DaiDaiToku Project, Ministry of Education, Sports, Science and Technology, Japan.

References

Abercrombie R and Mori J (1994) Local observations of the onset of a large earthquake: 28 June 1992 Landers, California. *Bulletin of the Seismological Society of America* 84: 725–734.

Akaike H (1980) Likelihood and Bayes procedure. In: Bernardo JM, DeGroot MH, Lindley DV, and Smith AFM (eds.) *Bayesian Statistics*, pp. 143–166. Valencia, Spain: University Press.

Aki K (1967) Scaling law of seismic spectrum. *Journal of Geophysical Research* 72: 1217–1231.

Aki K (1968) Seismic displacements near a fault. *Journal of Geophysical Research* 73: 5359–5376.

Aki K and Richards PG (2002) *Quantitative Seismology,* 2nd edn. Sausalito, CA: University Science Book.

Anderson JG (2003) Strong-motion seismology. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *Earthquake and Engineering Seismology, Part B*, pp. 937–965. London: Academic Press.

Andrews DJ (1976) Rupture velocity of plane strain shear cracks. *Journal of Geophysical Research* 81: 5679–5687.

Aochi H and Fukuyama E (2002) Three-dimensional nonplanar simulation of the 1992 Landers earthquake. *Journal of Geophysical Research* 107: doi:10.1029/2000JB000061.

Aochi H, Madariaga R, and Fukuyama E (2003) Constraint of fault parameters inferred from nonplanar fault modeling. *Geochemistry Geophysics Geosystems* 4: doi: 10.1029/2001GC000207.

Aoki Y, Segall P, Kato T, Cervelli P, and Shimada S (1999) Imaging magma transport during the 1997 seismic swarm off the Izu Peninsula, Japan. *Science* 286: 927–930.

Archuleta RJ (1984) A faulting model for the 1979 Imperial Valley earthquake. *Journal of Geophysical Research* 89: 4559–4585.

Benioff H, Press F, and Smith S (1961) Excitation of the free oscillations of earth by earthquakes. *Journal of Geophysical Research* 66: 605–619.

Beresnev IA (2003) Uncertainties in finite-fault slip inversions: To what extent to believe? (A critical review). *Bulletin of the Seismological Society of America* 93: 2445–2458.

Bernard P and Madariaga R (1984) A new asymptotic method for the modeling of near-field accelerograms. *Bulletin of the Seismological Society of America* 74: 539–557.

Beroza GC (1991) Near-source modeling of the Loma Prieta earthquake: Evidence for heterogeneous slip and implications for earthquake hazard. *Bulletin of the Seismological Society of America* 81: 1603–1621.

Beroza GC (1995) Seismic source modeling. *Reviews of Geophysics* 33(Supplement): 299–308.

Beroza GC and Mikumo T (1996) Short slip duration in dynamic rupture in the presence of heterogeneous fault properties. *Journal of Geophysical Research* 101: 22449–22460.

Beroza GC and Spudich P (1988) Linearized inversion for fault rupture behavior: Application to the 1984 Morgan Hill, California, earthquake. *Journal of Geophysical Research* 93: 6275–6296.

Boatwright J and Choy GL (1986) Teleseismic estimates of the energy radiated by shallow earthquakes. *Journal of Geophysical Research* 91: 2095–2112.

Bouchon M (1976) Teleseismic body wave radiation from a seismic source in a layered medium. *Geophysical Journal of the Royal Astronomical Society* 47: 515–530.

Bouchon M (1981) A simple method to calculate Green's functions for elastic layered media. *Bulletin of the Seismological Society of America* 71: 959–971.

Bouchon M (1997) The state of stress on some faults on the San Andreas system as inferred from near-field strong motion data. *Journal of Geophysical Research* 102: 11731–11744.

Bouchon M, Toksöz MN, Karabulut H, *et al.* (2002) Space and time evolution of rupture and faulting during the 1999 İzmit (Turkey) earthquake. *Bulletin of the Seismological Society of America* 92: 256–266.

Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research* 75: 4997–5009.

Chi WC, Dreger D, and Kaverina A (2001) Finite-source modeling of the 1999 Taiwan (Chi-Chi) earthquake derived from a dense strong-motion network. *Bulletin of the Seismological Society of America* 91: 1144–1157.

Cho I and Nakanishi I (2000) Investigation of the three-dimensional fault geometry ruptured by the 1995 Hyogo-Ken Nanbu earthquake using strong-motion and geodetic data. *Bulletin of the Seismological Society of America* 90: 450–467.

Cochard A and Madariaga R (1994) Dynamic faulting under rate-dependent friction. *Pure and Applied Geophysics* 142: 419–445.

Cohee BP and Beroza GC (1994) Slip distribution of the 1992 Landers earthquake and its implications for earthquake source mechanics. *Bulletin of the Seismological Society of America* 84: 692–712.

Cotton F and Campillo M (1995) Frequency domain inversion of strong motions: Application to the 1992 Landers earthquake. *Journal of Geophysical Research* 100: 3961–3975.

Courboulex F, Virieux J, Deschamps A, Gilbert D, and Zollo A (1996) Source investigation of a small event using empirical Green's functions and simulated annealing. *Geophysical Journal International* 125: 768–780.

Das S and Aki K (1977) A numerical study of two-dimensional spontaneous rupture propagation. *Geophysical Journal of the Royal Astronamical Society* 50: 643–668.

Das S and Kostrov BV (1990) Inversion for seismic slip rate history and distribution with stabilizing constraints: Application to the 1986 Andrean of Islands earthquake. *Journal of Geophysical Research* 95: 6899–6913.

Das S and Henry C (2003) Spatial relation between main earthquake slip and its aftershock distribution. *Reviews of Geophysics* 41, doi: 10.1029/2002RG000119.

Das S and Scholz CH (1983) Why large earthquakes do not nucleate at shallow depths. *Nature* 305: 621–623.

Day SM (1982a) Three dimensional simulation of spontaneous rupture: The effect of nonuniform prestress. *Bulletin of the Seismological Society of America* 72: 1881–1902.

Day SM (1982b) Three dimensional finite differential simulation of fault dynamics: Rectangular faults with fixed rupture velocity. *Bulletin of the Seismological Society of America* 72: 705–727.

Delouis B, Giardini D, Lundgren P, and Salichon J (2002) Joint inversion of InSAR, GPS, teleseismic, and strong-motion data for the spatial and temporal distribution of earthquake slip: Application to the 1999 İzmit Mainshock. *Bulletin of the Seismological Society of America* 92: 278–299.

Dziewonski AM and Woodhouse JH (1983) An experiment in systematic study of global seismicity: Centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981. *Journal of Geophysical Research* 88: 3247–3271.

Ellsworth WL and Beroza GC (1995) Seismic evidence for an earthquake nucleation phase. *Science* 268: 851–855.

Fletcher JB and Spudich P (1998) Rupture characteristics of the three M $\sim$ 4.7 (1992–1994) Parkfield earthquakes. *Journal of Geophysical Research* 103: 835–854.

Frankel A (2004) Rupture process of the M 7.9 Denali fault, Alaska, earthquake: Subevents, directivity, and scaling of high-frequency ground motions. *Bulletin of the Seismological Society of America* 94: S234–S255.

Fuchs K and Müller G (1971) Computation of synthetic seismograms with the reflectivity method and comparison with observations. *Geophysical Journal of the Royal Astronomical Society* 23: 417–433.

Fukahata Y, Nishitani A, and Matsu'ura M (2004) Geodetic data inversion using ABIC to estimate slip history during one earthquake cycle with viscoelastic slip-response functions. *Geophysical Journal International* 156: 140–153.

Fukuyama E and Irikura K (1986) Rupture process of the 1983 Japan Sea (Akita-Oki) earthquake using a waveform inversion method. *Bulletin of the Seismological Society of America* 76: 1623–1640.

Fukuyama E and Mikumo T (1993) Dynamic rupture analysis: Inversion for the source process of the 1990 Izu-Oshima, Japan, earthquake (M = 6.5). *Journal of Geophysical Research* 98: 6529–6542.

Fukuyama E, Mikumo T, and Olsen KB (2003) Estimation of the critical slip-weakening distance: Theoretical background. *Bulletin of the Seismological Society of America* 93: 1835–1840.

Geller RJ (1976) Scaling relations for earthquake source parameters and magnitudes. *Bulletin of the Seismological Society of America* 66: 1501–1523.

Graves RW and Wald DJ (2001) Resolution analysis of finite fault source inversion using one- and three-dimensional Green's functions. Part 1: Strong motions. *Journal of Geophysical Research* 106: 8785–8766.

Guatteri M and Spudich P (1998) Coseismic temporal changes of slip direction; the effect of absolute stress on dynamic rupture. *Bulletin of the Seismological Society of America* 88: 777–789.

Guatteri M and Spudich P (2000) What can strong-motion data tell us about slip-weakening fault-friction laws. *Bulletin of the Seismological Society of America* 90: 98–116.

Guatteri M, Spudich P, and Beroza GC (2001) Inferring rate and state friction parameters from a rupture model of the 1995 Hyogo-ken Nanbu (Kobe) Japan earthquake. *Journal of Geophysical Research* 106: 26511–26521.

Gülen L, Pinar A, Kalafat D, *et al.* (2002) Surface fault breaks, aftershock distribution and rupture process of the 17 August 1999 İzmit, Turkey, earthquake. *Bulletin of the Seismological Society of America* 92: 230–244.

Gutenberg B and Richter CF (1944) Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* 34: 185–188.

Hartzell SH (1978) Earthquake aftershocks as Green's functions. *Geophysical Research Letters* 5: 1–4.

Hartzell S (1989) Comparison of seismic waveform inversion results for the rupture history of a finite fault: Application to the 1986 North Palm Springs, California, earthquake. *Journal of Geophysical Research* 94: 7515–7534.

Hartzell SH and Heaton TH (1983) Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America* 73: 1553–1583.

Hartzell S and Langer C (1993) Importance of model parameterization in finite fault inversions: Application to the 1974 M_w

8.0 Peru earthquake. *Journal of Geophysical Research* 98: 22123–22134.
Hartzell S and Liu P (1995) Determination of earthquake source parameters using a hybrid global search algorithm. *Bulletin of the Seismological Society of America* 85: 516–524.
Hartzell S, Liu P, and Mendoza C (1996) The 1994 Northridge, California, earthquake: Investigation of rupture velocity, risetime, and high-frequency radiation. *Journal of Geophysical Research* 101: 20091–20108.
Hartzell SH, Stewart GS, and Mendoza C (1991) Comparison of L1 and L2 norms in a teleseismic waveform inversion for the slip history of the Loma Prieta, California, earthquake. *Bulletin of the Seismological Society of America* 81: 1518–1539.
Haskell NA (1964) Total energy and spectral density of elastic wave radiation from propagating faults. *Bulletin of the Seismological Society of America* 54: 1811–1841.
Haskell NA (1969) Elastic displacements in the near-field of a propagating fault. *Bulletin of the Seismological Society of America* 59: 865–908.
Hashimoto M (2005) How close to the Kobe Earthquake did seismology come? *Proceedings of the 2005 Joint Meeting for Earth and Planetary Science*, Chiba, Japan S094–001.
Heaton TH (1990) Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors* 64: 1–20.
Hellweg M and Boatwright J (1999) Mapping the rupture process of moderate earthquakes by inverting accelerograms. *Journal of Geophysical Research* 104: 7319–7328.
Hernandez B, Cotton F, and Campillo M (1999) Contribution of radar interferometry to a two-step inversion of the kinematic process of the 1992 Landers earthquake. *Journal of Geophysical Research* 104: 13083–13099.
Hirata K, Baba T, Sugioka H, *et al.* (2003) Ocean-bottom near-field observation to the 2003 Tokachi-oki earthquake with cabled seafloor observatory. *Earth Observing System Transactions, American Geophysical Union* 84: S52L- 04.
Horikawa H, Hirahara K, Umeda Y, Hashimoto M, and Kusano F (1996) Simultaneous inversion of geodetic and strong-motion data for the source process of the Hyogo-ken Nanbu, Japan, earthquake. *Journal of the Physics of the Earth* 44: 455–471.
Ichinose GA, Somerville P, Thio HK, Matsushima S, and Sato T (2005) Rupture process of the 1948 Fukui earthquake (M 7.1) from the joint inversion of seismic waveform and geodetic data. *Journal of Geophysical Research* 110, doi: 10.1029/2004JB003437.
Ichinose GA, Thio HK, Somerville PG, Sato T, and Ishii T (2003) Rupture process of the 1944 Tonankai earthquake (Ms 8.1) from the inversion of teleseismic and regional seismograms. *Journal of Geophysical Research* 108, doi: 10.1029/2003JB002393.
Ida Y (1972) Cohesive force across the tip of a longitudinal-shear crack and Griffith's specific surface energy. *Journal of Geophysical Research* 77: 3796–3805.
Ide S (1999) Source process of the 1997 Yamaguchi, Japan, earthquake analyzed in different frequency bands. *Geophysical Research Letters* 26: 1973–1976.
Ide S (2001) Complex source processes and the interaction of moderate earthquakes during the earthquake swarm in the Hida-Mountains, Japan, 1998. *Tectonophysics* 334: 35–54.
Ide S (2002) Estimation of radiated energy of finite-source earthquake models. *Bulletin of the Seismological Society of America* 92: 2294–3005.
Ide S (2003) Fracture surface energy of natural earthquakes from the viewpoint of seismic observations. *Bulletin of the Earthquake Research Institute* 78: 59–65.
Ide S and Aochi H (2005) Earthquake as multi-scale dynamic ruptures with heterogeneous fracture surface energy. *Journal of Geophysical Research* 110, doi: 10.1029/2004JB003591.
Ide S and Beroza GC (2001) Does apparent stress vary with earthquake size? *Geophysical Research Letters* 28: 3349–3352.
Ide S, Beroza GC, and McGuire JJ (2005) Imaging earthquake source complexity. In: In: Levander A and Nolet G (eds.) *Geophysical Monograph Series* 157, Seismic Earth: Array Analysis of Broadband Seismograms, pp. 117–135. Washington, DC: American Geophysical Union.
Ide S and Takeo M (1996) The dynamic rupture process of the 1993 Kushiro-oki earthquake. *Journal of Geophysical Research* 101: 5661–5675.
Ide S and Takeo M (1997) Determination of constitutive relations of fault slip based on seismic wave analysis. *Journal of Geophysical Research* 102: 27379–27391.
Ihmlé PF (1998) On the interpretation of subevents in teleseismic waveforms: The 1994 Bolivia deep earthquake revisited. *Journal of Geophysical Research* 103: 17919–17932.
Iio Y (1992) Slow initial phase of the P-wave velocity pulse generated by microearthquakes. *Geophysical Research Letters* 19: 477–480.
Iio Y (1995) Observations of the slow initial phase generated by microearthquakes - Implications for earthquake nucleation and propagation. *Journal of Geophysical Research* 100: 15333–15349.
Imanishi K and Takeo M (2002) An inversion method to analyze rupture processes of small earthquakes using stopping phases. *Journal of Geophysical Research* 107, doi: 10.1029/2001JB000201.
Imanishi K, Takeo M, and Ellsworth WL, *et al.* (2004) Source parameters and rupture velocities of microearthquakes in western Nagano, Japan, determined using stopping phases. *Bulletin of the Seismological Society of America* 94: 1762–1780.
Ishii M, Shearer PM, Houston H, and Vidale JE (2005) Extent, duration and speed of the 2004 Sumatra–Andaman earthquake imaged by the Hi-Net array. *Nature* 435: 933–936.
Iwata T (1991) Details of strong ground motions near earthquake foci and studies of source processes using these records. *Zisin 2* 44: 315–327 (in Japanese with English abstract).
Iwata T and Sekiguchi H (2002) Source model of the 2000 Tottori-ken Seibu earthquake and near-source strong ground motion. *Proceedings of the 11th Japan Earthquake Engineering Symposium*, 125–128 (in Japanese with English abstract).
Jackson DD (1979) The use of a priori data to resolve non-uniqueness in linear inversion. *Geophysical Journal of the Royal Astronomical Society* 57: 137–157.
Jackson DD and Matsu'ura M (1985) A Bayesian approach to nonlinear inversion. *Journal of Geophysical Research* 90: 581–591.
Ji C, Helmberger DV, Wald DJ, and Ma K-F (2003) Slip history and dynamic implications of the 1999 Chi-Chi, Taiwan. earthquake. *Journal of Geophysical Research* 108, doi: 10.1029/2002JB001764.
Ji C, Wald DJ, and Helmberger DV (2002a) Source description of the 1999 Hector Mine, California, earthquake. Part I: Wavelet domain inversion theory and resolution analysis. *Bulletin of the Seismological Society of America* 92: 1192–1207.
Ji C, Wald DJ, and Helmberger DV (2002b) Source description of the 1999 Hector Mine, California, earthquake. Part II: Complexity of slip history. *Bulletin of the Seismological Society of America* 92: 1208–1226.
Kakehi Y and Irikura K (1996) Estimation of high-frequency wave radiation areas on the fault plane by the envelope inversion

of acceleration seismograms. *Geophysical Journal International* 125: 892–900.
Kakehi Y and Irikura K (1997) High-frequency radiation process during earthquake faulting – envelope inversion of acceleration seismograms from the 1993 Hokkaido-Nansei-Oki, Japan, earthquake. *Bulletin of the Seismological Society of America* 87: 904–917.
Kakehi Y, Irikura K, and Hoshiba M (1996) Estimation of high-frequency wave radiation areas on the fault plane of the 1995 Hyogo-ken Nanbu earthquake by the envelope inversion of acceleration seismograms. *Journal of the Physics of the Earth* 44: 505–517.
Kanamori H (2004) The diversity of the physics of earthquakes. *Proceedings of the Japan Academy Series B* 80: 297–316.
Kanamori H and Anderson DL (1975) Theoretical basis of some empirical relations in seismology. *Bulletin of the Seismological Society of America* 65: 1073–1095.
Kanamori H and Brodsky EE (2001) The physics of earthquakes. *Physics Today* 54: 34–39.
Kanamori H and Stewart GS (1978) Seismological aspects of the Guatemala earthquake of February 4, 1976. *Journal of Geophysical Research* 83: 3427–3434.
Kennett BLN (2001) *The Seismic Wavefield Volume I: Introduction and Theoretical Development*. Cambridge, UK: Cambridge University Press.
Kennett BLN and Kerry NJ (1979) Seismic waves in a stratified half space. *Geophysical Journal of the Royal Astronomical Society* 57: 557–583.
Kikuchi M (1991) Complexity of earthquake source processes. *Zisin 2* 44: 301–314 (in Japanese with English abstract).
Kikuchi M and Kanamori H (1982) Inversion of complex body waves. *Bulletin of the Seismological Society of America* 72: 491–506.
Kikuchi M and Kanamori H (1986) Inversion of complex body waves – II. *Physics of the Earth and Planetary Interiors* 43: 205–222.
Kikuchi M and Kanamori H (1991) Inversion of complex body waves – III. *Bulletin of the Seismological Society of America* 81: 2335–2350.
Kikuchi M and Kanamori H (1994) The mechanism of the deep Bolivia earthquake of June 9, 1994. *Geophysical Research Letters* 21: 2341–2344.
Kikuchi M and Kanamori H (2004) Notes on teleseismic body-wave inversion program. http://www.eri.u-tokyo.ac.jp/ETAL/KIKUCHI/ (accessed Dec 2006).
Kikuchi M, Nakamura M, Yamada M, Fushimi M, Tatsumi Y, and Yoshikawa K (1999) Source parameters of the 1948 Fukui earthquake inferred from low-gain strong-motion records. *Zisin 2* 52: 121–128 (in Japanese with English abstract).
Kikuchi M, Nakamura M, and Yoshikawa K (2003) Source rupture process of the 1944 Tonankai earthquake and the 1945 Mikawa earthquake derived from low-gain seismograms. *Earth Planets and Space* 55: 159–172.
Kind R (1976) Computation of reflection coefficients for layered media. *Journal of Geophysics* 42: 191–200.
Kohketsu K (1985) The extended reflectivity method for synthetic near-field seismograms. *Journal of the Physics of the Earth* 33: 121–131.
Koketsu K, Yoshida S, and Higashihara H (1998) A fault model of the 1995 Kobe earthquake derived from the GPS data on the Akashi Kaikyo Bridge and other datasets. *Earth Planets and Space* 50: 803–811.
Kostrov VV (1974) Seismic moment and energy of earthquakes, and seismic flow of rock. *Izvestiya Earth Physics* 1: 23–40.
Krüger F and Ohrnberger M (2005) Tracking the rupture of the Mw = 9.3 Sumatra earthquake over 1,150 km at teleseismic distance. *Nature* 435: 937–939.
Langston CA and Helmberger DV (1975) A procedure for modelling shallow dislocation sources. *Geophysical Journal of the Royal Astronomical Society* 42: 117–130.
Lawson CL and Hanson RJ (1974) *Solving least squares problems*. Englewood Cliffs, NJ: Prentice-Hall.
Lee WHK (2002) Challenges in observational seismology. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *Earthquake and Engineering Seismology, Part A*, pp. 269–281. London: Academic Press.
Lee WHK, Shin TC, Kuo KW, Chen KC, and Wu CF (2001) CWB free-field strong-motion data from the 21 September Chi-Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America* 91: 1370–1376.
Liu P and Archuleta RJ (2004) A new nonlinear finite fault inversion with three-dimensional Green's functions: Application to the 1989 Loma Prieta, California, earthquake. *Journal of Geophysical Research* 109, doi: 10.1029/2003JB002625.
Ma K-F, Mori J, Lee S-J, and Yu SB (2001) Spatial and temporal distribution of slip for the 1999 Chi-Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America* 91: 1069–1087.
Madariaga R (1977) High-frequency radiation from crack (stress drop) models of earthquake faulting. *Geophysical Journal of the Royal Astronomical Society* 51: 625–651.
Mai PM and Beroza GC (2000) Source scaling properties from finite-fault-rupture models. *Bulletin of the Seismological Society of America* 90: 604–614.
Mai PM and Beroza GC (2002) A spatial random field model to characterize complexity in earthquake slip. *Journal of Geophysical Research* 107: doi: 10.1029/2001JB000588.
Mai PM, Spudich P, and Boatwright J (2005) Hypocenter locations in finite-source rupture models. *Bulletin of the Seismological Society of America* 95: 965–980.
Matsu'ura M (1991) Development of inversion theory in geophysics. *Zisin 2* 44: 53–62 (in Japanese with English abstract).
Matsu'ura M and Sato T (1997) Loading mechanism and scaling relations of large interplate earthquakes. *Tectonophysics* 277: 189–198.
McGuire JJ, Zhao L, and Jordan TH (2001) Teleseismic inversion for the second-degree moments of earthquake space-time distributions. *Geophysical Journal International* 145: 661–678.
McGuire JJ, Zhao L, and Jordan TH (2002) Predominance of unilateral rupture for a global catalog of large earthquakes. *Bulletin of the Seismological Society of America* 92: 3309–3317.
Mendez AJ and Anderson JG (1991) The temporal and spatial evolution of the 19 September 1985 Michoacan earthquake as inferred from near-source ground-motion records. *Bulletin of the Seismological Society of America* 81: 844–861.
Mendoza C and Hartzell SH (1988a) Inversion for slip distribution using teleseismic P waveforms: North Palm Springs, Borah Peak, and Michoacan earthquakes. *Bulletin of the Seismological Society of America* 78: 1092–1111.
Mendoza C and Hartzell SH (1988b) Aftershock patterns and main shock faulting. *Bulletin of the Seismological Society of America* 78: 1438–1449.
Menke W (1989) *Geophysical Data Analysis: Discrete Inverse Theory,* Revised Edition. San Diego, CA: Academic Press.
Mikumo T, Olsen KB, Fukuyama E, and Yagi Y (2003) Stress-breakdown time and slip-weakening distance inferred from slip-velocity functions on earthquake faults. *Bulletin of the Seismological Society of America* 93: 264–282.
Miyake H, Beroza GC, and Iwata T (2006) Frequency-dependent source processes for the 1989 Loma Prieta earthquake using a complex spectral inversion. *Journal of Geophysical Research* (submitted).

Miyatake T (1992) Dynamic rupture processes of inland earthquakes in Japan: Weak and strong asperities. *Geophysical Research Letters* 19: 1041–1044.

Miyazaki S, McGuire JJ, and Segall P (2003) A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array. *Journal of Geophysical Research* 108, doi: 10.1029/2001JB000456.

Miyazaki S, Segall P, McGuire JJ, Kato T, and Hatanaka Y (2006) Spatial and temporal evolution of stress and slip rate during the 2000 Tokai slow earthquake. *Journal of Geophysical Research* 111, doi: 10.1029/2004JB003426.

Mori J and Hartzell S (1990) Source inversion of the 1988 Upland, California, earthquake: Determination of a fault plane for a small event. *Bulletin of the Seismological Society of America* 80: 507–518.

Nakahara H, Nishimura T, Sato H, and Ohtake M (1998) Seismogram envelop inversion for the spatial distribution of high-frequency energy radiation from the earthquake fault: Application to the 1994 far east off Sanriku earthquake, Japan. *Journal of Geophysical Research* 103: 855–867.

Nakahara H, Sato H, Ohtake M, and Nishimura T (1999) Spatial distribution of high-frequency energy radiation on the fault of the 1995 Hyogo-Ken Nanbu, Japan, earthquake (Mw 6.9) on the basis of the seismogram envelope inversion. *Bulletin of the Seismological Society of America* 89: 22–35.

Nakayama W and Takeo M (1997) Slip history of the 1994 Sanriku-Haruka-Oki, Japan, earthquake deduced from strong-motion data. *Bulletin of the Seismological Society of America* 87: 918–931.

Oglesby DD, Archuleta RJ, and Nielsen SB (1998) Earthquakes on dipping faults: The effects of broken symmetry. *Science* 280: 1055–1059.

Oglesby DD and Day SM (2001) Fault geometry and the dynamics of the 1999 Chi-Chi (Taiwan) earthquake. *Bulletin of the Seismological Society of America* 91: 1099–1111.

Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America* 75: 1135–1154.

Okada T and Hasegawa A (2003) The M7.1 May 26, 2003 offshore Miyagi Prefecture earthquake in northeast Japan: Source process and aftershock distribution of an intra-slab event. *Earth Planets and Space* 55: 731–739.

Okada T, Umino N, Ito Y, Matsuzawa T, Hasegawa A, and Kamiyama M (2001) Source processes of 15 September 1998 M 5.0 Sendai, northeastern Japan, earthquake and its M 3.8 foreshock by waveform inversion. *Bulletin of the Seismological Society of America* 91: 1607–1618.

Okamoto T and Miyatake T (1989) Effects of near source seafloor topography on long-period teleseismic P waveforms. *Geophysical Research Letters* 16: 1309–1312.

Oliver J and Murphy L (1971) WWNSS: Seismology's global network of observing stations. *Science* 174: 254–261.

Olsen K, Madariaga R, and Archuleta R (1997) Three-dimensional dynamic simulation of the 1992 Landers earthquake. *Science* 278: 834–838.

Olson AH and Anderson JG (1988) Implications of frequency-domain inversion of earthquake ground motions for resolving the space-time dependence of slip on an extended fault. *Geophysical Journal* 94: 443–455.

Olson AH and Apsel RJ (1982) Finite faults and inverse theory with applications to the 1979 Imperial Valley earthquake. *Bulletin of the Seismological Society of America* 72: 1969–2001.

Peyrat S, Olsen K, and Madariaga R (2001) Dynamic modeling of the 1992 Landers earthquake. *Journal of Geophysical Research* 106: 26467–26482.

Press F, Ben-Menahem A, and Toksöz MN (1961) Experimental determination of earthquake fault length and rupture velocity. *Journal of Geophysical Research* 66: 3471–3485.

Quin H (1990) Dynamic stress drop and rupture dynamics of the October 15, 1979 Imperial Valley, California, earthquake. *Tectonophysics* 175: 93–117.

Rothman DH (1986) Automatic estimation of large residual statics corrections. *Geophysics* 51: 332–346.

Saikia CK (1994) Modified frequency-wavenumber algorithm for regional seismograms using Filon's quadrature: Modelling of Lg waves in eastern North America. *Geophysical Journal International* 118: 142–158.

Salichon J, Delouis B, Lundgren P, Giardini D, Costantini M, and Rosen P (2003) Joint inversion of broadband teleseismic and interferometric synthetic aperture radar (InSAR) data for the slip history of the Mw = 7.7, Nazca ridge (Peru) earthquake of 12 November 1996. *Journal of Geophysical Research*. 108, 10.1029/2001JB000913.

Salichon J, Lundgren P, Delouis B, and Giardini D (2004) Slip history of the 16 October 1999 Mw 7.1 Hector Mine earthquake (California) from the inversion of InSAR, GPS, and Teleseismic Data. *Bulletin of the Seismological Society of America* 94: 2015–2027.

Satake K (1989) Inversion of tsunami waveforms for the estimation of heterogeneous fault motion of large submarine earthquakes: The 1968 Tokachi-oki and 1983 Japan Sea earthquakes. *Journal of Geophysical Research* 94: 5627–5636.

Satake K (1993) Depth distribution of coseismic slip along the Nankai trough, Japan, from joint inversion of geodetic and tsunami data. *Journal of Geophysical Research* 98: 4553–4565.

Sato H, Nakahara H, and Ohtake M (1997) Synthesis of scattered energy density for nonspherical radiation from a point shear-dislocation source based on the radiative transfer theory. *Physics of the Earth and Planetary Interiors* 104: 1–13.

Scherbaum F (1996) *Of Poles and Zeros: Fundamentals of Digital Seismology*. Dordrecht: Kluwer.

Scholz CH (2002) *The Mechanics of Earthquakes and Faulting* 2nd edn. Cambridge, UK: Cambridge University Press.

Semmane F, Cotton F, and Campillo M (2005) The 2000 Tottori earthquake: A shallow earthquake with no surface rupture and slip properties controlled by depth. *Journal of Geophysical Research* 110, doi: 10.1029/2004JB003194.

Sekiguchi H, Irikura K, and Iwata T (1998) Detailed source process of the 1995 Hyogo-ken Nanbu (Kobe) earthquake using near-source strong ground motion data. *Proceedings. of the 10th Japan Earthquake Engineering Symposium* 1: 67–72.

Sekiguchi H, Irikura K, Iwata T, Kakehi Y, and Hoshiba M (1996) Minute locating of faulting beneath Kobe and the waveform inversion of the source process during the 1995 Hyogo-ken Nanbu, Japan, earthquake using strong ground motion records. *Journal of the Physics of the Earth* 44: 473–487.

Sekiguchi H and Iwata T (2002) Rupture process of the 1999 Kocaeli, Turkey, earthquake estimated from strong-motion waveforms. *Bulletin of the Seismological Society of America* 92: 300–311.

Shibazaki B, Yoshida Y, Nakamura M, Nakamura M, and Katao H (2002) Rupture nucleations in the 1995 Hyogo-ken Nanbu earthquake and its large aftershocks. *Geophysical Journal International* 149: 572–588.

Somerville P, Irikura K, Graves R, *et al.* (1999) Characterizing crustal earthquake slip models for the prediction of strong ground motion. *Seismological Research Letters* 70: 59–80.

Song S, Beroza GC, and Segall P (2005) Evidence for supershear rupture during the 1906 San Francisco earthquake. *Earth Observing System Transactions, American Geophysical Union* 86: S12A 05.

Spudich PKP (1992) On the inference of absolute stress levels from seismic radiation. *Tectonophysics* 211: 99–106.

Spudich P and Frazer LN (1984) Use of ray theory to calculate high-frequency radiation from earthquake sources having spatially variable rupture velocity and stress drop. *Bulletin of the Seismological Society of America* 74: 2061–2082.

Spudich P and Guatteri M (2004) The effect of bandwidth limitations on the inference of earthquake slip-weakening distance from seismograms. *Bulletin of the Seismological Society of America* 94: 2028–2036.

Steidl JH, Archuleta RJ, and Hartzell SH (1991) Rupture history of the 1989 Loma Prieta, California, earthquake. *Bulletin of the Seismological Society of America* 81: 1573–1602.

Takeo M (1985) Near-field synthetic seismograms taking into account the effects of anelasticity: The effects of anelastic attenuation on seismograms caused by a sedimentary layer. *Meteorology Geophysics* 36: 245–257 (in Japanese with English abstract).

Takeo M (1987) An inversion method to analyze the rupture processes of earthquakes using near-field seismograms. *Bulletin of the Seismological Society of America* 77: 490–513.

Tanioka Y, Ruff L, and Satake K (1996) The Sanriku-oki, Japan, earthquake of December 28, 1994 (Mw 7.7): Rupture of a different asperity from a previous earthquake. *Journal of Geophysical Research* 23: 1465–1468.

Tarantola A (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: Society of Industrial and Applied Mathematics.

Trifunac MD (1974) A three-dimensional dislocation model for the San Fernando, California, earthquake of February 9, 1971. *Bulletin of the Seismological Society of America* 64: 149–172.

Tsuboi S, Komatitsch D, Ji C, and Tromp J (2003) Broadband modeling of the 2002 Denali fault earthquake on the Earth Simulator. *Physics of the Earth and Planetary Interiors* 139: 305–312.

Umeda Y (1990) High-amplitude seismic waves radiated from the bright spot of an earthquake. *Tectonophysics* 175: 81–92.

Wald DJ (1996) Slip history of the 1995 Kobe, Japan, earthquake determined from strong motion, teleseismic, and geodetic data. *Journal of the Physics of the Earth* 44: 489–503.

Wald DJ and Graves RW (2001) Resolution analysis of finite fault source inversion using one- and three-dimensional Green's functions. Part 2: Combining seismic and geodetic data. *Journal of Geophysical Research* 106: 8767–8788.

Wald DJ and Heaton TH (1994) Spatial and temporal distribution of slip for the 1992 Landers, California earthquake. *Bulletin of the Seismological Society of America* 84: 668–691.

Wald DJ, Heaton TH, and Hudnut KW (1996) The slip history of the 1994 Northridge, California, earthquake determined from strong-motion, teleseismic, GPS, and leveling data. *Bulletin of the Seismological Society of America* 86: S49–S70.

Wald DJ, Helmberger DV, and Heaton TH (1991) Rupture model of the 1989 Loma Prieta earthquake from the inversion of strong-motion and broadband teleseismic data. *Bulletin of the Seismological Society of America* 81: 1540–1572.

Wald DJ, Kanamori H, Helmberger DV, and Heaton TH (1993) Source study of the 1906 San Francisco earthquake. *Bulletin of the Seismological Society of America* 83: 981–1019.

Wald D and Somerville PG (1995) Variable-slip rupture model of the Great 1923 Kanto, Japan, earthquake: Geodetic and body-waveform analysis. *Bulletin of the Seismological Society of America* 85: 159–177.

Ward SN and Barrientos SE (1986) An inversion for slip distribution and fault shape from geodetic observations of the 1983, Borah Peak, Idaho, earthquake. *Journal of Geophysical Research* 91: 4909–4919.

Wessel P and Smith WHF (1991) Free software helps map and display data. *Earth Observing System Transactions, American Geophysical Union* 72: 441.

Wiemer S and Katsumata K (1999) Spatial variability of seismicity parameters in aftershock zones. *Journal of Geophysical Research* 104: 13135–13151.

Wiens DA (1987) Effects of near source bathymetry on teleseismic P waveforms. *Geophysical Research Letters* 14: 761–764.

Wiens DA (1989) Bathymetric effects on body waveforms from shallow subduction zone earthquakes and application to seismic processes in the Kurile trench. *Journal of Geophysical Research* 94: 2955–2972.

Wu C and Takeo M (2004) An intermediate deep earthquake rupturing on a dip-bending fault: Waveform analysis of the 2003 Miyagi-ken Oki earthquake. *Geophysical Research Letters* 31, doi: 10.1029/2004GL021228.

Wu C, Takeo M, and Ide S (2001) Source process of the Chi-Chi earthquake: A joint inversion of strong motion data and Global Positioning System data with a multifault model. *Bulletin of the Seismological Society of America* 91: 1128–1143.

Yabuki T and Matsu'ura M (1992) Geodetic data inversion using a Bayesian information criterion for spatial distribution of fault slip. *Geophysical Journal International* 109: 363–375.

Yagi Y and Kikuchi M (2000) Source rupture process of Kocaeli, Turkey, earthquake of August 17, 1999, obtained by joint inversion of near-field data and teleseismic data. *Geophysical Research Letters* 27: 1969–1972.

Yagi Y, Kikuchi M, and Sagiya T (2001) Co-seismic slip, post-seismic slip, and aftershocks associated with two large earthquakes in 1996 in Hyuga-nada, Japan. *Earth Planets and Space* 53: 793–803.

Yamada T, Mori JJ, Ide S, Kawakata H, Iio Y, and Ogasawara H (2005) Radiation efficiency and apparent stress of small earthquakes in a South African gold mine. *Journal of Geophysical Research* 110, doi: 10.1029/2004JB003221.

Yao ZX and Harkrider DG (1983) A generalized reflection-transmission coefficient matrix and discrete wavenumber method for synthetic seismograms. *Bulletin of the Seismological Society of America* 73: 1685–1699.

Yoshida S (1986) A method of waveform inversion for earthquake rupture process. *Journal of the Physics of the Earth* 34: 235–255.

Yoshida S (1992) Waveform inversion for rupture process using a non-flat seafloor model: Application to 1986 Andreanof Islands and 1985 Chile earthquakes. *Tectonophysics* 211: 45–59.

Yoshida S (1995) Waveform inversion methods for the earthquake source. *Journal of the Physics of the Earth* 43: 183–209.

Yoshida S and Koketsu K (1990) Simultaneous inversion of waveform and geodetic data for the rupture process of the 1984 Naganoken-Seibu, Japan, earthquake. *Geophysical Journal International* 103: 355–362.

Yoshida S, Koketsu K, Shibazaki B, Sagiya T, Kato T, and Yoshida Y (1996) Joint inversion of near- and far-field waveforms and geodetic data for the rupture process of the 1995 Kobe Earthquake. *Journal of the Physics of the Earth* 44: 437–454.

Yu S-B, Kuo L-C, Hsu Y-J, *et al.* (2001) Preseismic deformation and coseismic displacements associated with the 1999 Chi-Chi, Taiwan, earthquake. 91: 995–1012.

Zeng Y, Aki K, and Teng TL (1993) Mapping of the high-frequency source radiation for the Loma Prieta earthquake, California. *Journal of Geophysical Research* 98: 11981–11993.

Zeng Y and Anderson JG (1996) A composite source model of the 1994 Northridge earthquake using genetic algorithms. *Bulletin of the Seismological Society of America* 86: S71–S83.

Zeng Y and Chen C-H (2001) Fault rupture process of the 20 September 1999 Chi-Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America* 91: 1088–1098.

Relevant Websites

http://www.consrv.ca.gov – California Strong Motion Instrumentation Program (CSMIP).

http://www.kik.bosai.go.jp – Digital Strong-Motion Seismograph Network.

http://www.iris.edu – Incorporated Research Institutions for Seismology (IRIS).

http://www.kyoshin.bosai.go.jp – Kyoshin Network, Japan (K-NET).

4.08 Fault Interaction, Earthquake Stress Changes, and the Evolution of Seismicity

G. C. P. King, Institute de Physique du Globe de Paris, Paris, France

4.08.1 Introduction

This chapter focuses on advances in our understanding of the Earth's lithosphere, over the last few decades. (In particular, the way in which studies of earthquakes have added to our knowledge, by shedding light on how earthquakes interact one with another). There is a surprising lack of consensus about how the lithosphere behaves. We know that it is inhomogeneous at scales ranging from microns to tens or hundreds of kilometers, is porous, and commonly anisotropic – remarks that apply with particular force to the continental lithosphere. Parts of the lithosphere are clearly brittle elastic and generate earthquakes, but the greater part of the lithosphere is not seismogenic with a mechanical behaviour that is poorly resolved. Even the brittle elastic part behaves strangely when compared with the engineering materials that we normally come across. For example, before half the failure stress is reached in a laboratory sample, it has dilated dramatically. This effect has not been demonstrated to occur at large scales, a disappointment to earthquake prediction attempts in the 1960s and 1970s, but serves to remind us that the behaviour of rocks at one scale may be a poor guide to their behaviour at another.

Commonly, rock sample studies may be only a metaphor for the behaviour at larger scales.

How does the nonseismogenic lithosphere behave? Visions vary on this. A persistent view has been, that only the upper 10–20 km of the continental lithosphere, the upper crust, is seismogenic and gives way to a lower crust that is viscous with no long-term strength. At greater depth, the Moho is reached and the upper mantle for perhaps 50 km is again strong.

This view is largely based on the results of laboratory tests conducted on homogeneous samples under very different strain rates, temperature conditions, and boundary conditions from those prevailing in the lithosphere. Nonetheless, the resulting model is mathematically and computationally tractable, and for geophysicists it has provided a satisfying explanation for a range of observations. Structural geologists have found the view difficult to accept. While not necessarily claiming that the entire lower crust and mantle is strong, a vision of highly localized shear zones (that only slip above a threshold stress) separated by strong, but deformable, blocks seems more appropriate to describe the structures observed in both active and deeply eroded mountain belts. Unlike the geophysicists view, this vision is, in general, only mathematically tractable when greatly simplified to a system of rigid blocks separated by narrow deformation zones. These models can be quite complex when describing the mechanics of fold-thrust belts (e.g. Suppe, 1985). The two views have been described respectively as 'dynamic' because they incorporate stresses while the latter are described as kinematic because only geometry (including displacement vectors) is included. In this chapter, only phenomena for which momentum terms in their underlying equations cannot be ignored are termed dynamic (seismic waves, dynamic fault rupture).

The two views have been to a large extent supported by the different data sets that the 'geophysicists' and 'structural geologists' customarily study, but both sides can find some discomfort in their simplifications. For example, recent GPS results, best interpreted in terms of blocks, are difficult to reconcile with earlier views that the geodetic data supported distributed deformation. Structural geologists can be perturbed by evidence for massive ductile deformation in the lower crust and upper mantle. Not all geoscientists view the same data; geochemists and hydrologists observed evidence for massive migration of water in the lithosphere and volcanologists have unequivocal evidence for the existence of massive regions of fluid rock, neither of which figure in the large-scale models outlined.

It was into this garden of confusion that the concepts of Coulomb stress interaction between earthquakes appeared in the 1980s. This met with astonishing success, while making the improbable simplification that the lithosphere is a homogeneous elastic half-space. Using this assumption, the stress on a second fault due to an earthquake on a first can be readily calculated. Section 4.08.2 begins by outlining this simple mathematical formulation for Coulomb stress interaction. Section 4.08.3 then provides examples of the method applied to both aftershock sequences and interactions between large events. Only aftershock sequences for stike-slip faults are illustrated since they offer the most straightforward examples. Section 4.08.4 starts to consider modifications of the simple early theory by the introduction of time-dependant failure through the concept of rate and state friction.

Readers who are already familiar with the basic concepts outlined in Sections 4.08.2–4.08.4 may wish to pass directly to later sections which consider unresolved problems and future perspectives in stress interaction studies. Section 4.08.5 considers the effects of long-term loading and lithospheric inhomogeneities together with the effects of fluids including the interactions between faults and volcanoes. Section 4.08.5 also describes why, in a medium subject to finite strain, the assumptions of infinitesimal strain or stress can never provide a complete description of fault interactions. Finally, the section discusses the accelerated seismic activity that appears over a wide region prior to major earthquakes. Section 4.08.6 touches on the question of dynamic triggering covered fully in Chapter 4.09, and Section 4.08.7 presents concluding remarks.

At the time of writing, there are at least 350 papers that can be considered relevant to stress interactions, too many to refer to in any detail. Particular topics are therefore illustrated using a suitable example with no attempt to reference other work on the same topic. For more complete reference lists, the reader is referred to reviews by Stein (1999), King and Cocco (2000), Freed (2005) and Steacy *et al.* (2005).

4.08.2 Stress Interactions between Faults in a Homogeneous Elastic Half-Space

Faults are widely assumed to be well approximated as very narrow planes across which some components of displacement are discontinuous while stress components across the plane are continuous. The system

is driven by stress boundary conditions which determine the nature of the faulting, strike-slip, normal or reverse. In turn the faults exhibit a frictional behavior that determines the stress level at which they will slip to generate an earthquake. As discussed later, each of these assumptions can be challenged, but they allow a simple theoretical and computational framework to be developed on the basis of dislocation theory where static displacements, strains, and stresses can computed by solving the elastostatic equation for a dislocation on an extended fault in an elastic, isotropic and homogeneous medium. This solution yields the Volterra equation (Volterra, 1907),

$$u_m(x_i) = \frac{1}{F}\int\int_{\Sigma} \Delta u_k(\xi_i)\, T^m_{k1}(\xi_i,\, x_i) n_i(\xi_i) \mathrm{d}\Sigma(\xi_i)$$

where F is the magnitude of a volume point force applied at ξ in the direction m. n_k are the components of the unit vector normal to the surface element $\mathrm{d}\Sigma$. The static displacement $u_m(x_i)$ is computed as a function of the dislocation $\Delta u_k(\xi_i)$ and the static traction $T^m_{kl}(\xi_i, x_i)$ on the fault plane Σ. The previous equation becomes

$$u_m(x_i) = \frac{1}{F}\int\int_{\Sigma} \Delta u_k \left[\lambda \delta_{kl} \frac{\partial \mathrm{U}^n_m}{\partial \xi_n} + \left(\frac{\partial \mathrm{U}^k_m}{\partial \xi_l} + \frac{\partial \mathrm{U}^l_m}{\partial \xi_k}\right)\right] n_l \,\mathrm{d}\Sigma$$

where the summation convention applies (e.g., Aki and Richards, 1980). The computation of the static displacement requires knowledge of the fault geometry, the slip distribution, and the strain nuclei (Mindlin, 1936) $U^k_m(\xi_i, x_i)$ (i.e., the static Green's functions). Using these solutions and Hooke's law analytical expressions for the static displacement, strain and stress fields caused by a finite rectangular fault either at the Earth surface or at depth have been derived (Okada, 1985).

4.08.2.1 Coulomb Failure

A number of criteria can be used to characterize failure in rocks of which the Coulomb failure criterion is widely used (e.g., Scholz, 1998). It requires that both the shear and normal stress on a preexisting or an incipient fault plane satisfy conditions analogous to those of friction on a preexisting surface.

Failure initiates and spreads on the plane when the Coulomb stress C_f, sometimes referred to as the Coulomb failure function (CFF), exceeds a specific value

$$C_\mathrm{f} = \tau_\beta + \mu(\sigma_\beta + \rho) \qquad [1]$$

where τ_β is the shear stress on the failure plane, σ_β is the normal stress (positive for extension), p is the pore fluid pressure, and μ the coefficient of friction. The value of τ_β must always be positive in this expression. However, the processes of resolving shear stress onto an assigned plane may give positive or negative values. The difference in sign of τ_β indicates whether the potential for slip on the plane is right or left lateral.

If the failure plane is orientated at β to the σ_1 axis (see **Figure 1**), we can express, under plane-stress conditions, the normal stress applied to the plane in terms of the principal stresses:

$$\sigma_\beta = \frac{1}{2}(\sigma_1 + \sigma_3) - \frac{1}{2}(\sigma_1 - \sigma_3)\cos 2\beta \qquad [2]$$

Two expressions are required for shear stress, with the one giving positive values being chosen. One is for left-lateral and the other for right-lateral shear stress.

$$\begin{aligned} \tau^L_\beta &= \frac{1}{2}(\sigma_1 - \sigma_3)\sin 2\beta \\ \tau^R_\beta &= -\frac{1}{2}(\sigma_1 - \sigma_3)\sin 2\beta \end{aligned} \qquad [3]$$

where σ_1 is the greatest principal stress and σ_3 is the least principal stress. For left- and right-lateral Coulomb stresses, eqn [1] then becomes

$$\begin{aligned} C_\mathrm{f} &= \frac{1}{2}(\sigma_1 - \sigma_3)(\sin 2\beta - \mu\cos 2\beta) \\ &\quad - \frac{1}{2}\mu(\sigma_1 + \sigma_3) + \mu p \end{aligned} \qquad [4]$$

where the sign of the shear stress (eqn [3]) is taken to indicate left- or right-lateral motion. By differentiating the last eqn [4] as a function of β, one finds that

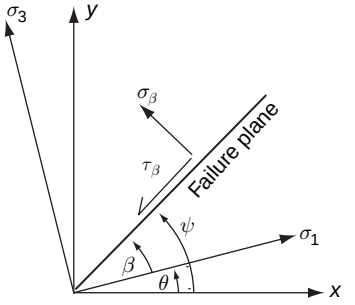


Figure 1 The axis system used for calculations of Coulomb stresses. Extension and left-lateral motion on the failure plane are assumed to be positive. The sign of τ_β is reversed for calculations of right-lateral Coulomb failure on specified planes.

the maximum Coulomb stress C_f^{max} occurs for angles when

$$\tan 2\beta = -\mu \quad [5]$$

Pore fluid pressure modifies the effective normal stress across the failure plane as shown in eqn [1]. When rock stress is changed more rapidly than fluid pressure can change through flow, pore fluid pressure, p, can be related to confining stress in the rock by Skemptons coefficient $B\left(p = -B\sigma_{kk}/3\right)$, where B varies between 0 and 1. Equation [1] and subsequent expressions (such as [4]) can therefore be rewritten on the assumption that σ_β represents the confining stress as well as the normal stress on the plane $\left(p = -B\sigma_\beta\right)$:

$$C_f = \tau_\beta + \mu'\sigma_\beta \quad [6]$$

where the effective coefficient of friction is defined by $\mu' = \mu(1-B)$. Although useful, it must be remembered that the expression is an approximation since the normal stress across a plane need not be so simply related to confining stress (e.g., Cocco and Rice, 2002). Further problems associated with fluid flow are discussed later.

The failure condition is inherently two-dimensional (2-D) with the intermediate stress σ_2 playing no part. Thus, the underlying physics can be illustrated in 2-D. To generalize the mathematics to 3-D it is only necessary to determine the orientation of the plane of greatest and least principal stresses in the appropriate coordinate system and to apply the failure conditions in that plane.

4.08.2.2 2-D Case: Coulomb Stress on a Plane of Specified Orientation

In a system where the x- and y-axes and fault displacements are horizontal, and fault planes are vertical (containing the z-direction), stress on a plane at an angle ψ from the x-axis (see **Figure 2**) resulting from a general stress field of any origin is given by

$$\begin{aligned}\sigma_{11} &= \sigma_{xx}\cos^2\psi + 2\sigma_{xy}\sin\psi\cos\psi + \sigma_{yy}\sin^2\psi \\ \sigma_{33} &= \sigma_{xx}\sin^2\psi - 2\sigma_{xy}\sin\psi\cos\psi + \sigma_{yy}\cos^2\psi \\ \tau_{13} &= \frac{1}{2}\left(\sigma_{yy}-\sigma_{xx}\right)\sin 2\psi + \tau_{xy}\cos 2\psi\end{aligned} \quad [7]$$

where these relations hold for a left-lateral mechanism $\left(\tau_{13} = \tau_{13}^L\right)$ and similar expressions can be derived for a right-lateral mechanism (as stated in [3]). The Coulomb stress for left-lateral C_f^L and right-lateral C_f^R motion on planes orientated at y with respect to the x-axis can now be written in the following manner:

$$C_f^L = \tau_{13}^L + \mu'\sigma_{33} \quad [8a]$$

$$C_f^R = \tau_{13}^R + \mu'\sigma_{33} \quad [8b]$$

Equation [8b] is illustrated in **Figure 2(a)** using a dislocation (earthquake fault) source. An elliptical slip distribution is imposed on the (master) fault in a uniform, stress-free, elastic medium. The contributions of the shear and normal components to the failure condition, and the resulting Coulomb stresses, for infinitesimal faults parallel to the master fault (commonly referred to as target faults) are shown in separate panels. Such a calculation represents the Coulomb stress on these planes resulting only from slip on the master fault (i.e., the figure shows a map of stress changes) and is independent of any knowledge of the prevailing regional stresses or any preexisting stress fields from other events. The signs in the calculation are chosen such that a positive Coulomb stress indicates a tendency for slip in the same right-lateral sense as the fault of interest. Negative Coulomb stresses indicate a reduction of this tendency. It is important to appreciate that because τ_{13} changes sign between eqns [8a] and [8b], a negative Coulomb stress for right-lateral fault motion is not the same as a tendency for left-lateral slip.

The distribution of increases and decreases of Coulomb stress shows features common to all subsequent figures. Lobes of increased shear stress appear at the fault ends, corresponding to the stress concentrations that tend to extend the fault. Off-fault lobes also appear, separated from the fault by a region where the Coulomb stresses have not been increased. If the master fault were infinitesimal in length and thus behaving as a point source, the off-fault lobes would be equal in amplitude to the fault-end lobes at all distances. The normal stress field is similar to the more familiar dilitational field with maxima and minima distributed antisymmetrically across the fault, but we consider only the component of tension normal to the fault. The influence of the normal stress on the Coulomb stress distribution is to reduce the symmetry of the final distribution and to increase the tendency for off-fault failure.

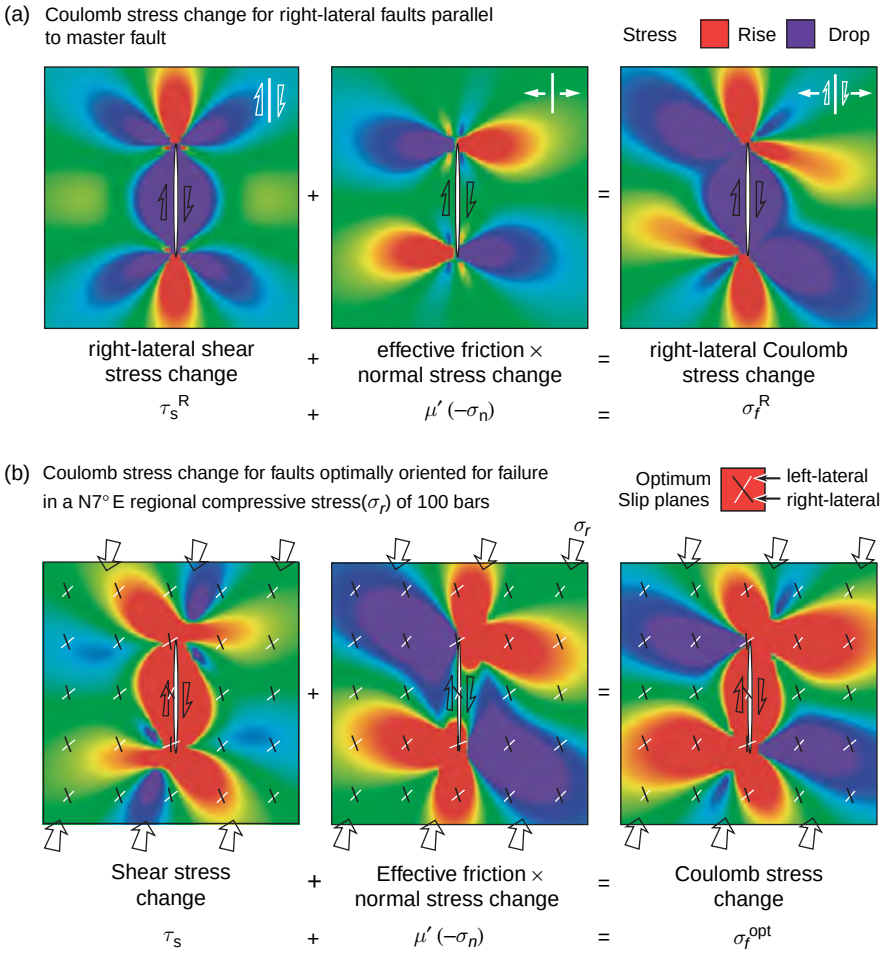


Figure 2 Coulomb stress changes caused by a right lateral strike-slip fault in plane strain. (a) Coseismic shear, normal, and Coulomb stress changes for right-lateral fault parallel to the master fault. (b) Shear, normal, and Coulomb stress changes on optimally oriented faults (black and white segments) with respect to a regional stress characterized by an N7° E (open arrows) compression of 100 b. The slip distribution on the fault plane is uniform, but it tapers at the fault ends. The effective friction coefficient μ' is taken to be 0.4. Redrawn from King GCP Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

4.08.2.3 2-D Case: Change of Coulomb Stress on Optimally Orientated Faults

The calculation in the previous section is general and need not be due to an earthquake. If a stress or stress change is known, then the Coulomb stress or stress change can be calculated on any plane. The more general case where a change of stress, such as that due to an earthquake, is imposed on a preexisting stress field can be calculated. The failure planes are not specified *a priori*, but are computed accounting for the interaction between the two fields. Optimal planes are regarded as those simultaneously favored by tectonic loading and the stress induced by the motion on the master fault. Here we define such planes in terms of uniformly varying stress fields. The optimum directions are determined by the stress change due to that earthquake $\sigma_{ij}^q \left(=\Delta\sigma_{ij}\right)$ and by existing regional stresses σ_{ij}^r which give a total stress σ_{ij}^t:

$$\sigma_{ij}^t = \sigma_{ij}^r + \sigma_{ij}^q \qquad [9]$$

The orientation of the principal axes resulting from the total stress are therefore derived using

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\sigma_{xy}^t}{\sigma_{xx}^t - \sigma_{yy}^t}\right) \qquad [10]$$

where θ is the orientation of one principal axis to the x-axis as shown in **Figure 1** and the other is at $\theta \pm 90°$. From these two directions, the angle of

greatest compression θ_1 must be chosen. The optimum failure angle ψ_o is then given by $\theta_1 \pm \beta$ where one plane is left and the other right lateral, with the sign of shear stress being chosen appropriately. Whereas the optimum planes are determined from σ^t_{ij}, the normal and shear stress changes on these planes are determined only by the earthquake stress changes σ^q_{ij}. Thus the changes in stress on the optimum planes become

$$\sigma_{33} = \sigma^q_{xx} \sin^2 \psi_\circ - 2\sigma^q_{xy} \sin \psi_\circ \cos \psi_\circ + \sigma^q_{yy} \cos^2 \psi_\circ$$
$$\tau_{13} = \frac{1}{2}\left(\sigma^q_{yy} - \sigma^q_{xx}\right) \sin 2\psi_\circ + \tau^q_{xy} \cos 2\psi_\circ \qquad [11]$$

and the Coulomb stress changes on the planes are equal and given by

$$C_f^{opt} = \tau_{13} + \mu' \sigma_{33} \qquad [12]$$

The two optimum planes correspond to left-lateral and right-lateral shear. The Coulomb stress change is the same on both, so that the Coulomb stress is the same whether calculated for left- or right-lateral planes in expression [12]. It is important to emphasize that we calculate the change of Coulomb stress on planes that are optimum after the earthquake with the optimum orientations being calculated from the earthquake stress field plus the regional stress field. The Coulomb stress changes caused by the earthquake stress changes are then resolved onto these planes. Where earthquake stresses are large, they can rotate the principal axes.

The results of a calculation to find optimum orientations and magnitudes of Coulomb stress changes are shown in **Figure 2(b)**. The slip on the master fault is the same as before (**Figure 2(a)**). A uniform 100 bar compressional stress is introduced with the orientation shown. White lines indicate optimum left-lateral orientations and black lines, right-lateral orientations. The shear and normal stress contributions to the Coulomb stress change are again shown in separate panels.

It can be seen from expression [9] that only the deviatoric part of the regional stress determines the orientation of principal axes, and hence the optimum stress orientations. Thus it is sufficient to apply the regional stress as a simple uniaxial compression or extension. This assumes, however, that the intermediate principal stress is vertical and thus a 2-D description is complete. In general, this is not true, since the relative magnitudes of the principal stresses control the focal mechanisms which need not be strike-slip.

The relative amplitude of the regional stress to the earthquake stress drop ($\Delta\tau$) might be expected to have an effect. This is explored for the strike-slip case in **Figure 3**, which shows the Coulomb stress change C_f^{opt} on optimally orientated right-lateral planes in which the regional field is equal to the stress drop $\Delta\tau$ (left panel) and 10 times $\Delta\tau$ (right panel). These examples span likely conditions. It is evident that, except close to the master fault, the orientations of the optimal planes and Coulomb stress changes on these planes are little altered. The optimal orientations are essentially fixed by the regional stress, except very close to the fault, where the stress change caused by slip on the master fault is comparable to the regional stress. Thus, for stress drops similar in amplitude to the regional stress, Coulomb stress changes close to the fault are always positive. (For a regional stress of zero, they are positive everywhere.)

The effects of varying the orientation of regional stresses and changing the coefficient of friction μ' are shown in **Figure 4**. Possible changes of regional stress orientation are limited since the main fault must move as a result of the regional stress; the 30° range covers the likely range. Similarly, values of friction between 0.0 and 0.75 span the range of plausible values. All of the panels show the same general features, fault-end and off-fault Coulomb stress lobes. Thus, modeling is most sensitive to the regional stress direction, almost insensitive to the regional stress amplitude, and modestly sensitive to the coefficient of effective friction. In other words, while the relative magnitude of the regional stress with respect to the earthquake stress drop controls the changes in the orientations of the optimal planes close to the master fault (as shown in **Figure 4**), the orientation of the regional stress is the most important factor controlling the orientations of the optimal planes for failure far away from the master fault.

4.08.2.4 3-D Case: Strike-Slip and Dip-Slip Conditions

The application to vertical strike-slip faults is easy because a 2-D problem (plane stress configuration) can be solved with the vertical components of the regional stress tensor neglected. However, the application to dip-slip faults requires the solution of a 3-D problem where the ratio between vertical and horizontal components of regional stress tensor must be known. This means that, while in the foregoing discussion we have assumed that only strike-slip vertical faults are present and thus stress components σ_{zz}, σ_{xz}, σ_{yz} could be neglected, in a more

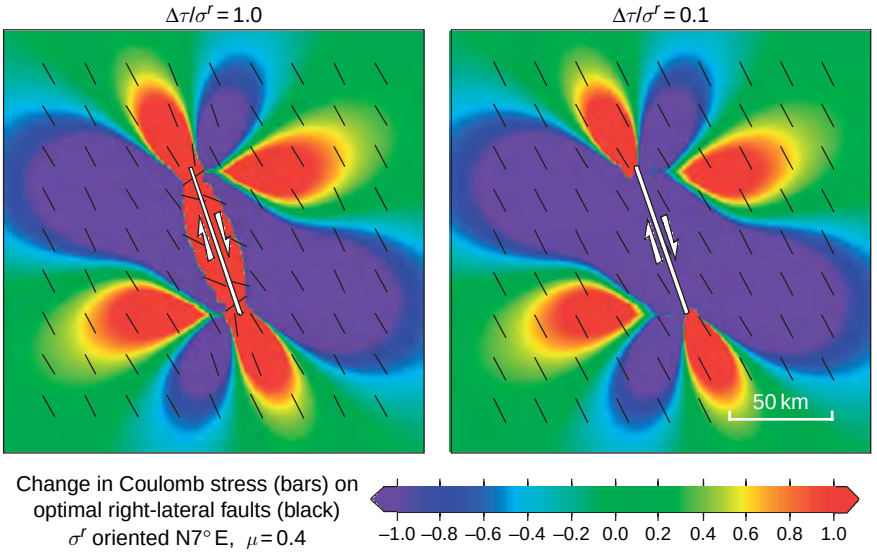


Figure 3 Coulomb stress changes for different magnitudes of the regional stress. The left panel shows the Coulomb stress changes for an earthquake stress drop similar in amplitude to the deviatoric regional stress, while the right panel shows the same calculations for a larger (10 times) regional stress amplitude. The former case represents the stress perturbation caused by an earthquake relieving all of the regional stress. In this case, the resulting optimally oriented planes (black segments) for failure rotates from the direction predicted by the regional deviatoric stress near the causative fault, where the induced coseismic stress is larger. Far away from the fault (left panel), where the regional stress amplitude is larger than the induced one, the optimally oriented planes follow the regional stress (see also the right panel). Redrawn from King GCP Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

general configuration these stress components cannot be ignored.

If the regional stress components are known, the total stress can be computed through eqn [9], which is used to find the orientation and magnitude of the principal stresses. We can therefore calculate the orientation of the plane containing σ_1 and σ_3 and hence the optimum orientations of slip planes where the change of Coulomb failure stress will be found. The calculation is straightforward and an example is shown in **Figure 5**. If we take horizontal stresses similar in amplitude to those employed before, we can examine the change of mechanism as the vertical stress is changed. The horizontal stresses are chosen to be 200 b (EW) and 400 b (NS). For **Figure 5(a)**, the vertical stress is 100 b, in **Figure 6(b)** the vertical stress is 300 b, and in **Figure 6(c)** the vertical stress is 500 b. The predominant mechanisms are reverse faulting in **Figure 6(a)**, strike-slip faulting in **Figure 5(b)**, and normal faulting in **Figure 5(c)**. If the vertical stresses are attributed to overburden pressure, then the figures correspond to depths of 300, 900, and 1500 m, respectively. This is clearly incorrect. Events do not occur at such shallow depths nor are systematic changes of focal mechanism with depth observed. It would seem that a confining pressure that increases with depth must be added to the horizontal stress such that the differential stresses remain similar. It is clear, however, that the mechanisms are sensitive to small variations in the stresses chosen; thus, to correctly predict mechanisms from stresses, the stress regime must be known with precision. Direct measurements of stress throughout the seismogenic zone are never available, thus predicting mechanisms from stress must be indirect. In practice, the most useful stress information comes from some sort of average of the stress orientation required to generate the observed focal mechanisms in the region. As discussed in Section 4.08.3.8, calculation of stress changes on individual fault planes determined from focal mechanisms might be preferred.

4.08.2.5 Sensitivity to the Main-Shock Focal Mechanism

The last section showed that the Coulomb distribution can be sensitive to incorrect assumptions about

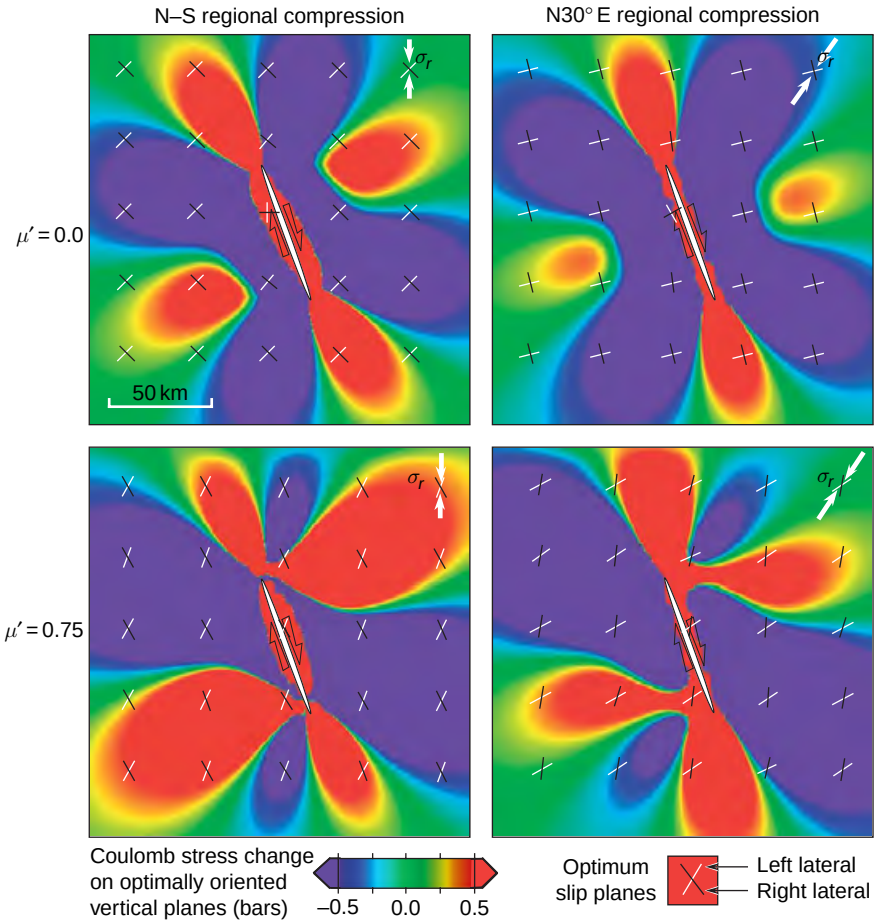


Figure 4 Coulomb stress changes for different orientations of the regional deviatoric stress (left and right panels) and for different values of the effective friction coefficient (upper and bottom panels). Both configurations of the regional stress field refer to horizontal s_1 and s_3 directions. Stress changes are caused by a 70-km-long and 12.5-km-wide vertical right-lateral fault with 5 m of tapered slip. The regional stress amplitude is 100 b as in **Figure 3**. The regional stress orientation controls the size of the off-fault to fault-end lobes. The effective friction coefficient controls the internal angle between right- and left-lateral slip planes. Redrawn from King GCP Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

regional stress, since this can cause target fault focal mechanisms to be incorrectly predicted. While nearby events are clearly sensitive to all of the details of the rupture process, this is not the case at some distance from the fault. This is particularly important for studying fault interaction between large-magnitude earthquakes. Completely incorrect focal mechanisms clearly result in incorrect results. Errors in dip (within an acceptable range) are much less serious. **Figure 6(a)** shows the distribution for a 45° dipping normal fault, and **Figure 6(b)** shows the distribution for a vertical dike with the same opening displacement as the horizontal slip vector of the dip-slip fault. Except at distances comparable to the source dimensions the two distributions are identical. This illustrates that, except close to a dip-slip fault, the fault dip is not important. In general, except in the near field (or at distances of few fault lengths), a double couple point source with the appropriate seismic moment is all that is needed.

While large errors in the strike of the focal planes are serious, small errors only result in (an approximately) commensurate rotation of the Coulomb stress distribution. This is shown in **Figure 6(c)**. It can be remarked at this point that it is much harder to evaluate the fault parameters close to dip-slip faults, making aftershock studies of such events harder than for strike-slip events. We therefore choose to illustrate aftershock studies for strike-slip events and large event interactions for mainly dip-slip events.

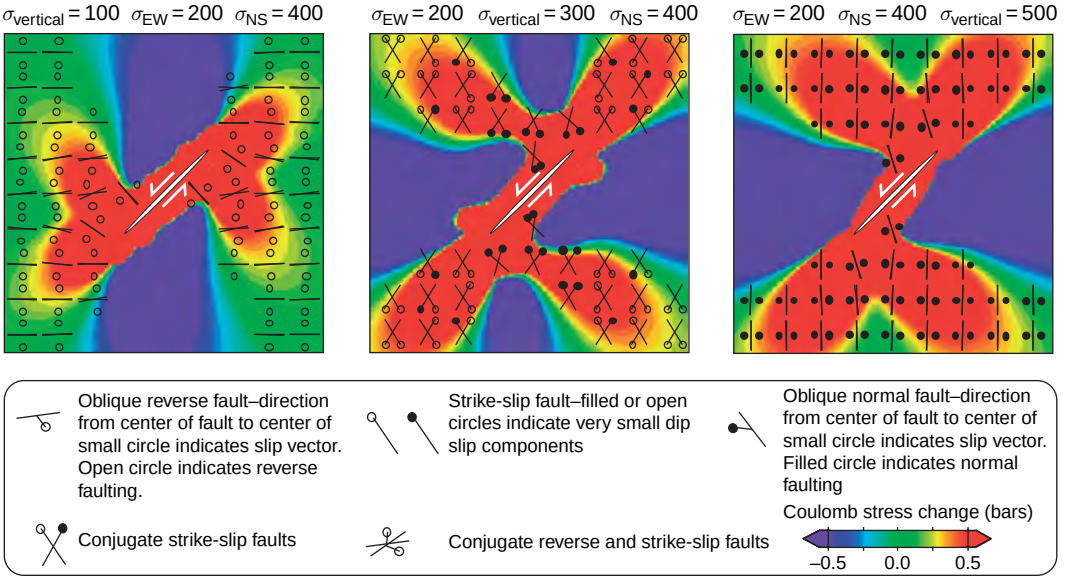


Figure 5 Coulomb stress changes and optimally slip planes for different orientations of the regional stress field. The vertical stress coincides with the least principal stress (100 b) in the left panel, with the intermediate principal stress (300 b) in the middle panel, and with the greatest principal stress (500 b) in the right panel. Therefore, the predicted faulting mechanisms on optimally oriented planes are strike-slip in the middle panel, and inverse or normal in the left and right panels, respectively. Reproduced from King GCP and Cocco M (2000) Fault interaction by elastic stress changes: New clues from earthquake sequences. *Advances in Geophysics* 44: 1–38, with permission from Elsevier.

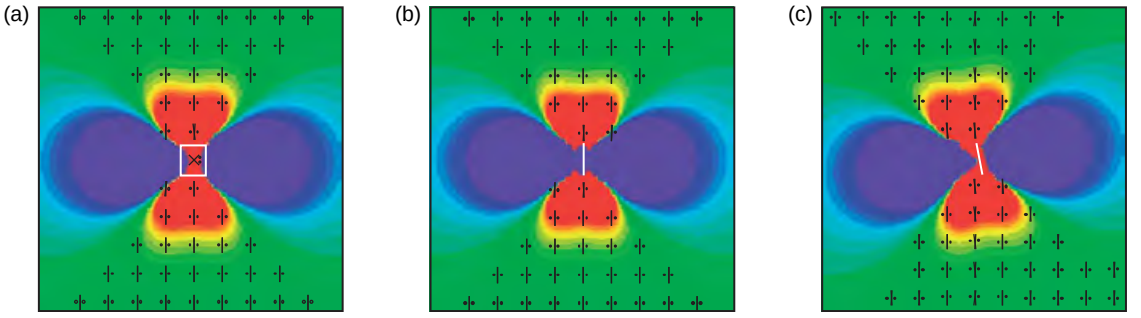


Figure 6 Coulomb stress changes caused by a 45° dipping normal fault (a), a vertical dike with the same opening displacement as the horizontal slip vector of the dip slip fault (b), and a 45° dipping normal fault with a different strike (c), Except at distances comparable to the source dimensions, the two distributions are identical, suggesting that, except close to a dip-slip fault, the fault dip is less important than fault strike; in this latter case, errors only result in an approximately commensurate rotation of the Coulomb stress distribution. Symbols as in **Figure 6**. Reproduced from King GCP and Cocco M (2000) Fault interaction by elastic stress changes: New clues from earthquake sequences. *Advances in Geophysics* 44: 1–38, with permission from Elsevier.

4.08.3 Examples of Coulomb Interactions

4.08.3.1 Coulomb Stress Changes and Aftershocks

The methods outlined above can be applied to aftershock distributions. **Figure 7** shows the 1979 Homestead Valley aftershock sequence. There are four lobes of Coulomb stress rise and four lobes of Coulomb stress drop. The lobes at the ends of the fault extend into the fault zone, while the off-fault lobes are separated from the fault over most of its length by a zone where the Coulomb stress is reduced. The distributions of aftershocks are consistent with these patterns. Many events are associated with increases of Coulomb stress of less than 1 b, while reductions of the same amount apparently suppress them. Relatively few events fall in the regions

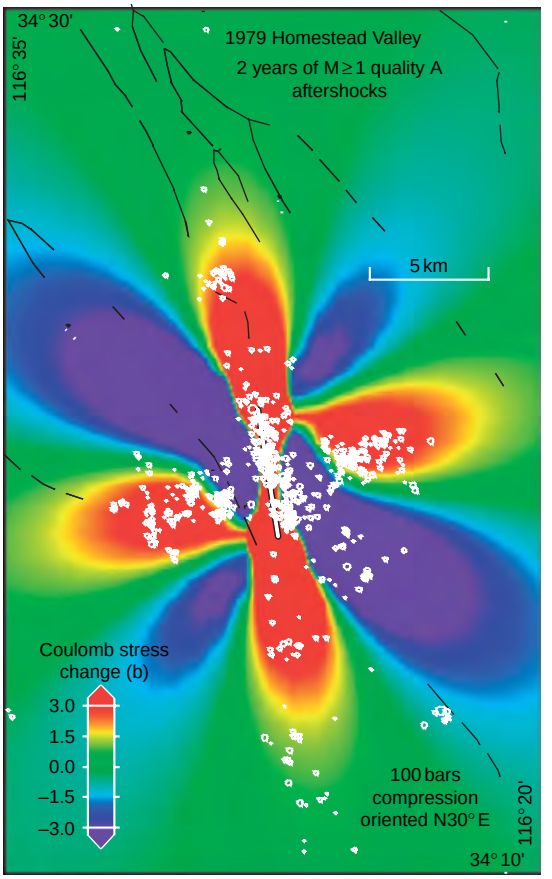


Figure 7 Coulomb stress changes caused by the strike-slip 15 Mar. 1979 Homestead Valley earthquake sequence and aftershock distribution (the same shown in **Figure 1**). The calculations are carried out in a half-space with the values of Coulomb stress plotted in the figures being calculated at half the depth to which the faults extend. The fault is 5.5 km long and 6 km wide and slip is 0.5 m and it is tapered at the fault ends. Redrawn from King GCP Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

of lowered Coulomb stress, and the clusters of off-fault aftershocks are separated from the fault itself by a region of diminished activity. The distributions of Coulomb stresses can be modified as described earlier by adjusting the regional stress direction and changing μ'. However, any improvements in the correlation between stress changes and aftershock occurrence are modest. Consequently, we have chosen to show examples with an average μ' of 0.4. Whatever values we adopt, we find that the best correlations of Coulomb stress change to aftershock distribution are at distances greater than a few kilometers from the fault. Closer to the fault, unknown details of fault geometry and slip distribution influence the stress changes. Correlation between aftershock distribution and Coulomb stress changes on a vertical cross section can also be observed and are discussed by King *et al.* (1994b).

4.08.3.2 Stress Changes Associated with the Landers Earthquake

Although stress interactions had been previously observed for other events, the clear correlations associated with the Landers earthquake (M_w 7.3) suggested to the scientific community that Coulomb calculations might indeed prove an effective method of relating large events with each other and relating large events to their aftershock sequences. The Landers Earthquake and associated events remain the best example for illustrating the techniques, and we reproduce here the modeling processes following King *et al.* (1994b). Since the main events were strike-slip (except for Cedar Mountain), the calculations are essentially two-dimensional and an 'effective' regional stress is easy to establish since only the deviatoric part in the horizontal plane is needed.

4.08.3.3 Coulomb Stress Changes Preceding the Landers Rupture

Figure 8 shows the Coulomb stress changes caused by the four $M > 5$ earthquakes within 50 km of the epicenter that preceded the Landers Earthquake. The Coulomb stresses are due to the 1975 $M_L = 5.2$ Galway Lake, 1979 $M_L = 5.2$ Homestead Valley, 1986 $M_L = 6.0$ North Palm Springs, and 1992 $M_L = 6.1$ Joshua Tree earthquakes. These progressively increased Coulomb stresses by 1 b at the future Landers epicenter. Together, they also produced a narrow zone of Coulomb stress increase of 0.7–1.0 b, which the future 70-km-long Landers Rupture followed for 70% of its length. The Landers fault is also nearly optimally oriented for most of its length. It is noteworthy that the three largest events are roughly equidistant from the future Landers epicenter: the right-lateral Homestead Valley and Joshua Tree events enhanced stress as a result of the lobes beyond the ends of their ruptures, whereas the North Palm Springs event enhanced rupture as a result of an off-fault lobe. Increasing the effective friction μ' from 0.4 (used in **Figure 8**) to 0.75 slightly increases the Coulomb stress rise on the future Landers fault plane while dropping the friction to zero slightly reduces it.

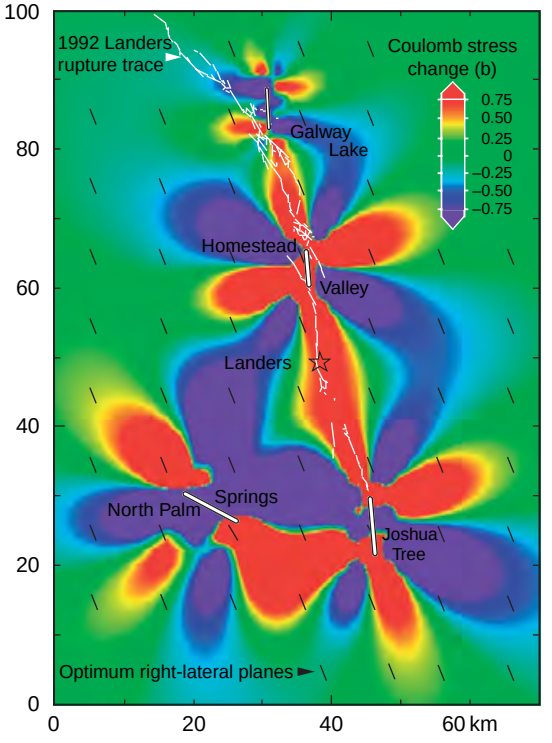


Figure 8 Coulomb stress changes caused by the four earthquakes that preceded the 1992 Landers main shock within 50 km of the epicenter. The source parameters used in this calculation are listed in King *et al*. (1994b). These events raised the stress at the future Landers epicenter and on most (70%) of the Landers rupture. Redrawn from King GCP Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

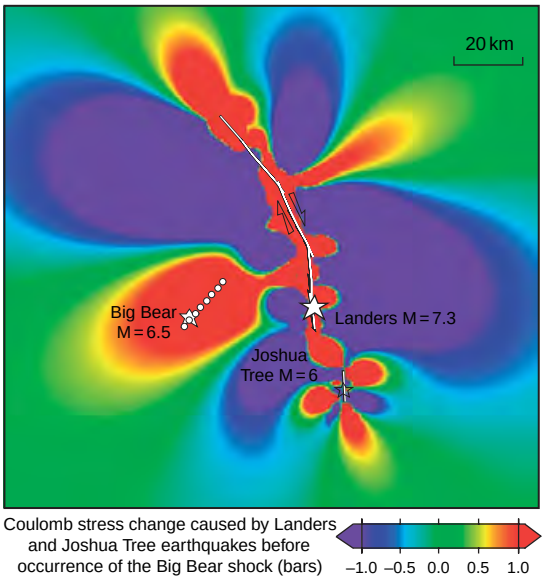


Figure 9 Coulomb stress changes caused by the 1992 Joshua Tree and Landers earthquakes. The location of the Big Bear aftershock and its faulting mechanism are consistent with the induced coseismic stress field. Redrawn from King GCP Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

4.08.3.4 Stress Changes Following the Landers Rupture but before the Big Bear Earthquake

Unlike the earthquake sources modeled so far, which are approximated by an elliptical slip distribution on a single plane, there is more information about the $M = 7.4$ Landers source and in the following the rupture is modeled with 13 fault segments to produce a slip distribution that is consistent with surface fault mapping, geodetic data, radar data, and the modeling of seismic data.

The stress changes caused by the Landers event are shown in **Figure 9**. The largest lobe of increased Coulomb stress is centered on the epicenter of the future $M_L = 6.5$ Big Bear event, where stresses were raised 2–3 b. The Big Bear earthquake was apparently initiated by this stress rise 3 h 26 min after the Landers main shock. The Coulomb stress change at the epicenter is greatest for high effective friction but remains more than 1.5 b for $\mu = 0$. There is no surface rupture or Quaternary fault trace associated with the Big Bear earthquake. Judging from its epicenter and focal mechanism, the plane that apparently ruptured was optimally aligned for left-lateral failure, with the rupture apparently propagating northeast and terminating where the Landers stress change became negative.

In addition to calculating the stress changes caused by the Landers rupture, an estimate can be made of the slip on the Big Bear fault needed to relieve the shear stress imposed by the Landers rupture. This is achieved by allowing the Big Bear rupture plane to slip freely as in the stress field caused by the Landers rupture. The potential slip along the Big Bear fault is 60 mm (left-lateral), about 5–10% of the estimated slip that occurred. These calculations suggest that the Big Bear slip needed to relieve the stress imposed by Landers was a significant fraction of the total slip that later occurred. Thus, from consideration of the stress changes and the kinematic response to those changes, it has been considered reasonable to propose that stresses from the Landers event played a major role in triggering the Big Bear shock.

4.08.3.5 Stress Changes Caused by the Landers, Big Bear, and Joshua Tree Ruptures

The Big Bear earthquake was the largest of more than 20 000 aftershocks located after the Landers earthquake, large enough to result in significant stress redistribution at the southwestern part of the Landers rupture zone. Consequently, the distribution of later events cannot be examined without considering its effect. Although smaller, a similar argument can be applied to the Joshua Tree event, whose aftershock sequence was not complete at the time of the Landers rupture. In **Figure 10**, the combined Coulomb stress changes for the Joshua Tree, Landers, and Big Bear earthquakes are plotted. This distribution is shown together with all well-located $M_L \geq 1$ events that occurred during the following 25 days.

Most $M_L > 1$ aftershocks occur in regions where the failure stress is calculated to have increased by $\geq$0.1 b, and few events are found where the stress is predicted to have dropped (**Figure 10**). Even when all seismicity within 5 km of the Landers, Big Bear, and Joshua Tree faults is excluded, more than 75% of the aftershocks occur where the stress is predicted to have risen by >0.3 b. In contrast, less than 25% of the aftershocks occur where the stress dropped by >0.3 b.

The Landers earthquake sequence suggests that a series of smaller events that preceded the Landers earthquake prepared a region of slightly elevated stress along much of the fault that the main event then followed (**Figure 8**). It is also clear that the

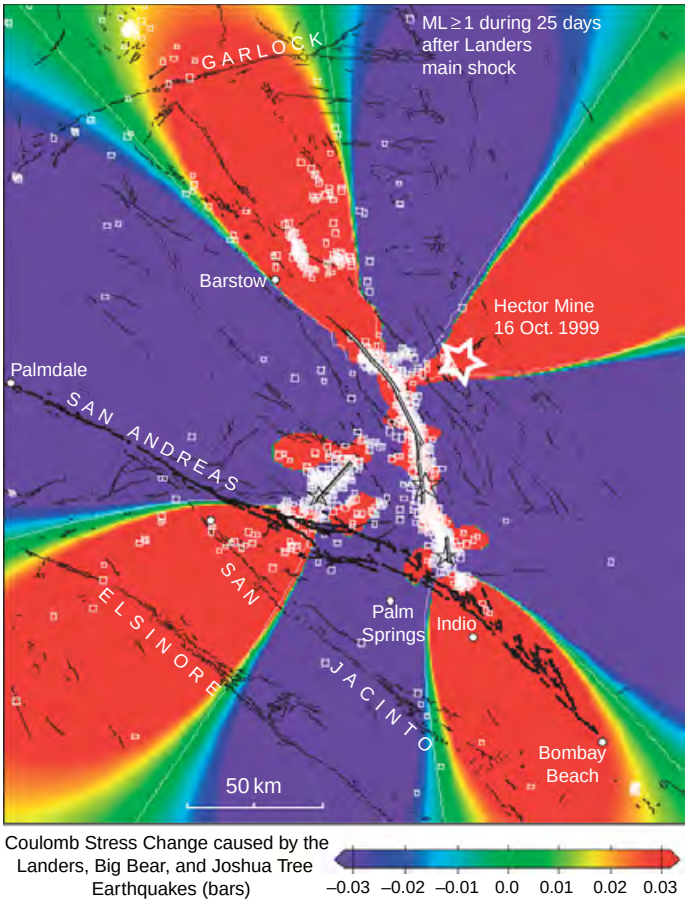


Figure 10 Coulomb stress changes caused by the Joshua Tree, Landers, and Big Bear earthquakes mapped at a depth of 6.25 km (halfway down the fault) and ($M \geq 1$) aftershock distribution during the 25 days following the Landers main shock. Source parameters are those used by King et al. (1994b). The location (open star) and the surface rupture (solid black line) of the M 7.1, 1999, Hector Mine earthquake epicenter is shown on the map. The Coulomb stress changes are plotted with a reduced amplitude (color) scale with respect to those of the previous figures. Redrawn from King GCP, Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

stresses created by the main event largely controlled the aftershock distribution, including the location of the large Big Bear aftershock.

4.08.3.6 Interactions between Large Earthquakes: Western Turkey and the Aegean

A number of studies have shown that large events appear to interact over large areas (see reviews quoted in the introduction for more detailed reference lists). Here we illustrate the effect taking an example from Turkey and the western Aegean. Since 1912, 29 events ($M_s \geq 6.0$) have occurred in the region shown in **Figure 11**. The area is of particular interest because the events do not lie along a single fault and the mechanisms vary and interaction results from both on- and off-fault Coulomb stress changes. Thus, unlike the sequence further east along the North Anatolian Fault (Stein *et al.*, 1997), the interactions cannot be attributed to slip along a deep extension of a single fault. The events are predominantly strike-slip and normal faulting, but some reverse faulting occurs. The later events have reliable focal mechanisms and seismic moments controlled by waveform modeling. Earlier events do not, but a combination of geological information to identify the faults involved and damage information closely controls the possible mechanisms. Details of the mechanism information can be found in Nalbant *et al.* (1998), from which the examples presented here are abstracted; this paper splits the time period into nine intervals such that all interactions can be seen. Here we only show three example time windows. They are 1912–44 (**Figure 12(a)**), 1912–67 (**Figure 12(b)**), and 1912–96 (**Figure 12(c)**). These are insufficient to demonstrate all the interactions which require the original figures to be examined, but allow the reader to gain an impression of the size of the events and the distances between them. The results, however, can be summarized. Out of 29 events, 16 (1935, 1939, 1944.1, 1953, 1957, 1964, 1967.2, 1968, 1969, 1970, 1975, 1978, 1981.1, 1981.2, 1982, 1983) occurred in regions where stress was increased by more than 0.1 b. If less certain information about earlier events is used and smaller stress increases are presumed to be significant, then 23 of the 29 events appear to have responded to Coulomb stress interactions. Stein *et al.* (1997) examined interactions between 10 events extending to the east of the region shown in **Figure 10(c)**. Of these, all but one (8 out of 10) occurred in regions of substantially enhanced Coulomb stress. For the East Anatolian Fault, Nalbant *et al.* (2002) (**Figure 10(c)**) have shown that 9 out of 10 events ($M > 6.7$) occurred in regions of enhanced Coulomb stress. Taken together with the Nalbant *et al.* (1998) work, 33 out of 49 events since 1912 along the North and East Anatolian Faults and the Aegean system show clear Coulomb interactions. If less certain information is included, this rises to 41 out of 49. Of great importance, not one substantial event occurred in regions where Coulomb stress was reduced. Such regions have been dubbed stress shadows (Harris, 1998); these mauve regions cover a substantial fraction of the area of the map in **Figure 12(c)** and, with some reservations discussed later, can be classified as unlikely to experience an earthquake. Interaction studies of this sort are always limited by a lack of knowledge of events that preceded the time period studied; in few cases, a study can extend over a greater time period than the return periods of the earthquakes. An exception is the study discussed in Section 4.08.5.1.

The 1999 Izmit Earthquake occurred in one of the four regions (see figure caption) where known active faults and increased Coulomb stress coincided. When the Izmit earthquake struck, Coulomb stress on faults in the eastern Marmara Sea increased by between 1 and 5 b as well as by about 1 bar to the east of the Izmit rupture zone (Hubert-Ferrari *et al.*, 2000). On 12 November 1999, a second M 7.2 earthquake (Duzce) occurred in this eastern region (see **Figure 11**). It is important to note that this region was in a stress shadow prior to the Izmit earthquake which changed its status. Other examples of this sort are discussed by Nalbant *et al.* (1998). Thus, while stress shadows are in general less likely to host future earthquakes, their status can change rapidly if an event occurs in an adjacent region of high Coulomb stress. The present situation in the Sea of Marmara region is discussed in Section 4.08.5.1, where the effects of long-term loading are also included. A ~100 km stretch of fault south of Istanbul can now be identified as being likely to host a future earthquake.

The success of the Izmit and Duzce predictions together with the prediction of an M 8.7 event following the 26 December 2004 M 9.3 great Sumatra earthquake (McCloskey *et al.*, 2005) provide three examples of Coulomb methods forecasting future events and form an important part of the evidence establishing regions of enhanced Coulomb stress in places hosting known active faults as a powerful approach to establishing medium-term seismic hazard for a region.

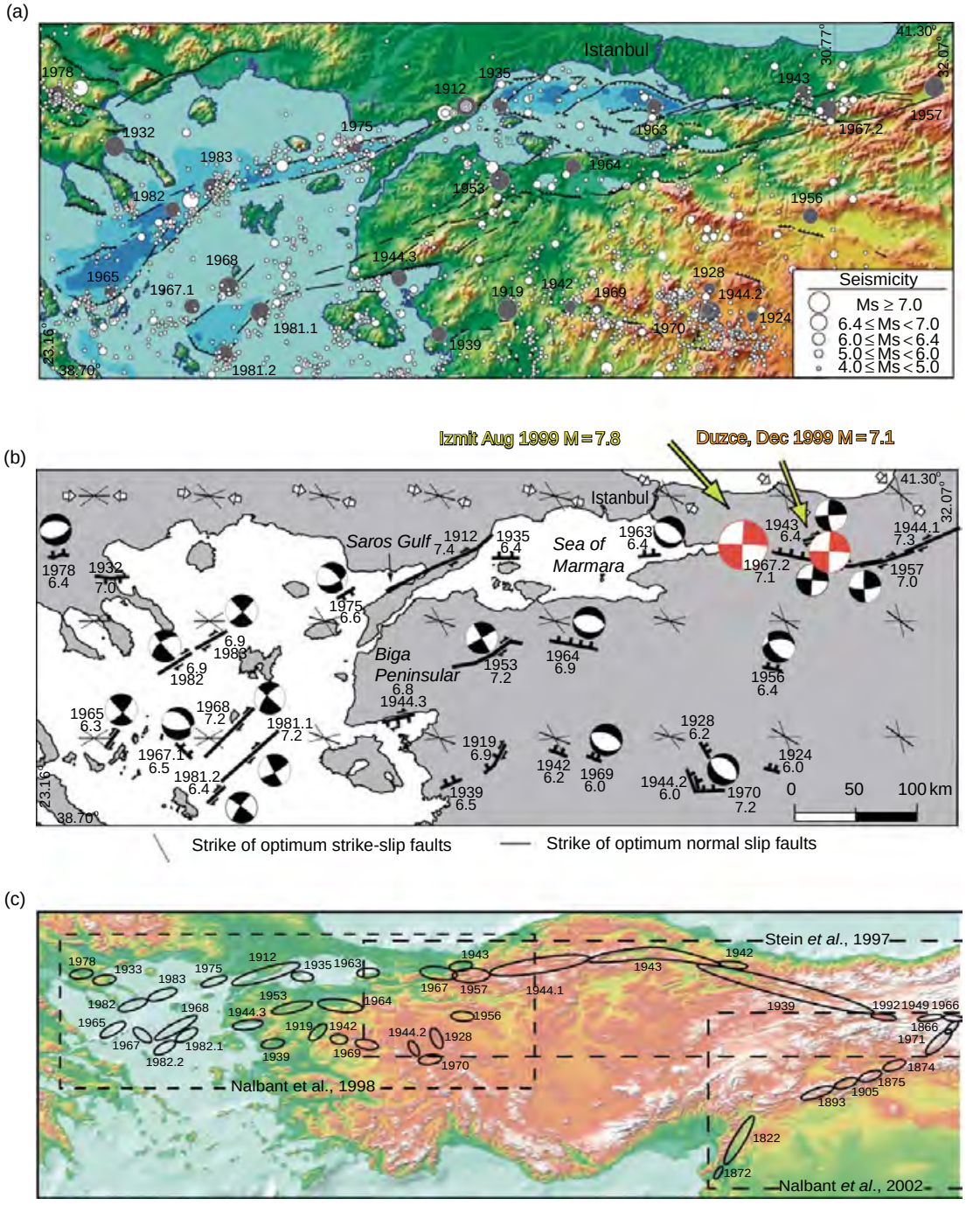


Figure 11 (a) Tectonic map of western Turkey and North Aegean Sea and distribution of seismicity between 1900 and 1996. (b) Fault plane solutions and active faults in the area. The orienation of the regional stress field used in the study is shown. Short solid lines indicate the orientation of the principal compressive stress. This is also the strike of optimally orientated normal faulting. The directions (two conjugate) for optimum strike-slip faulting are shown by lighter lines. (c) The regions and events studied by Nalbant *et al*. (1998), Stein *et al*. (1997), and Nalbant *et al*. (2002). (a and b) Reproduced from Nalbant SS, Hubert A and King GCP (1998) Stress coupling between earthquakes in northwest Turkey and the North Aegean Sea. *Journal of Geophysical Research* 103: 24,469–24,486, with permission from American Geophysical Union.

4.08.3.7 Interactions between Very Large Earthquakes

Most of the earthquakes discussed above had lengths (L) that were not much greater than their down-dip width (W). As a consequence, an earthquake on one segment of fault could significantly modify the stress on all of a nearby segment. The distance to which the stress field extends, however, depends on the smallest fault dimension; thus, when the L-to-W ratio becomes very large, the stress change affects only a small part of an adjacent fault segment at most (**Figure 13**), and thus the simple stress triggering described earlier is less likely to occur. Only if the epicenter where the next event initiates is close to the boundaries of slip of the previous earthquake is triggering unambiguous. For the North Anatolian Fault sequence, six events had L-to-W ratios greater than 10 and initiated close to the termination of a previous rupture. One event in 1943 initiated 230 km from the end of the rupture termination of the 1939 event and has not been regarded as being triggered by static stress changes (Stein *et al.*, 1997). Because of the reduced role of stress interaction for very large events, other effects, including loading stresses (Section 4.08.5.1), seem to become more important for massive earthquakes.

4.08.3.8 Is It Better to Calculate Stress Transfer to Known Faults?

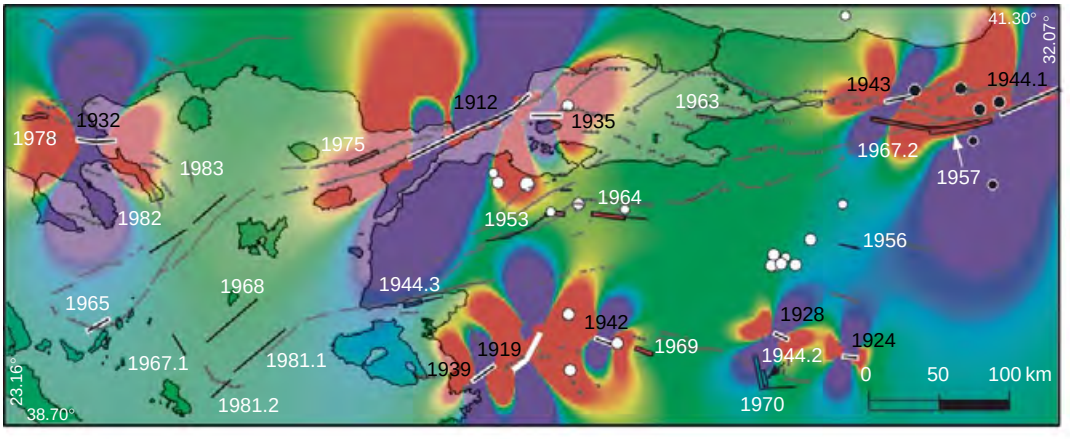
Most of the early studies of both aftershocks and stress interactions between large events used the formulation of section 4.08.2.2 to calculate optimum fault orientations based on a regional stress field. Where the fault planes of large events were known, this was commonly backed up by calculation of Coulomb stress change on the known fault. Some recent work has abandoned looking at changes on optimum planes considering it to be better to resolve the stress onto a target fault plane (Steacy *et al.*, 2005). This is correct if an averaged fault surface represents the part of the fault where triggering occurs. However, faults are inherently complex, en echelon and multistranded on many scales, and rupture may initiate on local complexity. Thus, it is unknown whether the stress change resolved on the fault plane as defined by its average surface trace or focal mechanism is more appropriate than the optimum direction. It should be recalled that stress directions are derived from averages of fault directions and focal mechanisms for a region and hence may represent better those parts of a fault where slip can be triggered.

For small events and aftershocks, very large numbers of events can be examined if the optimum Coulomb condition is assumed providing a much larger data set than events of known mechanisms. This could be seen as an advantage. However, since a certain knowledge of the stress system is never available, most authors have assumed that more than one mechanism is possible. An overall compressional or extensional regional stress field may allow some events to be strike-slip and not just prevailing dip-slip mechanisms. As a result, the maximum Coulomb stress for the two possible fields is chosen. For example, the largest Coulomb stress increase is for either strike-slip or reverse fault events. This enlarges the Coulomb lobes; thus, although some extra events can be captured, better correlations may not be real.

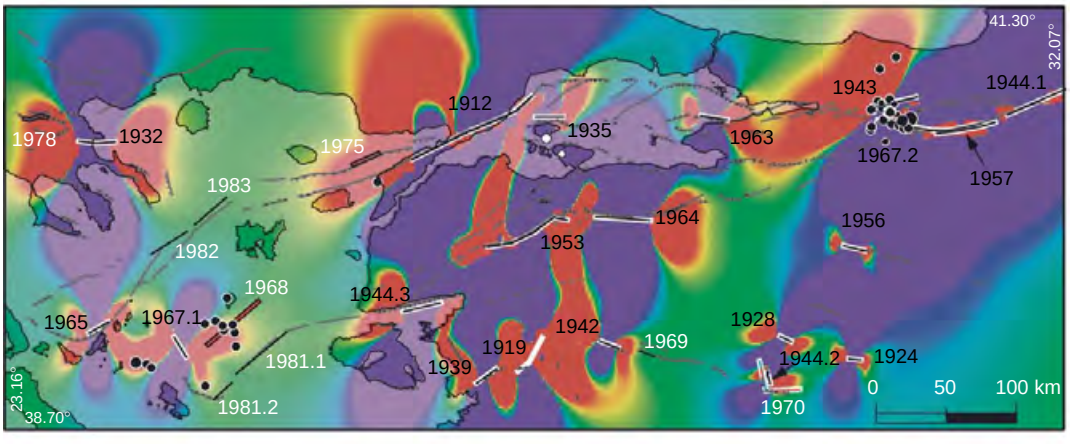
Some of the errors associated with the study of aftershocks may be the failure to account for the stress changes resulting from the aftershocks themselves. Thus, the stress field from the main shock alone is inadequate to represent the stress field as an aftershock sequence evolves. The aftershock sequence can release a substantial fraction of the magnitude of the main event (Marsden and Nalbant, 2005). Clearly, the stress changes for the Big Bear aftershock could not be ignored when examing the seismicity following the Landers earthquake. However, most aftershocks are not localized in small clusters, and as the event sizes reduce, their area of influence diminishes rapidly. Consequently, except for some of the larger aftershocks, this should do no more than blur the boundaries of the main-event Coulomb distribution. It certainly cannot account for events that clearly fall in regions where Coulomb stress is reduced by the main event.

As noted earlier for a number of cases, aftershock focal mechanisms are available and a number of workers have carried out direct calulations and examined each of the two possible nodal planes and carried out various statistical tests. Perhaps surprisingly, the results have not proved to be much better than the earlier studies (see references in Steacy *et al.* (2005)). In such tests, some aftershock sequences have worked quite well such as Landers, Superstition Hills, Kobe, while others, notably Loma Prieta, have not. Later discussion may provide some of the reasons for the lack of improvement and why studies of events such as Loma Prieta have not proved successful.

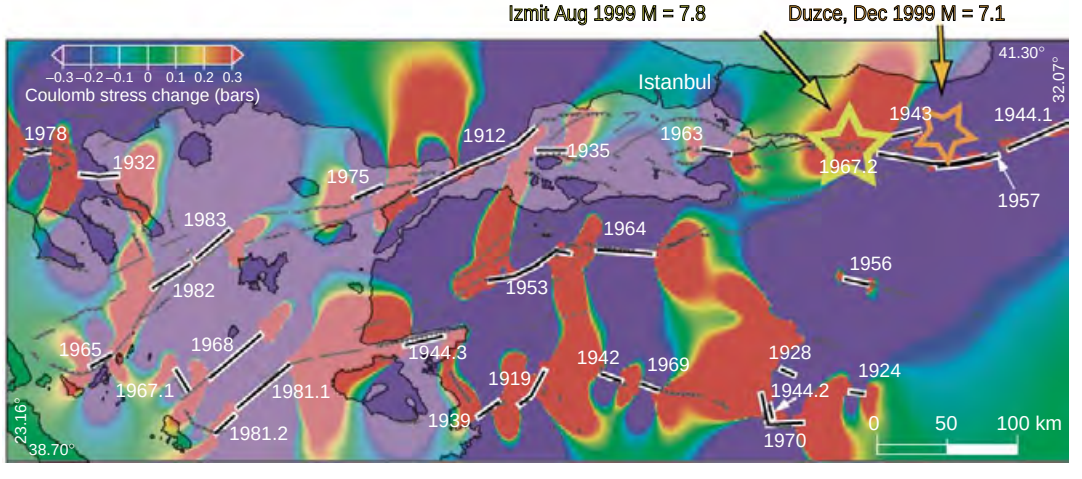
(a)


41.30°
32.07°
1912
1935
1963
1943
1944.1
1932
1978
1975
1967.2
1983
1957
1964
1982
1953
1956
1968
1944.3
1965
1928
1942
1924
1967.1
1981.1
1919
1969
1944.2
1939
0 50 100 km
23.16°
38.70°
1981.2
1970

(b)


41.30°
32.07°
1912
1943
1944.1
1978
1932
1935
1963
1975
1967.2
1983
1957
1964
1956
1982
1953
1968
1944.3
1965
1967.1
1928
1942
1919
1969
1981.1
1944.2
1924
1939
0 50 100 km
23.16°
38.70°
1981.2
1970

(c)


Izmit Aug 1999 M = 7.8
Duzce, Dec 1999 M = 7.1
–0.3 –0.2 –0.1 0 0.1 0.2 0.3
Coulomb stress change (bars)
Istanbul
41.30°
32.07°
1978
1912
1943
1944.1
1932
1935
1963
1975
1967.2
1957
1983
1964
1956
1982
1953
1965
1968
1944.3
1928
1942
1969
1924
1967.1
1981.1
1919
1944.2
1939
0 50 100 km
23.16°
38.70°
1981.2
1970

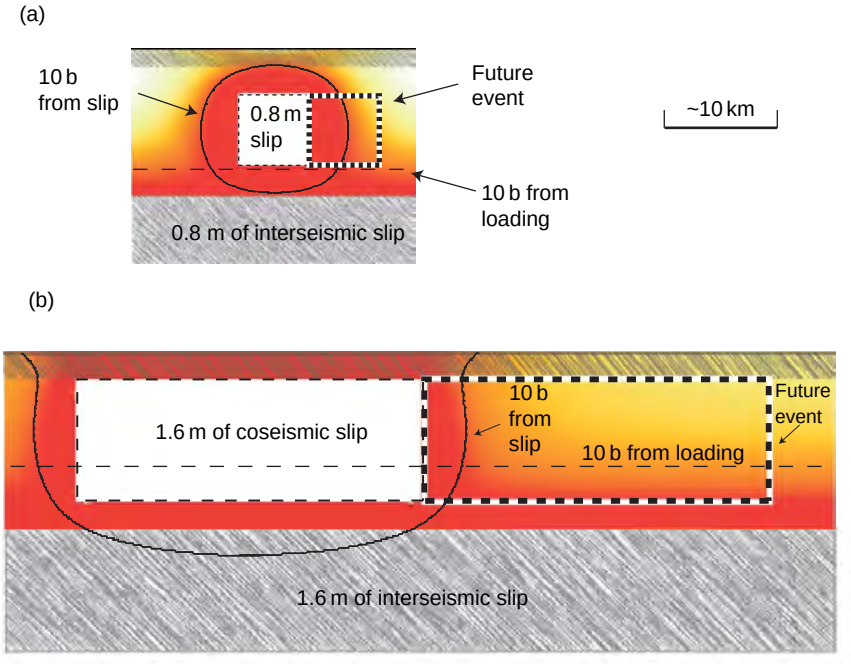


Figure 13 The relative contributions of long-term loading and seismic slip for events with *L*-to-*W* ratios of 1 and 4 for (a) and (b), respectively. The solid contour in each case indicates 10 b of increase due to the earthquake and the dashed line 10 b due to loading. Hatching indicates regions that slip aseismically. See also figure 6 of Lin and Stein (2004).

4.08.4 Introducing Time into the Failure Criterion

The results discussed so far result from the assumption that earthquakes can be represented by dislocation surfaces, that faults obey a Coulomb friction failure condition, that the lithosphere can be treated as a homogeneous, elastic half-space, and that the loading system consists of a simple stress field. The examples provided, plus many others in the literature, support the notion that stress changes due to motion on one fault can result in motion on a second. Despite examples demonstrating that other effects must also intervene, stress interaction must be an important element in any attempt to explain the most prominent feature of global earthquake statistics. Earthquake pairs are clustered in space and time with the occurrence of one event increasing the probability of a second, with the probability decaying with distance and time from the first event (Kagan and Jackson, 1991). In the development so far, time has been absent and will now be considered in subsequent sections where more complex models are adopted. Since the failure criteria on

Figure 12 Coulomb stress changes on optimally oriented strike-slip and normal faults caused by earthquakes in subsequent time windows. The causative faults are shown in black lines enclosed with a white outline. Stress-loaded faults are red enclosed with a black outline. The black segments show those faults that are loaded by less than ±0.1 b. Solid circles indicate aftershocks of some of the events, while open circles show background activity. (a) Stress changes caused by the events in the time window 1912–44 (1912, 1919, 1924, 1928, 1932, 1935, 1939, 1942, 1943, and 1944.1). (b) Stress changes caused by the events in the time window 1912–67 (the previous events plus those in 1944.2, 1944.3, 1953, 1956, 1957, 1963, 1964, 1965, 1967.1, and 1967.2). (c) Stress changes caused by the events in the time window 1912–96 (all the previous events plus those in 1968, 1969, 1970, 1975, 1978, 1981.1, 1981.2, 1982, 1983). Four regions both showed increased Coulomb stress and included faults showing clear Holocene activity, part of the western Sea of Marmara, the western part of the Biga Peninsular, the Saroz Gulf, and the Izmit Bay region where the August 1999 (M_s 7.8) earthquake occurred. Reproduced from Nalbant SS, Hubert A and King GCP (1998) Stress coupling between earthquakes in northwest Turkey and the North Aegean Sea. *Journal of Geophysical Research* 103: 24,469–24,486, with permission from American Geophysical Union.

faults are time dependent, the lithosphere is inhomogeneous and loading conditions are not simple stress fields; attempts to account for such effects introduce a wide range of possible models.

4.08.4.1 Time-Dependent Rock Failure

Failure in a rock sample does not occur at the peak stress, but occurs in a 'post-peak' region when the system is mechanically weakening faster than the loading system can respond (Scholz, 1990). There is a mechanical breakdown associated with this weakening and which requires finite displacement before failure can occur. Because this breakdown, or nucleation is quasi-static, it requires a finite time to occur. It is this 'kinematic' feature of failure that introduces time into failure criteria. For rock samples, this is associated with the growth and interaction of defects, with the largest defects being the most important (e.g., Ashby and Sammis, 1990). The effect is to be expected both for new failure in rock and for new failure on an existing fault surface. In the lithosphere, there is no way to establish *a priori* the scale at which the breakdown process occurs; however, for large earthquake faults, this appears to involve large volumes of rock (Manighetti *et al.*, 2004).

Since earthquakes are commonly regarded to result from new slip on existing surfaces, the behavior of such surfaces has been carefully studied in the laboratory, giving rise to 'rate- and state-dependent friction models'. These have been widely adopted and are effective in explaining many observations. While the mathematical formulation undoubtedly captures basic features of the underlying mechanical processes, it is important to remember that failure in the lithosphere could involve different processes from a saw cut in granite. It is important to appreciate that 'rate- and state-dependent friction models' can only partially characterize behavior for a 'post-peak' region of a failure curve. At low stresses stress failure is not time dependent.

4.08.4.2 Rate- and State-Dependent Friction Equations

Many formulations of rate- and state-dependent friction law have been proposed (e.g., Scholz, 1998, and references therein). In general, these constitutive laws can be described by two coupled equations. The first one relates the sliding resistance τ to the slip velocity V and the state variables θ_i; it is usually named the governing equation,

$$\tau = F[V, \theta_i, A, B, \mu_0, \sigma_n]$$

where A and B are two positive parameters that depend on the material properties, temperature and pressure. μ_o is a reference friction value and σ_n is the normal stress. The second law provides the time evolution of the state variable and it is called the evolution equation. For a single state variable, this becomes

$$\frac{\partial \theta}{\partial t} = G[V, L, \theta, B]$$

The state variable θ provides a memory for the sliding surface. A widely used formulation for these laws includes a single state variable.

$$\tau = \mu_0 \sigma_n + A \ln \frac{V}{V^*} + B \ln \frac{\theta}{\theta^*}$$

$$\frac{\partial \theta}{\partial t} = 1 - \frac{\theta}{L} V$$

These equations have been proposed to model earthquake after effects and they have been used to predict aftershock rates (e.g., Dieterich, 1994), to study stress shadowing (e.g., Harris and Simpson, 1996) as well as stress triggering (e.g., Gomberg *et al.*, 1998). Subevent triggering can occur over several hours, days, or weeks, as observed, for instance, during the 1992 Landers earthquake. During such short triggering times, the remote tectonic load is not playing any role. For longer periods, loading must be considered. The relation between short-term effects of rate and state friction and longer-term loading is illustrated in **Figure 14** (Stein *et al.*, 1997). In **Figure 14(a)**, the Coulomb stress ahead of the fault is shown to increase steadily as a result of tectonic loading at a point x until it increases abruptly because of an earthquake on an adjacent fault after which it returns to the same rate. In a simple model of repeating earthquakes, a new event occurs when a threshold is crossed. The threshold is shown to be blurred to indicate that in practice the recurrence time is irregular and has to be expressed in statistical terms – a mean and standard deviation, for example. Thus, at any given time, the probability of an earthquake in any time period depends on where the stress on the fault lies in the earthquake cycle. Shortly after an earthquake the probability of an event is low and as the mean recurrence time is approached the probability of a second event being triggered increases. The simple result of the stress step is to apparently advance the proximity to the mean recurrence time and thus increase the probability that an event will occur.

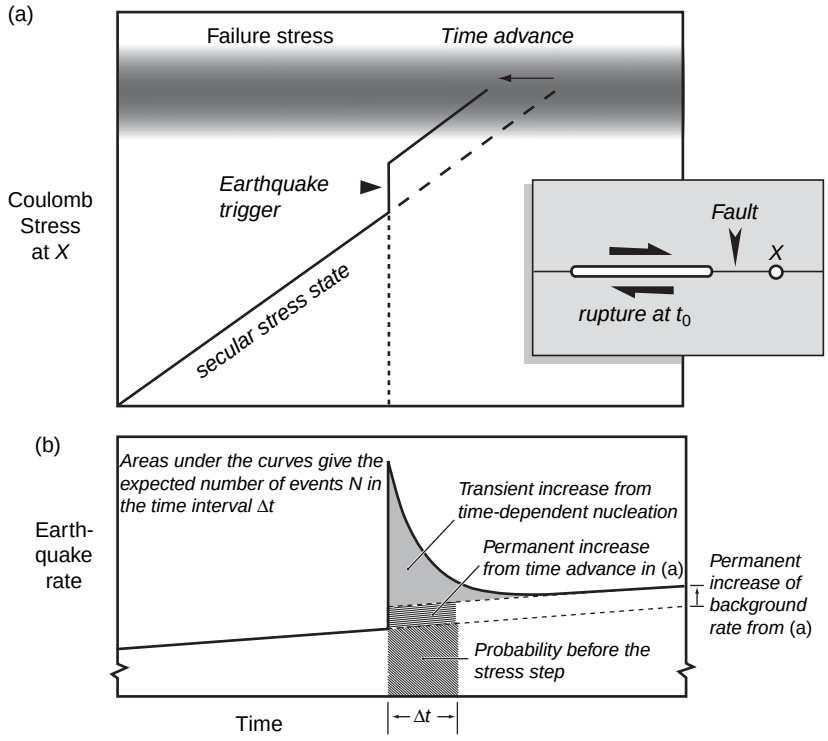


Figure 14 Schematic illustration of a sudden stress increase on a fault at point x near to an earthquake rupture at time t_o (inset). (a) The stress increase advances the time to the next rupture. If the stress was closer to failure at the time of the stress step, the stress step could have triggered the event immediately. (b) The associated short-term (transient) and long-term (permanent) increase in the rate of earthquake occurrence. This can also be interpreted as an increase of probability of a large event. Reproduced from Stein R, Barka AA, and Dieterich JH (1997) Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International* 128: 594–604, with permission from Blackwell.

The transient effects predicted by rate and state friction behave differently (**Figure 14(b)**). It is assumed that the stress on the fault may be at different distances from the peak stress threshold. If an abrupt stress increase lifts the fault past that threshold it is only a matter of time before it will fail, with no further increase of loading being required. This then results in a much greater increase of probability than the simple recurrence time model. (Stein, 1999)

Numerous studies have adopted this approach. A study that uses the rate and state model to predict both the rates of small earthquakes and the probabilities of larger events has been published by Toda *et al.* (2005), accompanied by an animation that follows the evolution of seismicity in southern California between 1986 and 2003. The study provides an articulate description of the use of rate and state friction. In a key conclusion, they suggest that changes of stress level modify the rate of seismicity rather than suppressing it completely. This might suggest that stress shadows do not exist only as substantial reductions of seismicity rate. Bearing in mind that rate and state friction is only established as an empirical description of final yielding for a post-peak region (see Section 4.08.4.1), whether background seismicity can be completely suppressed or only reduced in an epicentral region can hardly be established by the sparse data.

4.08.5 Loading Models and Lithospheric Properties

4.08.5.1 Elastic Half-Space Loading Models

In the introduction, two categories of lithospheric models were described: one with faults or localized shear zones extending to the asthenosphere and the second with viscous deformation occurring in the lower crust and below. Both of the approaches with various adaptions have been introduced to

explain the long-term effects of loading and shorter-term time-dependent effects following big earthquakes.

From the point of view of Coulomb calculations, the most widely adopted models assume that a fault continues below seismogenic depths, and between earthquakes moves at a constant rate. Where a sufficient span of GPS data is available, the model used to explain geodetic observations can be used directly to generate long-term loading. An example is shown in **Figures 15(a)–(d)** adapted from Pondard *et al.* (2007) that considers loading and earthquakes since 1600 for the Sea of Marmara region. **Figures 15(a)** and **b** show the stress loading based on the rate defined by GPS measurements which are also consistent with longer-term geological rates. **Figure 15(c)** shows the stresses due to the earthquakes alone, and **Figure 15(d)** combines these with the loading (**Figure 15(b)**). A ~100 km stretch of fault in the Sea of Marmara is now stressed and represents an earthquake hazard to Istanbul. Stress change and loading models of this kind have also been adopted to provide the parameters needed to estimate the probabilistic seismic hazard discussed in Section 4.08.4.2 (e.g., Parsons, 2004).

The creeping fault at depth is assumed to have a velocity-hardening behavior and can consequently be expected to slip at a greater rate following an earthquake (e.g., Çakir 2003, Perfettini and Avouac 2004). This afterslip should introduce steadily increasing stresses following an earthquake and particularly for events with large L-to-W ratios should cause the stressed region to expand with time (e.g., Stein *et al.*, 1992). The effect of this evolving stress field offers a way to explain the time evolution of aftershocks that differs from that suggested by rate and state friction (Section 4.08.4.2). However, viscous relaxation at depth and fluid flow processes are also candidates, as discussed in Sections 4.08.5.2 and 4.08.5.4.

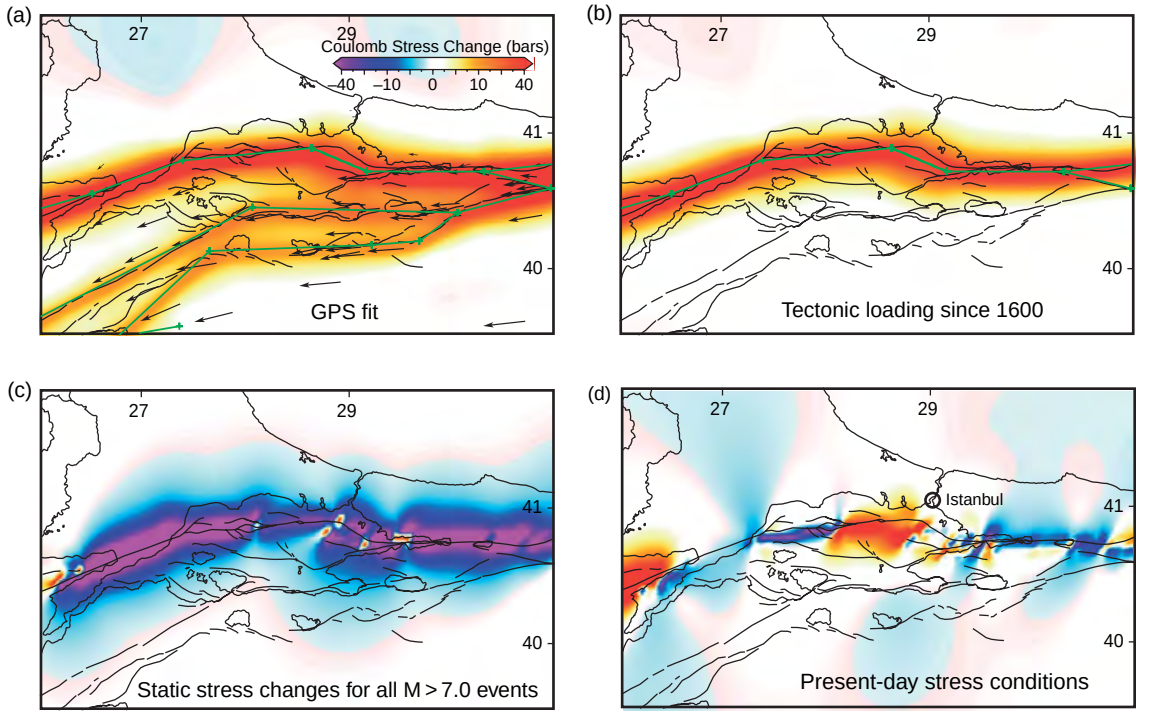


Figure 15 Accumulated loading and stress loading for the Marmara Sea region from 1600 to 2000. (a) Loading rate based by fitting GPS vectors for all branches of the Anatolian Fault around the Sea of Marmara. The panel only shows a part of region studied to determine the rates (Flerit *et al.*, 2003). The inset scale bar corresponds to a time period of 400 years and is the same for all subsequent panels. The GPS vectors are shown as arrows, and green lines identify the slipping faults beneath the locking depth. (b) The loading for the northern branch alone for 400 years. (c) Stress changes resulting from all earthquakes >M 7.0 since 1600. Smaller events are of little significance. (d) The combined effect (summed) of loading and stress changes due to events. The stressed region south and west of Istanbul is expected to host a future ~M 7 earthquake.

4.08.5.2 2-D Plate Models

Models with an elastic plate over a viscous half-space have proved less easy to adopt. Indeed, it is not clear how they could be fully implemented. Thus, only their possible effects have been modeled using end members. One end member considers the plate to be driven at a distance by displacement boundary conditions and viscous effects are ignored. Faults or deformation zones evolve in the plate as a result of purely elastic interactions between a distribution of initial defects which grow and interact by obeying a simple failure law. Studies of this sort adopt automaton modeling (e.g., Ferguson *et al.*, 1999; Olami *et al.*, 1992). These approaches assume that the seismogenic layer is 100% seismogenic and all earthquakes are a consequence of earlier events. This assumption also underlies models that call on 'self-organized criticality' (SOC) and is an underlying justification for concepts such as 'epidemic-type aftershock sequence' models (Helmstetter and Sornette, 2002; Ogata, 1989, 1988) Some authors suggest that these models have a universal application (e.g., Helmstetter and Sornette, 2002; Helmstetter *et al.*, 2003; Ogata, 2004; Sornette and Helmstetter, 2002).

While these techniques may produce a useful statistical approach to characterizing earthquake catalogs and some insight into interaction processes, a problem lies in differences in interaction between faults with small and large L/W ratios and their coupling to loading processes resulting from stable sliding both within and beneath the seismogenic zone. This is summarized in **Figure 16**. **Figure 16(a)** shows that the stick-slip region does not come fully to the surface and even within the seismogenic region, patches of stable sliding can occur. This effect can be seen in the depth distribution of slip contributed by events at different depths (**Figure 16(b)**). Near to the surface and near the base of the seismogenic layer, stick-slip events contribute only a modest amount of slip (King *et al.*, 1994a). In creeping regions such as central California, only a few percent of the slip is seismic and occurs on a few patches (**Figure 16(c)**). Events in such regions are too widely separated to interact (**Figure 16(d)**). Unlike the situation shown in **Figure 13** in the center of a noncreeping seismogenic zone where interactions are strong. It seems that complete stress coupling and stick-slip behavior can only occur in a narrow region in the center of a seismogenic zone . Since events of $M \sim 6$ completely cut the seismogenic zone, only much smaller events in a narrow depth range can be considered to be dominated by elastic interactions alone. The related problem of large earthquakes not belonging to the same fractal set as small ones is considered by Scholz (1990) and Pacheco *et al.* (1992).

A second end member considers only the effects of stress relaxation on viscous layers at depth following an earthquake in the seismogenic zone. This has been proposed to explain at least some aspects of the time-dependent effects associated with aftershocks and delayed triggering of more distant large events (e.g., Chery *et al.*, 2001; Freed and Lin, 2001; Pollitz *et al.*, 2004). The effect however is almost indistinguishable in terms of the evolving stress field to the assumption that increased post-earthquake slip occurs on a fault at depth. The approach has proved successful to explain earthquake interactions in Mongolia, for example; however, the success does not demonstrate that the approach is necessarily correct. However, the vertical component of the displacement field determined from satellite-based radar shows that such an approach cannot be correct (Peltzer *et al.*, 1998).

If the viscous layer is considered to be much deeper (e.g., the base of the lithosphere), the approach can be combined with the concept of creeping faults below the seismogenic depth and the two approaches converge. However, whatever its role in explaining the triggering of distant large earthquakes over long time periods, the effect of relaxation at the base of the lithosphere becomes too small to explain the temporal evolution of aftershock sequences or earthquake clusters.

4.08.5.3 Inhomogeneous Lithospheric Models – The Effects of Voids or Fissures

The most dramatic inhomogeneity that can be introduced in a solid is a void, which can contain volcanic magma, water or gas. A familiar feature of volcanic or thermal regions is the earthquakes that occur very close to a volcanic vent or hot springs. The *b*-value of the Gutenberg–Richter relation associated with such earthquakes is commonly high – a large number of small events occur for each larger event. Expressed in terms of moment release, small events account for as much deformation as large events unlike the seismicity of nonvolcanic regions (see also Section 4.08.5.5). The same character is shared by the seismicity of creeping faults and in both cases the deformation associated with the seismicity is tiny compared with

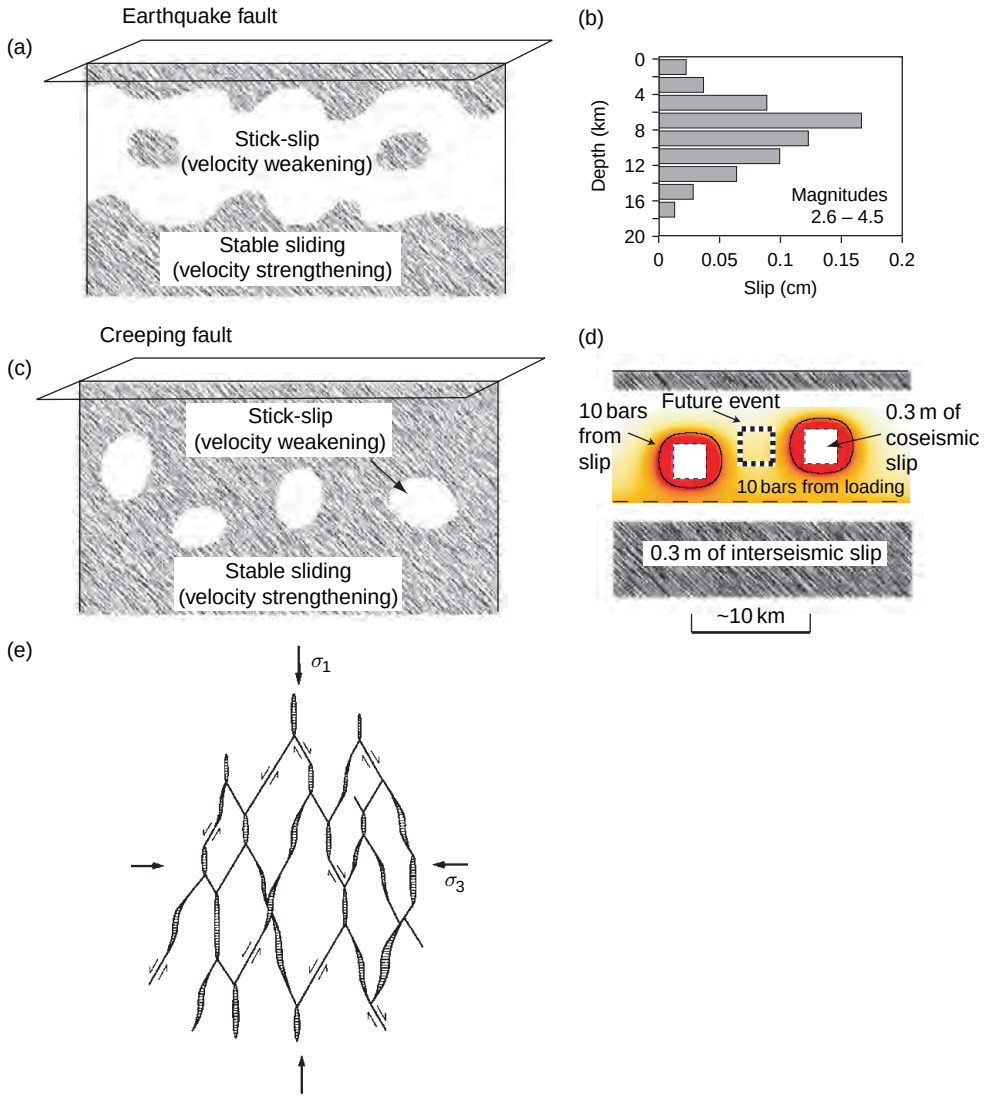


Figure 16 (a) Strike-parallel longitudinal section showing schematically regions of stick-slip and stable sliding. (b) The depth distribution of slip for events with magnitudes between 2.6 and 4.5 for the San Francisco Bay Area. The seismic slip is concentrated around 8 km. Similar results are found for other magnitude ranges (King *et al*., 1994a), indicating that at shallow and greater depths in the seismogenic zone much of the deformation is aseismic as indicated in (a). A discussion of the depth of seismicity within the seismogenic zone can be found in Pacheco *et al*. (1992). (c) Strike-parallel longitudinal section showing schematically regions of stick-slip and stable sliding for a fault where aseismic creep predominates. (d) For a creeping fault, small stick-slip regions are too far apart to interact by stress coupling. (e) A mesh model for swarm seismicity involving interlinked minor faults and extension fractures infilled with dikes, hydrothermal veins, etc. (cross-hatched), developed in relation to the principal compressive stresses, $\sigma_1 > \sigma_2 > \sigma_3$. Modified from Hill DP (1977) A model for earthquake swarms. *Journal of Geophysical Research* 82: 1347–1352.

that due to nonseismic processes (see also **Figure 16**). The most straightforward explanation is that earthquakes in volcanic regions like those on creeping faults are not interacting, being seperated by voids that deform aseismically. Sibson (2002) offers a hypothetical geometry for such processes based on the work of Hill (1977).

4.08.5.4 Interactions between Volcanic Eruptions and Earthquakes

Since volcanic fissures or magma chambers can be treated using dislocation theory; the techniques described for the interaction between large earthquakes can be modified to examine interactions between volcanoes and earthquakes and vice versa.

Unlike earthquake faults, however, the plumbing of volcanoes is less well understood. It is not clear whether an eruption should be regarded as resulting from an increase of magma pressure forcing magma to the surface or a reduction normal stress allowing fissures to open, allowing magma to reach the surface. It is similarly not clear whether the magma is contained in quasi-spherical chambers or in fissures. The various possibilities result in different stress fields whether examining the triggering of eruptions by earthquakes or the reverse. Notwithstanding these possible ambiguities, plausible relations between static stress changes associated with earthquakes and eruptions have been proposed for Djibouti and Italy (e.g., Noir *et al.*, 1997; Nostro *et al.*, 1998). Relations between volcanoes and earthquakes including dynamic and viscoelastic effects are reviewed by Hill *et al.*, (2002).

4.08.5.5 Fluid Movement in Nonvolcanic Regions

Substantial evidence is accumulating for the mobilization of water in the seismogenic crust following major earthquakes both close to the active fault and in a region extending for several fault lengths from the epicenter. Direct evidence is derived from water expelled at the surface by the compressional elastic rebound of normal faulting earthquakes (Muir-Wood and King, 1993). Monitored changes of river flows finds that the increased flow progressively decays over 6–10 months (**Figure 17**). Comparing the volume of water expelled to the volumetric strain calculated from a source known from seismic and geodetic data indicates that a substantial proportion of the coseismic stress is relaxed by fluid flow. The volume expelled and the time constants involved suggest that the water is held in a connected system of high aspect ratio cracks. The importance of fluid flow is also suggested from GPS and InSAR studies of post-earthquake relaxation (Jonsson *et al.*, 2003; Peltzer *et al.*, 1998). As noted earlier, in these cases the relative effects of relaxation in a viscous lower crust, post-earthquake creep on a lower crustal fault, and the effects of water flow are not easily separated.

The importance of the fluid flow suggested by these studies indicates that it must play some controlling role in aftershock sequences, and the time period of about 8 months is similar to the length of aftershock sequences (Muir-Wood and King, 1993). Two effects may control seismicity. The first is a change of the effective friction due to changes of fluid pressure (eqn [6]) (e.g., Nur and Booker, 1972), thus altering the rate and distribution of seismicity predicted from Coulomb stress changes alone. While the Coulomb stress distrubution may not be greatly changed, fluid pressure changes may be an important alternative to rate and state friction to explain temporal changes. Furthermore, increases of pressure could trigger events in regions where reduced Coulomb stress predicts an absence of activity.

The suggestion that water is contained in high aspect ratio cracks and changes in pressure over a period of several months suggests that the effective rock modulus will change over a similar period. Since water is held in cracks whose orientation reflects the prevailing stress field, the modulus changes can be expected to be anistropic. Such readjustments may be expected to generate stress fields and hence seismicity that are not simply related to a regional stress field or the Coulomb stress change field following an earthquake. Several publications on the topic can be found in the review papers in the introduction. Of particular interest are recent observations of a correlation between the monsoon and midcrustal seismicity in the Himalayas (Bollinger *et al.*, 2007) and an aftershock sequence driven by high-pressure carbon dioxide rising from depth (Miller *et al.*, 2004).

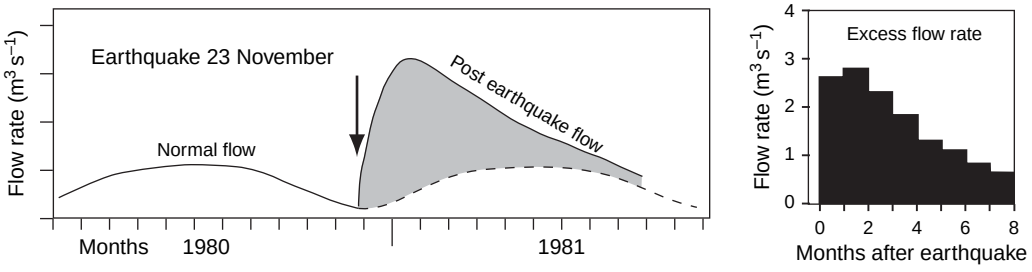


Figure 17 Excess flow for the Caposele spring resulting from compressional elastic rebound following M_s 6.9 Irpinia earthquake 1980. The form of the excess flow is typical for normal faulting events (Muir-Wood and King, 1993).

4.08.5.6 Inhomogeneities Resulting from Loading

Prior to the popularity of Coulomb stress triggering, other explanations for off-fault aftershocks had been suggested. For example off-fault events have been explained as being related to bends or offsets of faults (see King (1986) among various authors) with the fragmentation needed for long-term slip to traverse such barriers requiring deformation to occur away from the main faults (King and Nabelek, 1985). Underlying this approach is concept that a system of faults must collectively move such as to relieve the applied boundary conditions. Kinematic rules must be obeyed because faults accommodate finite deformation, which imposes requirements that cannot be considered by stress models or by the mathematical simplifications of infinitesimal strain. For simple shear displacement boundary conditions, motion can in principle be accommodated by strike-slip faults with a single strike direction (**Figure 18(a)**). For simple compression or extension, antithetic dip-slip faults can also, in principle, accommodate long-term motion (**Figure 18(b)**), provided the faults completely cut from the surface to a depth at which stress is relaxed by viscous flow. No other boundary conditions can be accommodated in such a simple way, even in principle (**Figure18(c)**). Where two faults intersect, a third fault direction must exist to accommodate finite motion. At its simplest, a bend becomes a triple junction. In 3-D, five directions must exist to accommodate finite brittle strain (King, 1983).

In the figures, it can be seen that the three deformation systems have the same principle axis of shortening, but in the case of **Figure 18(c)** several different focal mechanisms must operate to generate the observed azimuthal strain distribution (strain rosette). The assumption in Section 4.08.2 that the principle stress axes uniquely define the orientation of fault planes is not correct. The effect of complex boundary conditions has been seen most clearly for the aftershocks of the Irpina 1980 (Italy) earthquake (Deschamps and King, 1985) and in events both before and after the 1989 Loma Prieta (USA) earthquake (Beroza and Zoback, 1993).

A consequence of multifault deformation is that the concept of a temporally constant regional stress field is not appropriate. The stresses must be both inhomogeneous and evolve with time. The observation of complex fault mechanisms is a consequence of the kinematics of finite deformation and cannot be explained to be the consequence of low friction on faults alone. Whether or not fault friction is low, multiple mechanisms are the only way to accommodate complex boundary conditions. Although stresses are inhomogeneous, the strain that is relieved by a collection of mechanisms is simple. **Figure 18(d)** shows strain rosettes for the whole San Francisco Bay Area. The deformation due to small events summed in magnitude ranges is the same as that predicted by the overall plate motion. The form is such that the region must be associated with a number of fault mechanisms. This can be seen more clearly for strain rosettes for different regions around the San Francisco Bay Area shown in **Figure 19**. For the Calaveras and Hayward faults, the strain can be relieved by simple strike-slip faulting. Along the San Francisco peninsular this is not the case; the strain conditions differ. The Golden Gate is associated with substantial extension, the central peninsular and Loma Prieta region with substantial shortening. The Golden Gate region is thought to be where the great 1906 earthquake initiated. This event could have started on a fault with a substantial normal component of slip and would be triggered by stresses very different from those that would be appropriate if the fault could be represented by the average strike and mechanism of the San Andreas. This expands further on the remarks in Section 4.08.3.8 that suggest that resolving Coulomb stresses onto a plane determined from field measurements may not be appropriate. However, where a number of fault planes must operate to accommodate the strain boundary conditions, a single time independent regional stress cannot exist, it must be both temporally and spatially variable. There is consequently no reason why a optimum fault direction based on a constant regional stress should be correct either.

4.08.5.7 The Distribution of Small Events before Large Earthquakes

The aftershocks that result from a large earthquake are unambiguously due to stress increases that result from the event (Section 4.08.3). Over longer periods of time, there is also clear evidence for earlier large earthquakes setting up the stress conditions that control future large events. In the case of Loma Prieta, smaller events also appear to have loaded the site of the future event (**Figure 8** and associated text). The effect of long-term loading is also important and becomes progressively more significant for events with a large L-to-W ratio where interactions become less important

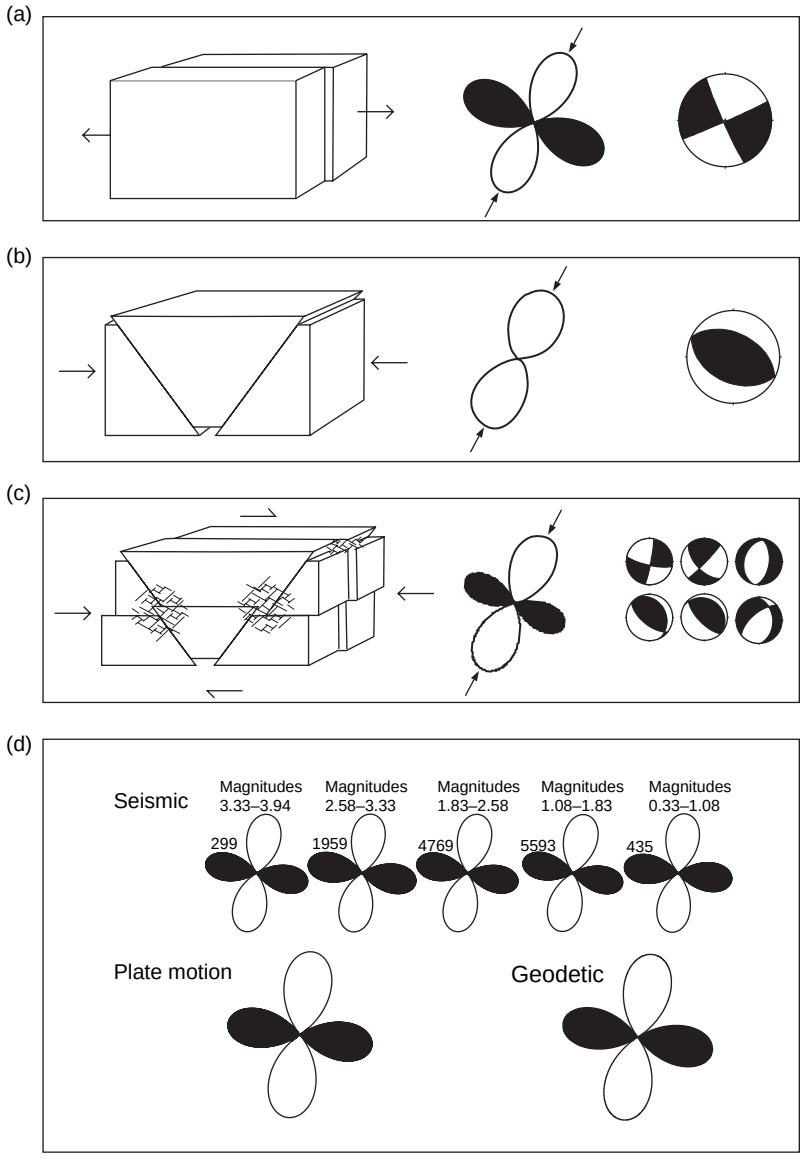


Figure 18 The kinematics of finite strain. (a) Simple shear displacement boundary can be released on a single fault. The associated strain rosette and focal mechanism are shown. (b) Simple contraction can be relieved (under certain circumstances) by simple dip-slip faulting. The associated strain rosette and focal mechanism are shown. (c) More complex displacement boundary conditions require multidirection faulting indicated schematically. The strain rosette can only result from numerous earthquakes with different mechanisms operating together. The mechanisms shown are typical of the heterogeneous aftershocks for the Loma Prieta earthquake. (d) The seismic strain can be represented by examining earthquakes in magnitude windows each of which includes earthquakes that change average fault length by a factor of 2. The distribution of finite strain released by small events has the same form as that predicted from plate motion and determined geodetically. Parts of the figure are adapted from Amelung F and King GCP (1997a) Large scale tectonic deformation inferred from small earthquakes. *Nature* 386: 702–705.

(Section 4.08.3.7). Since the 1980s, considerable interest has been directed to the study of accelerated moment release (AMR) (e.g., Jaumé and Sykes, 1996), where activity is found to increase over a region with dimensions several times greater than the fault length (L). A range of explanations have been offered to explain the observation. Two relate directly to the concept of stress interactions. In the first, stress triggering between small events is considered to progressively create a stress field sufficiently homogeneous over a large

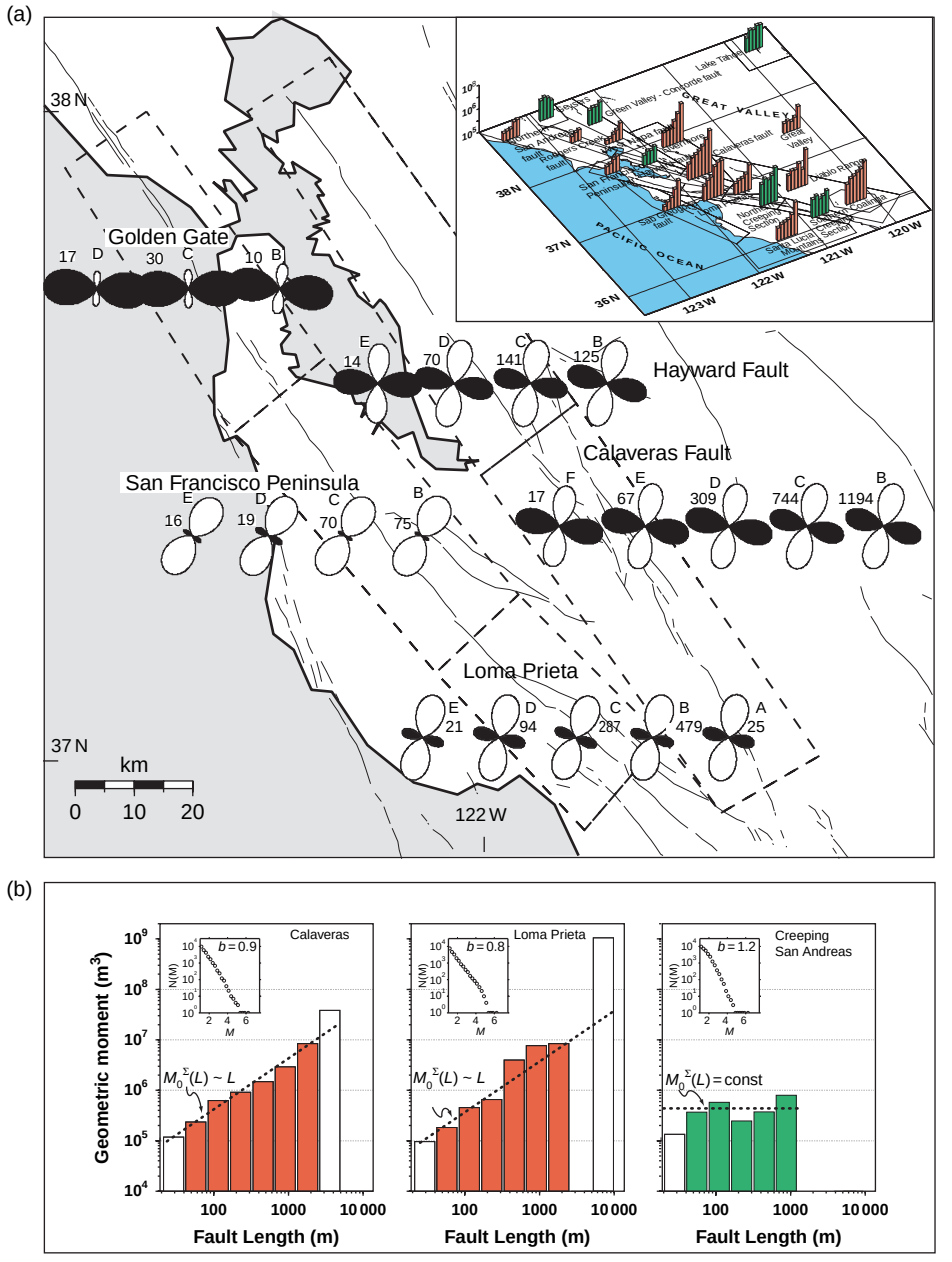


Figure 19 Earthquake kinematics for the San Francisco Bay Area. (a) The strain rosettes are created by separating events into bins with average fault dimensions differing by factors of 2 (as described in the caption to **Figure 18**). Although the overall finite strain summed for the whole region is the same as the applied boundary conditions (**Figure 18(d)**), there are substantial local variations. (b) The scaling relations for the summed geometric moment (potency) in the bins described above for example regions. For noncreeping regions, the moment release doubles for each factor of two increase of fault dimensions. For creeping faults and volcanic or thermal spring regions, the moment release is the same for all bins. Insets show the equivalent *b*-values. The inset to (a) shows the observed scaling for regions in North California. The figures are simplified from Amelung F and King GCP (1997a) Large scale tectonic deformation inferred from small earthquakes. *Nature* 386: 702–705; Amelung F and King GCP (1997b) Earthquake scaling laws for creeping and non-creeping faults. *Earth and Planetary Science Letters* 24: 507–510. The rosettes are labeled with letters between A and F which correspond to magnitude bins A, 1.08–2.58; B, 2.58–3.33; C, 3.33–3.94; D, 3.94–4.54; E, 4.54–5.14; F, 5.14–5.75.

enough region to host a big earthquake. The process is considered to be a manifestation of SOC. This attractive model is unfortunately difficult to relate to the observations of the geologists, geodesists, or seismologists who observe large earthquakes repeating on readily identified faults. Some of these problems have been addressed by the conditional criticality model (Bowman and Sammis, 2004).

A second explanation attributes the increase of activity as the consequence of the stress in a wide region increasing toward failure to be the consequence of the final stages of reloading a fault. This can be thought of as the final refilling of an earthquake shadow resulting from an earlier event. The final process of loading is shown schematically in **Figure 20** (King and Bowman, 2002; Scholz, 1990), with activity progressively increasing as the stress shadow fills.

Following an earthquake, the region that must have been loaded prior to the event can be calculated in a similar way to that used to identify post-earthquake Coulomb stress increases, but by slipping the fault in the oposite direction to the slip that occurred in the event. It should be appreciated that this back-slip approach (Savage, 1983) models only the large scale stress field that drives the earthquake. More local stress fields associated with fault complexity that progressively produce topography are not modeled in this way (e.g., Perfettini and Avouac, 2004). For all the events with magnitude >6.5 in California, an acceleration of activity has been found. When compared to random catalogs, the individual probability that the AMR for each event is a physical effect is modest. However, the cumulative probability that all events show AMR as a result of random processes is less than 1%. The AMR in the far-field region of the 26 December 2004 Sumatra earthquake prior to the event (**Figure 21**) had less than 2% probability of being random (Mignan *et al.*, 2006). While for Californian events with L-to-W ratios of ~3.5, effects other than loading can possibly be invoked to explain AMR, for the Sumatra event with an L-to-W ratio of ~35, loading seems the only plausible explanation for the observed pre-event increase of seismic activity.

4.08.6 Dynamic Triggering

This chapter concerns static stress triggering – its successes and failures. When discussing the latter, the possibility of events being triggered by large stresses associated with the passage of seismic waves must be considered. There is no doubt that dynamic triggering occurs; indeed, the propagation of rupture on a fault plane is a process of continuous triggering. At distances too great to result from static stress changes, events occur that are clearly related to the distant earthquake. The events in Alaska associated with 26 December 2004 Sumatra earthquake are a stunning example (West *et al.*, 2005).

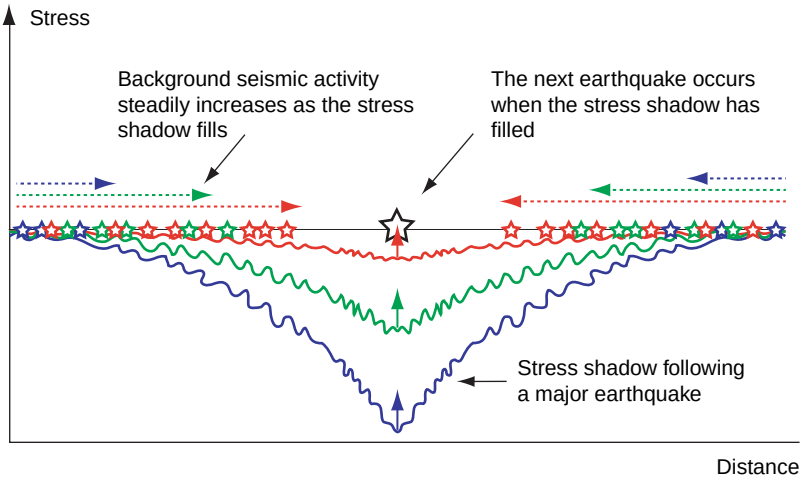


Figure 20 Following a large earthquake, a stress hole or shadow is created around the associated fault (blue line). This progressively fills (green, then red lines), until the next major earthquake occurs. For the filling of the stress shadow for events with large L-to-W ratios, the effect of aseismic creep at depth predominates. As the stress shadow fills, activity increases over a wide region (Scholz (1990); King and Bowman (2002)).

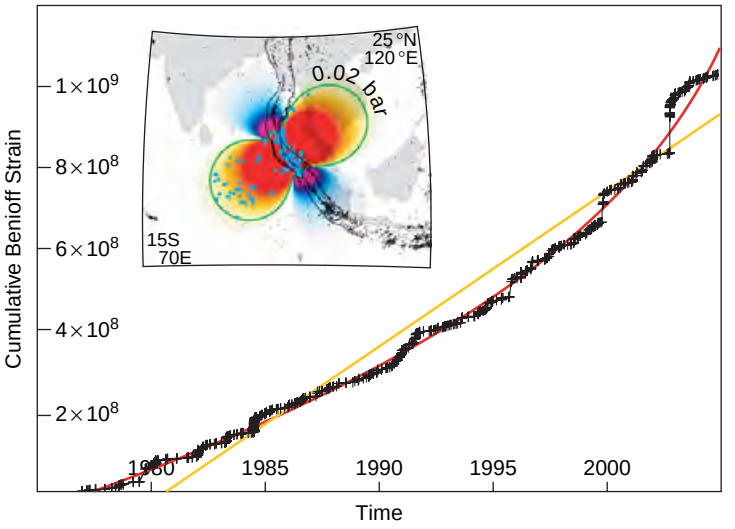


Figure 21 AMR for the great Sumatra earthquake of 26 Dec. 2004. The Cumulative Benioff Strain and the time (in years) are, respectively, the *y*-axis and *x*-axis. The power law fit (red curve) has less than a 2% probability of occurring randomly. The pre-stress field calculated by the backslip earthquake method for the high slip part of the 26 Dec. 2004 event is shown in the inset. The red contour (0.02 b) outlines the area of events that provides the AMR. The fault used has four pure dip-slip segments each with a dip of 11°, a down-dip width of 200 km. Reproduced from Mignan A, King GCP, Bowman R, Lacassin R, and Dmowska R (2006) Seismic activity in the Sumatra-Java region prior to the December 26, 2004 (Mw = 9.0-9.3) and March 28, 2005 (Mw = 8.7) earthquakes. *Earth and Planetary Science Letters* 217: 339–408, with permission from Elsevier.

The dynamic triggering of events, however, cannot be explained by a failure occurring at a simple stress threshold as discussed in Section 4.08.2. Except near to the epicenter, the peak dynamic stresses associated with propagating seismic waves are much greater than those associated with the static stress change. Threshold triggering would result in every very big earthquake, immediatedly triggering events all over the world. Because failure takes time (Section 4.08.4), this does not occur. Even large transient stresses are less important than sustained stresses.

If delayed failure suppresses dynamic triggering, why does it occur? At great distances, triggering is commonly associated with fluid processes, either water or magma, suggesting that special conditions are needed for it to occur (Hill *et al.*, 2002). These are discussed in Chapter 4.09. Until the processes involved are better understood, events that appear in the negative lobes of static Coulomb stress change, must be considered to be a possible consequence of dynamic stresses.

4.08.7 Conclusions

Relating the occurrence of one earthquake to another using changes of Coulomb stress is clearly successful. The distribution of aftershocks can be effectively explained in many cases. Interactions between large events have also been demonstrated and used to successfully predict destructive earthquakes (in Turkey and Sumatra). A problem with this approach is the modest length of most historical catalogs. This information can be recovered by geological methods (fault geomorphology and trenching combined with steadily improved methods of dating), which will in due course greatly improve our understanding of earthquake hazard. Of equal importance, the concept of stress shadows can be used to suggest where destructive events are less likely to occur (Section 4.08.3.6). However, where these stress shadows fall along major faults they do not last indefinitely. Adjacent earthquakes and tectonic loading over time restores the region to failure prior to the next substantial event. This is associated with increasing background activity in the region. Where it can be identified (Section 4.08.5.7), this provides a further approach to identifying the location of future damaging events. These results are of major social importance.

Stress interactions do not occur instantaneously – time is involved. The most successful approach to quantifying this invokes the concept of rate and state friction (Section 4.08.4.2). Following an abrupt stress

increase, the rate at which small events occur increases. A combination of Coulomb stress change and rate and state friction can explain both the distribution of aftershocks and the Omori law which describes their rate of decay. The same approach can be adapted to predict an enhanced earthquake probability of a second event following a first. Such earthquake probability forecasts form the most effective way of communication risk to goverment authorities, emergency services, and the general public, but it will be many years before the approach can be scientifically validated.

Despite the successes outlined, in many cases earthquakes occur where they are not predicted by Coulomb stress changes alone. Many possible explanations have been discussed. Many draw on the fact that lithosphere is not a homogeneous elastic half-space, faults are not fully described by dislocation surfaces, and stress boundary conditions do not capture important features of finite deformation. Dynamic triggering can also explain features that cannot be explained by static stress models.

The problem with advancing our understanding of how earthquakes interact lies in finding data that can be interpreted unambiguously. At present, there are more models possible than data to distinguish between them. However, improvements in the data needed to understand lithospheric mechanics should come to our aid. For example, recent advances in our ability to map faults with the aid of satellite images and determine their rates with novel techniques such as cosmogenic dating. Space geodesy, both GPS and InSAR, are also placing constraints on the behavior of the lithosphere. In due course, these will allow a more complete understanding of how and with what earthquakes interact. While we wait for the future, we are lucky that simple ideas work quite well and can certainly save lives.

Acknowledgments

This is IPGP contribution number 2214 and INSU contribution number 396.

References

Aki K and Richards P (1980) *Quantitative Seismology*. San Francisco: W. H. Freeman and Co.

Amelung F and King GCP (1997a) Large scale tectonic deformation inferred from small earthquakes. *Nature* 386: 702–705.

Amelung F and King GCP (1997b) Earthquake scaling laws for creeping and non-creeping faults. *Earth and Planetary Science Letters* 24: 507–510.

Ashby MF and Sammis CG (1990) The damage mechanics of brittle solids in compression. *PAGEOPH* 133: 489–521.

Beroza GC and Zoback MD (1993) Mechanism diversity of the Loma Prieta aftershocks and the mechanics of mainshock-aftershock interaction. *Science* 259: 210.

Bollinger L, Perrier F, Avouac J-P, Sapkota S, Gautam U, and Tiwari DR (2007) Seasonal modulation of midcrustal seismicity in the Himalayas of Nepal. *Geophysical Research Letters* 2006GL029192.

Bowman DD and Sammis C (2004) Intermittent criticality and the Gutenberg-richter distribution. *PAGEOPH* 161: 1945–1956.

Çakir Z, de Chabalier J-B, Rolando A, Bertrand M, Aykut B, and Gilles P (2003) Coseismic and early postseismic slip associated with the 1999 Izmit earthquake (Turkey), from SAR interferometry and tectonic field observations. *Geophysical Journal International* 155: 93–110.

Chery J, Carretier S, and Ritz J-F (2001) Postseismic stress transfer explains clustering of large earthquakes in Mongolia. *Earth and Planetary Science Letters* 194: 277–286.

Cocco M and Rice JR (2002) Pore Pressure and poroelasticity effects in Coulomb stress analysis of earthquake interactions. *Journal of Geophysical Research* 107(B2): 2030.

Deschamps A and King GCP (1985) Aftershocks of the Campania-Lucania (Italy) earthquake of 23 November 1980. *Bulletin of the Seismological Society of America* 76: 2483–2517.

Dieterich J (1994) A consitutive law for the rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research* 99: 2601–2618.

Ferguson CD, Klein W, and Rundle JB (1999) Spinodals, scaling and ergodicity in a threshold model with long-range stress transfer. *Physical Review E* 60: 1359–1373.

Flerit F, Armijo R, King GCP, Meyer B, and Barka A (2003) Slip partitioning in the Sea of Marmara pull-apart determined from GPS velocity vectors. *Geophysical Journal International* 154: 1–7.

Freed AM (2005) Earthquake triggering by static, dynamic and postseismic stress transfer. *Annual Review of Earth and Planetary Sciences* 33: 335–367.

Freed AM and Lin J (2001) Delayed triggering of the 1999 Hector Mine earthquake by viscoelastic stress transfer. *Nature* 411: 180–183.

Gomberg J, Beeler NM, Blanpied ML, and Bodin P (1998) Earthquake triggering by transient and static deformation. *Journal of Geophysical Research* 103: 24411.

Harris RA (1998) Introduction to special section: Stress triggers, stress shadows and implications for seismic hazard. *Journal of Geophysical Research* 103: 24 347–24 358.

Harris RA and Simpson RW (1996) In the shadow of 1857 - the effect of the great Ft. Tejon earthquake on subsequent earthquakes in southern California. *Geophysical Research Letters* 23: 229–232.

Helmstetter A and Sornette D (2002) Sub-critical and super-critical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research* 107(B10): 2237.

Helmstetter A, Sornette D, and Grasso J-R (2003) Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws. *Journal of Geophysical Research* 108(B10): 2046.

Hill DP (1977) A model for earthquake swarms. *Journal of Geophysical Research* 82: 1347–1352.

Hill DP, Pollitz F, and Newhall C (2002) Earthquake-volcano interactions. *Physics Today* 55: 41–47.

Hubert-Ferrari A, Barka A, Jacques E, and Leyman SA (2000) Seismic hazard in the Marmara Sea following the 17 August 1999 Izmit earthquake. *Nature* 404: 269–272.

Jaumé SC and Sykes LR (1996) Evolution of moderate seismicity in the San Francisco Bay region, 1850 to 1993: Seismicity changes related to the occurrence of large and great earthquakes. *Journal of Geophysical Research* 101: 765–790.

Jonsson S, Segall P, Pedersen R, and Björnsson G (2003) Post earthquake ground movements correlated to pore pressure transients. *Nature* 424: 179–183.

Kagan YY and Jackson DD (1991) Long-term earthquake clustering. *Geophysical Journal International* 104: 117–133.

King GC and Nabelek J (1985) The role of fault bends in faults in the initiation and termination of earthquake rupture. *Science* 283: 984–987.

King GCP (1983) The accommodation of large strains in the upper lithosphere of the earth and other solids by self-similar fault system: The geometrical origin of b-value. *Pure and Applied Geophysics* 121: 761–815.

King GCP (1986) Speculations on the geometry of the initiation and termination of earthquake rupture and the evolution of morphology and geological structures. *Pure and Applied Geophysics* 124: 567–585.

King GCP and Bowman DD (2002) The evolution of regional seismicity between large earthquakes. *Journal of Geophysical Research* 107: 2096.

King GCP and Cocco M (2000) Fault interaction by elastic stress changes: New clues from earthquake sequences. *Advances in Geophysics* 44: 1–38.

King GCP, Oppenheimer D, and Amelung F (1994a) Block versus continuum deformation in the Western United States. *Earth and Planetary Science Letters* 128: 55–64.

King GCP, Stein RS, and Lin J (1994b) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

Lin J and Stein RS (2004) Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strike-slip faults. *Journal of Geophysical Research* 109: B02303 (doi:10.1029/2003JB002607).

Manighetti I, King G, and Sammis CG (2004) The role of off-fault damage in the evolution of normal faults. *Earth and Planetary Sciences Letters* 217: 339–408.

Marsden D and Nalbant S (2005) Methods of measuring seismicity rate changes: A review of how the M 7.3 Landers earthquake affected the aftershock sequence of the M 6.1 Joshua Tree earthquake. *Pure and Applied Geophysics* 162: 1151–1185.

McCloskey J, Nalbant SS, and Steacy S (2005) Earthquake risk from co-seismic stress. *Nature* 434: 291.

Mignan A, King GCP, Bowman R, Lacassin R, and Dmowska R (2006) Seismic activity in the Sumatra-Java region prior to the December 26, 2004 (Mw = 9.0-9.3) and March 28, 2005 (Mw=8.7) earthquakes. *Earth and Planetary Science Letters* 217: 339–408.

Miller SA, Nalbant SS, and Steacy S (2004) Aftershocks driven by a high pressure CO2 source at depth. *Nature* 427: 724–727.

Mindlin R (1936) Force at a point in the interior of a semi-infinite solid. *Physics* 7: 195–202.

Muir-Wood R and King GCP (1993) Hydrological signatures of earthquake strain. *Journal of Geophysical Research* 98: 22 035–22 068.

Nalbant S, McCloskey J, Steacy S, and Barka AA (2002) Stress accumulation and increased seismic risk in eastern Turkey. *Earth and Planetary Science Letters* 195: 291–298.

Nalbant SS, Hubert A, and King GCP (1998) Stress coupling between earthquakes in northwest Turkey and the North Aegean Sea. *Journal of Geophysical Research* 103: 24 469–24 486.

Noir J, Jacques E, Békri S, Adler PM, Tapponnier P, and King GCP (1997) Fluid flow triggered migration of events in the 1989 Dobi earthquake sequence of Central Afar. *Geophysical Research Letters* 24: 2335–2338.

Nostro C, Steinn RS, Cocco M, Belardinelli ME, and Marzocchi W (1998) Two-way coupling between Vesuvius eruptions and southern Apennine earthquakes, Italy, by elastic stress transfer. *Journal of Geophysical Research* 103: 24 487–24 504.

Nur A and Booker JR (1972) Aftershocks caused by pore fluid-flow. *Science* 175: 885.

Ogata Y (1988) Statistical models for earthquake occurrence and residual analysis for point processes. *Journal of the American Statistical Association* 83: 9–27.

Ogata Y (1989) Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics* 169: 159–174.

Ogata Y (2004) Space-time model for regional seismicity and detection of crustal stress changes. *Journal of Geophysical Research* 109: B06308.

Ogata Y, Jones LM, and Toda S (2003) When and where the aftershock activity was depressed: Contrasting decay patterns of the proximate large earthquakes in southern California. *Journal of Geophysical Research* 108(B6): 2318.

Okada Y (1985) Surface deformation to shear and tensile faults in a half space. *Bulletin of the Seismological Society of America* 75: 1135–1154.

Olami Z, Feder HJS, and Christensen K (1992) Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Physical Review Letters* 68: 1244–1247.

Pacheco JF, Schoz CH, and Sykes L (1992) Changes in frequency-size relationship from small to large earthquakes. *Nature* 355: 71–73.

Parsons T (2004) Recalculated probability of M >7 earthquakes beneath the Sea of Marmara, Turkey. *Journal of Geophysical Research* 109: B05304 doi:10.1029/2003JB002667.

Peltzer G, Rosen P, Rogez F, and Hudnut K (1998) Poro-elastic rebound along the Landers 1992 earthquake surface rupture. *Journal of Geophysical Research* 103: 30131–30145.

Perfettini H and Avouac JP (2004) Stress transfer and strain rate variations during the seismic cycle. *Journal of Geophysical Research* 109: 6402, doi:10.1029/2003JB002917.

Pollitz F, William HB, and Marleen N (2004) A physical model for strain accumulation in the San Francisco Bay region: stress evolution since 1838. *Journal of Geophysical Research* 109: B11408 (doi:10.1029/2004JB003003).

Pondard N *et al.* (2007) Fault interaction in the Sea of Marmara pull-apart (North Anatolian Fault): Earthquake clustering and propagating sequences *Geophys. J. Int.* (in press).

Rundle JB, Klein W, Gross S, and Turcotte DL (1995) Boltzmann fluctuations in numerical simulations of nonequilibrium lattice threshold solids. *Physical Review Letters* 75: 1658–1661.

Savage JC (1983) A dislocation model of strain accumulation and release at a subduction zone. *Journal of Geophysical Research* 105: 3095–3102.

Scholz C (1990) *The Mechanics of Earthquakes and Faulting*, 439 pp. New York: Cambridge University Press.

Scholz CH (1998) Earthquakes and friction laws. *Nature* 391: 37–42.

Sibson RH (2002) Geology of the crustal earthquake source, edited, 455–470, pp Int'l Assoc. Seimol. & Phys. Earth's Interior, Committee on Education.

Sornette D and Helmstetter A (2002) On the occurrence of finite-time-singularities in epidemic models of rupture, earthquakes and starquakes. *Physical Review Letters* 89: 158501.
Steacy S, *et al.* (2005) Introduction to special section: Stress transfer, earthquake triggering and time-dependant seismic hazard. *Journal of Geophysical Research* 110: doi:10.1029/2005JB003692.
Stein R, Barka AA, and Dieterich JH (1997) Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International* 128: 594–604.
Stein RS (1999) The role of stress transfer in earthquake occurrence. *Nature* 402: 605–609.
Stein RS, King GCP, and Jian L (1992) Change in failure stress on the southern San Andreas fault system caused by the 1992 Magnitude = 7.4 Landers earthquake. *Science* 258: 1328–1332.
Suppe J (1985) *Principals of Structural Geology*, 537 pp. Englewood Cliffs NJ: Prentice-Hall.
Toda S, Ross SS, Keith R-D, and Serkan B (2005) Forecasting the evolution of seismicity in southern California: Animations built on earthquake stress transfer. *Journal of Geophysical Research* 110(B5): B05S16.
Volterra V (1907) Sur l'equilibre des corps elastiques multiplement connexes. *Annales Scientifiques De L Ecole Normale Superieure* 24: 401–517.
West M, Sanchez J, and Mc Nutt S (2005) Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake. *Nature* 308: 1144–1146.

Relevant Website

http://quake.usgs.gov – USGS, Earthquake Hazards Program – Northern California.

4.09 Dynamic Triggering

D. P. Hill, US Geological Survey, Menlo Park, CA, USA

S. G. Prejean, US Geological Survey, Anchorage, AK, USA

Nomenclature

c	constant in Omori's law
g	acceleration of gravity (m s^{-1})
h	vertical distance
k	constant in Omori's law
n_a	number of earthquakes in the interval t_a
n_b	number of earthquakes in the interval t_b
p	constant in Omori's law
r	distance
t	time
t_0	origin time of distant mainshock or arrival time of dynamic stresses
t_a	specified time interval after t_0
t_b	specified time interval before t_0
$\mathring{u}$	particle velocity
v_s	phase velocity of a seismic shear wave
C	cohesive strength (MPa)
E	expected value
G	shear modulus (MPa)
K_1, K_2	constant values for CFF
M	local magnitude
M_0	seismic moment
M_W	moment magnitude
P	pore pressure
R	radius in Mohr's circle (1/2)($\sigma_1 - \sigma_3$)
T	traction acting on a fault plane at angle in Mohr's circle diagram
T_p	peak dynamic stress
β	beta statistic
δ	time increment used in the β-statistic
ε_p	peak dynamic strain
θ	angle between a fault plane and the least principal stress σ_3
λ	mean recurrence time in a Poisson distribution
μ	coefficient of friction
ρ	density
σ	effective stress component (σ_n–P)
σ_1, σ_3	greatest and least principal stress components
σ_m	mean stress (1/2)($\sigma_1 + \sigma_3$)
σ_n	rock matrix stress component
σ_z	vertical principal stress
τ	shear stress component
Δ	epicentral distance (km)
Δt	time delay between t_0 and the onset of triggered seismicity
ΔCFF	change in Coulomb failure function

4.09.1 Introduction

One of the outstanding questions in earthquake seismology involves the nature of short-term processes that ultimately trigger slip on a given fault. The Earth's crust is pervasively laced with faults, and multiple lines of evidence indicate that the brittle crust is in some sense critically stressed and in a state of incipient failure nearly everywhere (Zoback and Zoback, 2002). Because long-wavelength stresses associated with the relative motion of major tectonic plates accumulate relatively slowly (rates of $\sim 10^{-5}$ MPa yr^{-1} or less), regional stresses may remain incrementally below the frictional strength of faults (typically of the order of ~10–100 MPa) for periods ranging from decades to centuries. Insight into the triggering question thus depends on understanding the influence of short-term, short-wavelength fluctuations in both the stress field and fault strength, such that the local stress state temporarily exceeds the local failure threshold leading to slip nucleation and an earthquake.

Sources of short- to mid-term stress fluctuations in the crust are many. Those most likely to be significant at seismogenic depths include (1) other earthquakes in the crust (Das and Scholz, 1981; Freed, 2005; Steacy *et al.*, 2005; Harris, 1998); (2) magmatic intrusions (Savage and Clark, 1982; Hill *et al.*, 2002; Manga and Brodsky, 2006); (3) anthropogenic activities such as reservoir-filling, mining, and fluid injection or withdrawal (McGarr *et al.*, 2002); (4) solid earth tides and ocean loading (Cochran *et al.*, 2004); and (5) seasonal meteorological factors such as snow loading and groundwater recharge (Christiansen *et al.*, 2005). Local fluctuations in fault strength may result from changes in pore fluid pressure within a fault zone (Lockner and Beeler, 2002), the nonlinear response of fault-zone friction to small perturbations in local stresses (Dieterich, 1979; Johnson and Jia, 2005), and subcritical crack growth (Atkinson, 1984).

In this chapter we are concerned with slip nucleation triggered either directly or indirectly by dynamic stress. Our principal focus will be on the evidence for triggering by dynamic stresses in the form of seismic waves (earthquake–earthquake interactions), but we also note recent evidence for tidal triggering as a quasi-static, low-amplitude reference point for the

spectral range of oscillatory stressing with the potential for dynamic triggering. Because many sites displaying remotely triggered seismicity are volcanic and geothermal areas, our scope includes the response of magmatic and hydrothermal systems to dynamic stresses as potential sources for locally triggered seismicity. We conclude with a perspective on key challenges to be met in advancing our understanding of dynamic triggering and its implications for active crustal processes.

4.09.1.1 Earthquake–Earthquake Interaction Modes

Earthquake–earthquake interactions in the form of mainshock–aftershock sequences have been recognized since the emergence of seismology as a quantitative science (Omori, 1894). With the global expansion of high-quality seismic networks coupled with the capabilities for resolving detailed spatial-temporal variations in deformation patterns afforded by satellite geodesy (continuous Global Positioning System (GPS) and interferometric synthetic aperature radar (InSAR) technologies, in particular), it has become increasingly clear that stress changes resulting from one earthquake are capable of triggering additional earthquake activity over a surprisingly wide range of distances and timescales. As currently understood, earthquake interactions are generally placed under one of three rather general stress transfer modes: (1) static, (2) quasi-static, or (3) dynamic.

4.09.1.1.1 Static and quasi-static stress triggering

Research on earthquake interactions associated with static stress triggering (the change in the static stress field from just before an earthquake to shortly after the dynamic stresses from the seismic waves have decayed away) began in the late 1960s and blossomed in the early 1990s with the growing availability of high-quality seismic and deformation data against which to test the patterns predicted by elastic dislocation models (Das and Scholz, 1981; Harris, 1998; King and Cocco, 2001; Stein, 1999). Shortly thereafter, extensions of the dislocation models were developed to account for gradual stress changes (quasi-static stress triggering) associated with viscous relaxation of the plastic lower crust and upper mantle in response to the sudden dislocation (an earthquake) across a fault in the overlying brittle crust (Pollitz and Sacks, 2002).

Both the static and quasi-stress interaction modes appeal to the permanent stress change produced by one earthquake nudging the stress field in the vicinity of a fault some distance away closer to the Coulomb failure threshold on that fault (e.g., Stein and Lisowski, 1983; Hudnut *et al.*, 1989; Harris, 1998; Kilb *et al.*, 2002; also *see* Section 4.09.4 and Chapter 4.08). This concept is commonly expressed in terms of a change in the Coulomb failure function, that is,

$$\Delta \mathrm{CFF} = \Delta\tau - \mu_s \Delta\sigma_n \quad [1]$$

where

$$\mathrm{CFF} = |\tau| - \mu_s\sigma_n - C \quad [2]$$

is the Coulomb failure function, τ and σ_n are the shear and normal stress components acting on the fault, respectively, μ_s is the static coefficient of friction, and C is the cohesive strength (e.g., Oppenheimer *et al.*, 1988; Harris, 1998; Kilb *et al.*, 2002; Cocco and Rice, 2002; also *see* Chapter 4.08). $\mathrm{CFF} = 0$ corresponds to Byerlee's law for frictional failure (Byerlee, 1980), and a positive change in CFF, or $\Delta\mathrm{CFF} > 0$ indicates that the stress state has moved incrementally toward $\mathrm{CFF} = 0$ and Coulomb failure.

Because static stress changes decay relatively rapidly with distance (as $\sim\Delta^{-3}$, where Δ is the distance from the epicenter), their triggering potential is generally regarded as limited to one or two source dimensions from a given earthquake. Viscoelastic relaxation following a large crustal earthquake is largely confined to the lower crust and asthenosphere such that quasi-static stress changes propagate as a two-dimensional stress change. Quasi-static stress changes thus decay more slowly with distance (as $\sim\Delta^{-2}$), their triggering potential thus extends to greater distances than static stress changes, and the relatively low viscoelastic propagation speeds result in delayed triggering (Pollitz and Sacks, 2002). These two interaction modes are the focus of Chapter 4.08 on stress transfer.

4.09.1.1.2 Seismic waves and dynamic stress triggering

The amplitudes of dynamic stresses propagating as seismic waves decrease relatively slowly with distance (as $\sim\Delta^{-2}$ for body waves and $\sim\Delta^{-3/2}$ for surface waves), and thus, their triggering potential extends from the near-field (the aftershock zone) to much greater distances than either the static or quasi-static stress changes as illustrated in **Figure 1** (Kilb *et al.*, 2000). Amplification by radiation directivity, which is not a factor in either static or quasi-static stress changes, can further enhance the amplitudes of dynamic stresses in a particular quadrant for earthquakes with unilateral

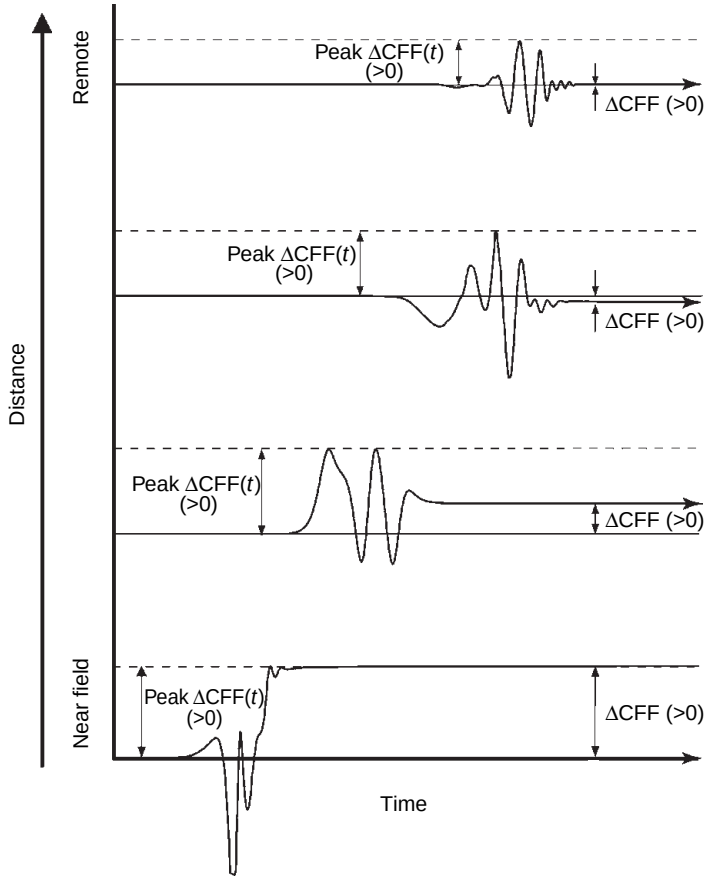


Figure 1 Diagram from Kilb *et al.* (2000) illustrating differences between static and dynamic stresses with distance in terms of idealized time histories for peak changes in dynamic (ΔCFF(*t*)) and static (ΔCFF) Coulomb stress. ΔCFF(*t*) is the time-dependent version of ΔCFF (eqn [1]). The double arrows show peak values for ΔCFF(*t*) and ΔCFF. Adapted from *Nature*.

fault rupture. Dynamic stresses are oscillatory, however, alternatively nudging the local stress field closer to and further from the local Coulomb failure stress over a range of frequencies leaving no permanent stress change. Indeed, it seems that Coulomb failure alone is not capable of explaining the full spectrum of observed dynamic triggering modes. In a number of cases, remotely triggered seismicity appears to be a secondary response to some more fundamental, aseismic process involving fluid activation or creep that was locally stimulated by the passing dynamic stresses.

4.09.2 Evidence for Dynamic Triggering

Conceptually, dynamic triggering is based on the inference of a causal link between dynamic stresses (seismic waves) from a large earthquake propagating through a given site (event '*a*') and the onset of local earthquake activity at that site (event '*b*'). Credible evidence in support of such a causal link emerges with multiple, well-documented observations that event '*b*' follows event '*a*' within some 'reasonable' time interval, Δt (**Figure 2**). Statistical methods, if carefully formulated, can add backbone to malleable qualifiers such as 'reasonable' and 'sufficiently', thereby enhancing objective credibility of this inference when well-posed tests indicate that the temporal sequence '*a*' then '*b*' cannot be dismissed as chance coincidences between random, statistically independent processes.

A number of tests for the statistical significance of dynamic triggering appear in the literature, the most common of which is the β-statistic of Matthews and Reasenberg (1988). As an example, **Figure 3** shows the β-statistic maps Gomberg *et al.* (2001) generated to evaluate patterns of dynamic triggering by the $M_W = 7.4$ Landers earthquake of 1992 and the

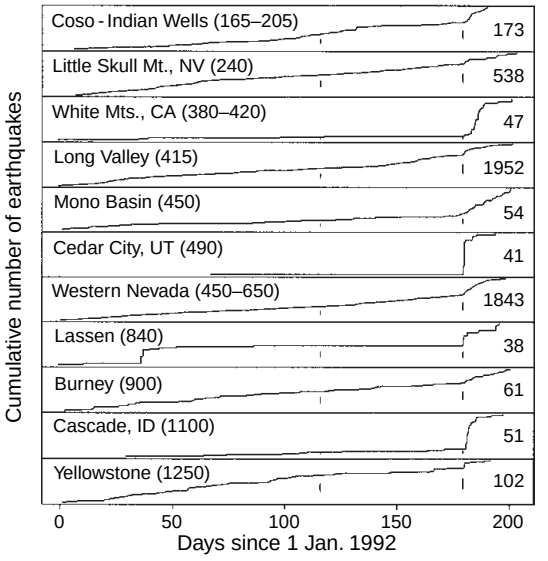


Figure 2 Evidence for remote dynamic triggering by seismic waves from the $M_W = 7.4$ Landers, CA, earthquake of 28 June 1992 in a plot of cumulative number of locatable earthquakes at 11 sites in the western United States for 211 days beginning with 1 January 1992 (Hill *et al.* 1993). Numbers in parentheses are distances in kilometers from the Landers epicenter. Numbers at the right indicate the total number of earthquakes in the respective areas for the entire time interval. The short vertical bars at day 114 mark the $M_W = 7.1$ Petrolia earthquake of 25 April 1992 and those at day 180 mark the $M_W = 7.4$ Landers earthquake. Reproduced from *Science*.

$M_W = 7.1$ Hector Mine earthquake of 1999 in southern California. Under this test, the value of β represents the number of standard deviations by which the seismicity rate during a specified time interval (2 weeks in the case of **Figure 3**), following a dynamic stressing event, exceeds an estimate of the background seismicity rate for a specified area. Values for $\beta > 2$ (two standard deviations) are generally considered to be statistically significant. Here, it is important to note that many of the isolated areas that show up as red ($\beta > 2$) in **Figure 3** do not correspond to areas of dynamic triggering from either the Landers or Hector Mine earthquakes. In Appendix 1, we describe this test and its limitations when applied to dynamic triggering and provide references to alternative approaches. In essence, elevated β-statistics are permissive of dynamic triggering, but they are not sufficient by themselves to establish compelling statistical significance. They can, however, serve as a useful exploratory tool. This cautionary note underscores the importance of understanding the assumptions underlying any test for statistical significance and the implications inherent in the choice of parameters used in a specific test.

4.09.2.1 Early Inferences on Dynamic Triggering

In an unpublished note dated 1955, Charles Richter wrote "The reverse effect – a major earthquake triggering a minor shock – is most probably [sic] within the immediate aftershock are [sic], but essentially by elastic wave propagation may set off action at a greater distance. If the distant effect is large enough, it may itself act as a trigger, so that there may be relay action, in which some of the later events are larger" (S. Hough, personal communication, 2005). Many of the early published inferences on dynamic triggering were associated with investigations into acoustic emissions and high-frequency seismic noise (see, e.g., Armstrong and Stierman (1989), Nicolaev and Troitskii (1987), and references therein, and Galperin *et al.* (1990)) and concerns over the possibility that large, underground nuclear explosions might trigger a damaging earthquake (Emiliani *et al.*, 1969). At the time, these studies were greeted within the scientific community by attitudes ranging from mild interest to strong skepticism. Skepticism prevailed largely because of questions regarding the statistical significance of isolated observations together with the lack of a compelling physical model. The fact that some of these studies were a part of efforts focused on even more controversial topics such as anomalous animal behavior induced by acoustic emissions as a means of earthquake prediction (e.g., Armstrong and Stierman, 1989) probably contributed to the skeptical reception within the mainstream scientific community.

In the end, however, early investigations on dynamic triggering were hampered by limited data. The situation began changing with the rapid expansion of continuously recorded, telemetered seismic networks around the globe through the 1980s and 1990s coupled with improvements in real-time processing and a growing number of broadband, high-dynamic-range digital installations set the stage for capturing a range of interesting seismic signals missed by the earlier instrumentation.

4.09.2.2 Widespread Evidence for Remote Dynamic Triggering in Western North America

Between 1980 and the spring of 1992, four $M \sim 7$ earthquakes shook the western United States

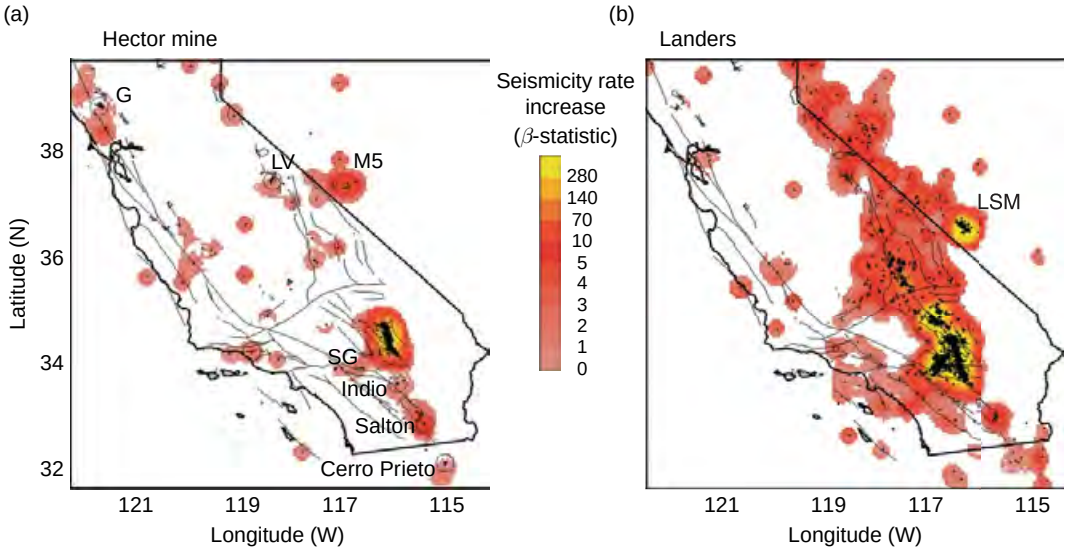


Figure 3 β-Statistic maps illustrating seismicity rate increases associated with the $M_W = 7.1$ Hector Mine earthquake of 16 October 1999 (a) and the $M_W = 7.4$ Landers earthquake of 28 June 1992 (b) from Gomberg *et al.* (2001). The plots show a spatially smoothed β-statistic over dimensions of 20 km for $M \geq 2.0$ earthquakes from the Northern California Seismic Network (NCSN) and Southern California Seismic Network (SCSN) catalogs. The rate change was calculated for a 2-week period following the respective mainshocks with respect to the background period 1987 through mid-1992 (pre-Landers). Thin lines are major fault traces. Circles in (a) mark sites of activity triggered by the Hector Mine earthquake. M5 in (a) reflects aftershock activity to an $M = 5.3$ earthquake that occurred ~2.5 months prior to the Hector Mine earthquake. G, the Geysers geothermal area; LV, Long Valley caldera; SG, San Gorgonio pass; LSM, Little Skull Mountain. Modified from *Nature*.

including the $M_W = 7.4$ Eureka, California earthquake of 8 November 1980 and the $M_W = 7.1$ Petrolia, California earthquake of 25 April 1992. Although many of the regional seismic networks were in place during this period, none of these earthquakes produced dynamic triggering that was noticed at the time.

4.09.2.2.1 The $M_W = 7.3$ Landers earthquake of 28 June 1992

On 28 June 1992, 2 months after the Petrolia earthquake, the $M_W = 7.3$ Landers earthquake ruptured a 70 km length of the Mojave Desert in southern California. Over the next few hours, it became clear that seismicity rates had increased at a number of sites across western North America at distances ranging from 200 km to as much as 1250 km (17 source dimensions). These sites included Long Valley caldera; Lassen Peak, Burney, CA; the Wasatch front in central Utah; Cascade Idaho; and Yellowstone National Park. Documented seismicity rate increases began within minutes to 33 h following the Landers mainshock (**Figures 2–4**). This large number of independent occurrences provided overwhelming evidence that remote triggering by dynamic stresses is indeed a viable process in the Earth (Hill *et al.*, 1993; Gomberg, 1996).

The Landers mainshock resulted from a unilateral rupture propagating to the north–northwest along a series of north–northwest striking dextral fault segments. All of the recognized sites of dynamic triggering were north of the Landers epicenter, suggesting that amplification enhanced by rupture directivity may influence the distribution of dynamic triggering – a suggestion borne out by subsequent instances of dynamic triggering (Gomberg *et al.*, 2001). Indeed, the absence of notable triggering by the four earlier $M \sim 7$ earthquakes may reflect the fact that they did not produce significant rupture directivity or, in the case of the $M = 7.4$ Eureka earthquake of 1980, that rupture propagation was directed to the southwest away from the continental United States. With the notable exception of a $M = 5.6$ earthquake beneath Little Skull Mountain in southern Nevada 240 km north of the Landers epicenter, the earthquakes triggered by Landers had magnitudes of $M \sim 3$ or less with a tendency for the maximum magnitude to decrease with increasing distance for the Landers epicenter (Anderson *et al.*, 1994).

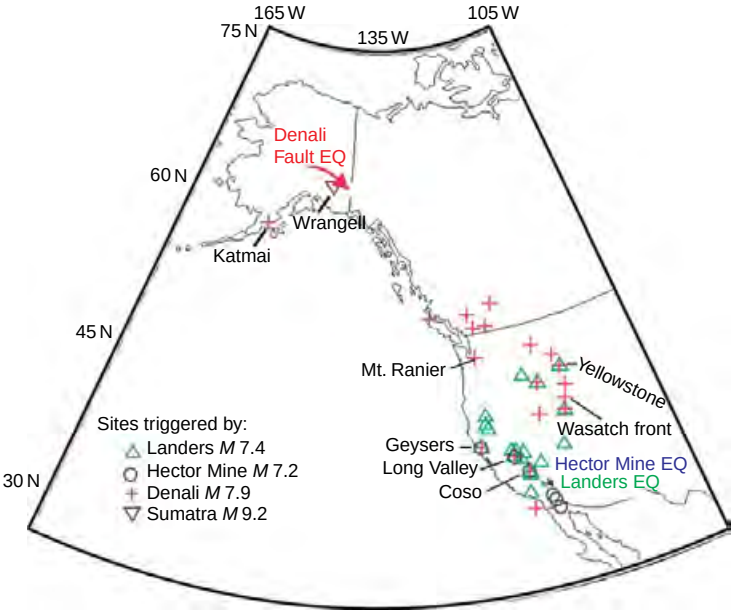


Figure 4 Map showing sites with triggered seismicity in western North America from the Landers (green triangles), Hector Mine (blue circles), and Denali Fault (red crosses) earthquakes. Mainshock ruptures are indicated by corresponding colored lines with arrow indicating rupture direction. Inverted gold triangle is Mt. Wrangell volcano, which showed a seismicity decrease following the Denali Fault earthquake (Sanchez and McNutt, 2004) but responded to dynamic stresses from the $M_W = 9.1$ Sumatra–Adaman Islands earthquake of 26 December 2004 with several brief bursts of triggered seismicity (West *et al.*, 2005).

All of the remote triggering occurred in areas of elevated background seismicity in transtensional tectonic settings. This was underscored, for example, by the absence of a triggered response along the creeping section of the San Andreas Fault through central California and Parkfield that, although a persistent source of small earthquakes, is a transpressional stress regime (Spudich *et al.*, 1995). Many of the triggered sites were areas of geothermal activity and/or Quaternary volcanism (Hill *et al.*, 1993; Anderson *et al.*, 1994; Gomberg, 1996).

4.09.2.2.2 The $M_W = 7.1$ Hector Mine earthquake of 16 October 1999

On 16 October 1999, an $M = 7.1$ earthquake ruptured a 40 km length of the Mojave Desert along a series of faults located just 20 km east of the Landers rupture (**Figures 3** and **4**). In this case, fault rupture was bilateral but with the dominant rupture direction to the south–southeast. The most energetic triggered response to the Hector Mine dynamic stresses was in the Salton Trough south of the epicenter (**Figures 3** and **4**), thus adding weight to the idea that rupture directivity and dynamic stress amplitudes are important factors in determining the distribution of dynamic triggering (Gomberg *et al.*, 2001; Hough and Kanamori, 2002).

The Salton Trough is an extensional tectonic regime characterized by frequent earthquake swarms (**Figure 3**), geothermal areas, and Quaternary volcanism. The triggered response to the Hector Mine earthquake, which included three $M > 4$ earthquakes, was centered beneath (1) the northern end of the Salton Trough in the vicinity of Indio ($\Delta \sim 87$ km), (2) the south end of the Salton Sea in the vicinity of the Obsidian Buttes volcanic field ($\Delta \sim 165$ km), and (3) in the Mexicali Valley in the vicinity of the Cerro Prieto geothermal field in Baja California at $\Delta \sim 260$ km (Hough and Kanamori, 2002; Glowacka *et al.*, 2002). Three sites northwest of the epicenter showed a weaker triggered response: the Coso volcanic field at $\Delta \sim 250$ km (Prejean *et al.*, 2004), the north flank of Mammoth Mountain at the west margin of Long Valley caldera at $\Delta \sim 450$ km (Johnston *et al.*, 2004a), and the Geysers geothermal field at $\Delta \sim 750$ km (Gomberg *et al.*, 2001).

With the exception of Mammoth Mountain and Cerro Prieto, the onset of the triggered seismicity coincided with arrival of the surface waves from the Hector Mine earthquake. The onset of seismicity

with respect to surface-wave arrivals was apparently delayed by $\Delta t \sim 20$ min at Mammoth Mountain and $\Delta t \sim 2$ h at Cerro Prieto. Notably, the triggered response at both Mammoth Mountain and Cerro Prieto included deformation transients detected by borehole dilatometers and tiltmeters, respectively (Johnston *et al.*, 2004a; Glowacka *et al.*, 2002).

4.09.2.2.3 The $M_W = 7.9$ Denali Fault earthquake of 3 November 2002

The $M = 7.9$ Denali Fault earthquake of 3 November 2002 produced the most widely recorded remote dynamic triggering through mid-2006. The earthquake, which was centered 65 km east of Denali National Park, Alaska, resulted from a complex rupture with maximum surface displacements reaching 8.8 m (Eberhart-Phillips *et al.*, 2003; Rowe *et al.*, 2004). With a seismic moment ($M_0 \sim 4 \times 10^{20}$ N m) comparable to that of the 1906 San Francisco earthquake, it was the largest earthquake to strike North America since the $M_W = 9.2$ Anchorage megathrust earthquake of 27 March 1964 (Plafker, 1969). The unilateral rupture produced pronounced directivity with peak amplification of dynamic stresses directed to the southeast through British Columbia and the western United States. With the exception of the Katmai volcanic cluster, 740 km southwest of the epicenter, all recognized sites of dynamic triggering were located southeast of the epicenter in the sector of maximum directivity that spanned the areas responding to the Landers and Hector Mines earthquakes (**Figure 4**). In all cases, the onset of dynamic triggering developed as a seismicity rate increase during passage of the Denali Fault earthquake Love and Rayleigh waves. In three cases, Mount Rainer, Long Valley caldera, and Yellowstone, delayed earthquake swarms followed the initial surge in triggered seismicity by $\Delta t \sim 2.5$ h, 23.5 h, and 8 days, respectively (Prejean *et al.*, 2004; Husen *et al.*, 2004b).

The Yellowstone volcanic system and the central Utah seismic zone, where peak dynamic stresses ranged from ~0.2 to 0.3 MPa, provided the most energetic recorded triggered response to the Denali earthquake (Husen *et al.*, 2004b; Pankow *et al.*, 2004). In these locations, the seismicity rate increases were clearly evident in earthquake catalogs with compelling β-statistics of the order of ~10–30. Yellowstone responded with pronounced changes in the hydrothermal system in addition to the widespread seismicity increase (Husen *et al.*, 2004a). The former included eruption pattern changes in 12 out of 22 instrumented geysers as well as transient temperature changes and boiling in a number of hot springs. Notably, geysers that responded to previous nearby (distances <200 km) $M \sim 7$ earthquakes did not respond to the Denali earthquake (Husen *et al.*, 2004b). Seismicity increased over much of the Park with some 250 located earthquakes detected within the first 24 h after the Denali earthquake, and it remained elevated over the next 30 days. Bursts of triggered seismicity within Yellowstone caldera tended to be concentrated in the vicinity of major geothermal areas.

Seismic activity increased at a number of sites along the Intermountain Seismic Belt (ISB) through central Utah following passage of the Denali surface waves (Pankow *et al.*, 2004). Tectonically, the Utah section of the ISB is characterized by east–west extension accommodated normal to oblique slip along northerly striking faults. The clusters of triggered seismicity through central Utah showed no obvious correlation with the Quaternary volcanic vents common in the southern part of the area. Many hot springs along the ISB appear to be associated with deep fault-zone circulation rather than mid-crustal magmatic sources. High-resolution hypocentral locations showed that the triggered clusters tended to be spatially separated from background seismicity. The triggered activity, which included an $M = 3.2$ earthquake, decayed to background levels over about 25 days. A site of triggered seismicity near Cedar City in southwestern Utah also responded with triggered seismicity following the 1992 Landers earthquake.

The remaining instances of dynamic triggering by the Denali Fault earthquake were not evident in regional earthquake catalogs but were uncovered by examining high-pass filtered seismograms from broadband and strong motion instruments from individual seismic stations within these networks. The results revealed six sites in southern British Columbia (Gomberg *et al.*, 2004), two sites in western Idaho (Husker and Brodsky, 2004), five sites along the west coast states including Mount Rainer in Washington, the Geysers geothermal field in northern California, Mammoth Mountain-Long Valley caldera in eastern California, and the Coso volcanic field in southern California (Prejean *et al.*, 2004). Two earthquakes with magnitudes $M = 1.8$ and 2.5 located ~125 km offshore in the southern California borderland (4003 km from the Denali epicenter) that occurred during passage of the Denali surface wave coda and may also represent triggered activity, although the statistical significance of such isolated events is difficult to establish (Prejean *et al.*, 2004).

The sole instance of dynamic triggering orthogonal to the sector of rupture directivity involved a burst of small ($M \leq 2.0$) earthquakes in the Katmai volcanic cluster 740 km southwest of the Denali epicenter that began during the surface wave train and died out over the next several hours (Moran *et al.*, 2004). Indeed, it appears that seismicity rates may have temporarily slowed at Mt. Wrangell and Mt. Veniaminof volcanoes following the Denali earthquake at distances of 247 and 1400 km southwest of the epicenter, respectively (Sanchez and McNutt, 2004). The Katmai volcanic field appears to be particularly susceptible to dynamic triggering in that it has responded to four previous earthquakes in the $M = 6.8$–7.0 range that were located at distances between 60 and 160 km (Moran *et al.*, 2004; Power *et al.* 2001).

4.09.2.3 The Search for Additional Evidence for Dynamic Triggering

The compelling evidence for dynamic triggering provided by the 1992 Landers earthquake spawned a search through both historic and instrumental records for evidence of dynamic triggering from previous earthquakes as well as for new instances of dynamic triggering associated with new large earthquakes (**Table 1**). As the evidence accumulates, it is becoming clear that dynamic triggering at remote distances is a relatively common phenomenon.

4.09.2.3.1 Anecdotal evidence for pre-instrumental dynamic triggering

In an extensive search through archive newspapers and correspondence for felt reports, Hough (2001) and Hough *et al.* (2003) find evidence for dynamic triggering at distances up to1000 km or more by two of the four $M_W > 7$ New Madrid, Missouri earthquakes of 1811–1812 and the $M_W > 7$ Charleston, South Carolina, earthquake of 1886. These results are particularly intriguing because they offer evidence for dynamic triggering in a low-seismicity-rate, low-strain-rate intraplate environment characterized by a transpressional stress regime (see Reinecker *et al.*, 2004).

Steeples and Steeples (1996) and Meltzner and Wald (2003) investigated the distribution of aftershocks and possible triggered earthquakes for the great ($M_W \sim 7.8$) San Francisco earthquake of 1906. They found evidence for nine earthquakes in the magnitude range $M \sim 4$ to ~ 6 within the first 48 h after the 1906 mainshock. The most distant, an $M \sim 4$ earthquake in western Arizona at $\Delta \sim 900$ km, occurred during the seismic waves from the 1906 mainshock and is a likely candidate for dynamic triggering. Earthquakes marginally beyond the aftershock zone and thus potential candidates for remote dynamic triggering include an $M \sim 6.1$ event near the south end of the Salton Sea ($\Delta \sim 700$ km, $\Delta\tau \sim 11.3$ h) and $M \sim 3.5$ and $M \sim 4.5$ earthquakes in western Nevada at $\Delta \sim 400$ and 410 km and $\Delta\tau \sim 32.8$ and 39.0 h, respectively.

4.09.2.3.2 Instrumental evidence from around the globe

Reviewing instrumental records from the dense seismic network around the Geysers geothermal field in the northern California Coast Ranges between 1988 and 1994, Stark and Davis (1995) and Gomberg and Davis (1996) found instances of abrupt, temporary

Table 1 Major earthquakes triggering multiple sites

Earthquake	Magnitude	Date	N sites	Max range	Trig Mmax	Ref.
Landers, CA	$M_w = 7.3$	28 Jun. 92	14	~1250 km	5.6	Hill *et al.*, 1993
Hector Mine, CA	$M_w = 7.1$	16 Oct. 99	5	~750 km?		Glowacka *et al.*, 2002; Gomberg *et al.*, 2001
Izmit, Turkey	$M_w = 7.4$	17 Aug. 00	Widespread in Greece	~1000 km	3.8	Brodsky *et al.*, 2000
Denali Fault, AK	$M_w = 7.9$	11 Mar. 02	~18	> 3660 km	3.2	Prejean *et al.*, 2004
Sumatra-Andaman Island	$M_w = 9.0$	26 Dec. 04	2	~11 000	<2?	West *et al.*, 2005; Ukawa, personal communication, 2005
Tokachi-oki	$M_w = 8.1$	25 Sep. 03	3	~1400 km	DLF	Miyazawa and Mori, 2005
SW Siberia	$M_w = 7.3$	27 Sep. 03	3	~4000 km	DLF	Miyazawa and Mori, 2005

DLF, deep, low-frequency tremors.

increases in the rate of small earthquakes ($M < 3$) to dynamic stresses from eight regional earthquakes with magnitudes between $M = 6.6$ and 7.7, and at distances between $\Delta = 212$ and 2500 km. The Geysers geothermal field, which seems particularly susceptible to dynamic triggering, subsequently responded to the 1999 Hector Mine earthquake (Gomberg *et al.*, 2001) and the 2002 Denali Fault earthquake located 3120 km to the northwest (Prejean *et al.*, 2004).

Just over 2 months before the Landers earthquake, an $M_W = 5.4$ earthquake on 13 April 1992 occurred within the Roer Valley Graben near Roermond, The Netherlands (Camelbeeck *et al.*, 1994). A series of earthquakes 40 km to the southeast (along strike with the northwest striking normal faults forming the graben) began some 3 h after the mainshock. Camelbeeck *et al.* (1994) suggest that these represent remotely triggered earthquakes.

Brodsky *et al.* (2000) document an increase in $M \geq 2$ seismicity over much of continental Greece abruptly as the surface waves from the $M = 7.4$ Izmit, Turkey, earthquake of 17 August 1999, swept through the region at distances from $\Delta \sim 400$ to $\sim$1000 km. The largest of the triggered earthquakes exceeded the catalog completeness threshold $M > 3.5$, and these larger events tended to be located away from areas characterized by high background seismicity. Greece is dominated by pervasive normal faulting with widely distributed hot springs, which derive their heat from deep crustal circulation along the normal faults rather than mid-crustal magma bodies (Brodsky *et al.*, 2000). Greece has no history of volcanism.

Mohamad *et al.* (2000) describe evidence for remotely triggered seismicity in a localized cluster along the Syrian–Lebanese border from the dynamic stresses generated by the $M_S = 7.3$ Gulf of Aqaba earthquake of 22 November 1995 that resulted from strike-slip rupture along a segment of the sinistral Dead Sea Fault some 500 km to the south. The triggered seismicity began 2.8 h after the mainshock and produced over 20 small earthquakes over the next 3.5 h, the largest of which was an $M = 3.7$ event. The triggered cluster occurred near a restraining bend in the sinistral Yammouneh fault some 50 km north of the volcanic and geothermal area in the Golan Heights.

The epicenters of $M_W = 6.5$ and 6.4 earthquakes in the South Iceland Seismic Zone on 17 and 21 June 2000, respectively, were separated by just 17 km in an east–west direction. Both earthquakes involved dextral slip on north-striking planes. As described by Arnadottir *et al.* (2004), the first and largest of these earthquakes triggered widespread seismicity on the Reykjanes Peninsula to distances of $\sim$100 km to the west as well as significant slip on at least three faults in the Reykjanes Peninsula while the second slightly smaller earthquake, although 17 km further west, did not. The triggered seismicity included three $M \sim 5$ earthquakes that followed the $M = 6.5$ event by seconds to minutes. Arnadottir *et al.* (2004) conclude that these $M \sim 5$ earthquakes were beyond the range for Coulomb failure due to static stress changes from the $M = 6.5$ mainshock and thus that they were likely triggered by dynamic stresses. Their analysis of continuous GPS geodetic data spanning the Reykjanes Peninsula indicates that the geodetic moment for the second of the triggered $M \sim 5$ earthquakes ($\Delta \sim 78$ km) was an order of magnitude larger than its seismic moment and that the large Coulomb stress change from this largely aseismic, slow earthquake was sufficient to trigger the third and most distant $M \sim 5$ event some 4 min later through quasi-static stress transfer. This, then, provides a possible example of earthquake triggering as a secondary response to a locally triggered deformation transient consistent with the slow fault slip model proposed by Anderson *et al.* (1994).

Singh *et al.* (1998) examined Mexican seismograms and earthquake catalogs spanning the interval 1920 through 1957 for evidence of dynamic triggering in the Valley of Mexico within the Trans Mexican Volcanic Belt from large (M 7.6–8.0) subduction-zone earthquakes along the southwest coast of Mexico. They identified 'abrupt' (within 2 days) seismicity increases following 7 out of 13 subduction-zone earthquakes at distances of $\Delta \sim 303$ to $\sim$588 km. In three cases, they identified seismicity increases after a delay of some 30 days following the subduction-zone earthquake. Three of the subduction-zone earthquakes produced no recognizable seismicity rate increases. The latter included an $M = 8.0$ earthquake at a distance of $\Delta \sim 405$ km and an $M = 7.7$ earthquake at $\Delta \sim 347$ km. The three delayed responses of $\Delta t \geq 30$ days certainly raise some interesting questions regarding statistical significance and a causal link between dynamic stresses and triggered seismicity.

In a study of possible dynamic triggering in the Taiwan region, Wen *et al.* (1996) examined the catalog for the Taiwan Telemetered Seismic Network (TTSN) from 1973 to 1994 for evidence of seismicity increases following 12 regional earthquakes with

magnitudes from $M = 6.5-7.1$ at distances of $\Delta \sim 138$–2959 km. In nine cases they found a fourfold increase in the number of $M \geq 4.5$ earthquakes in the 15-day period following an $M \geq 6.5$ earthquake with respect to the 15 preceding days and in 10 cases a sevenfold increase in $M_L \geq 4.0$ earthquakes. The significance of this result remains a question. Taiwan is dominated by a compressional–transpressional stress regime (Reinecker *et al.*, 2004) although, locally, the northern end of the island, which includes the Quaternary Tatun volcanic group and associated geothermal activity, is an extensional regime (Song *et al.*, 2000). Wen *et al.* (1996) do not address the spatial distribution of the triggered seismicity.

Somewhat surprisingly, evidence in the published literature for dynamic triggering in Japan is sparse (Harrington and Brodsky, 2006). Japan is well monitored by dense seismic and deformation networks, and it is populated by many active volcanoes and geothermal systems. Harrington and Brodsky (2006) suggest that this insensitivity to triggering may result from compressional tectonics that dominates the country or perhaps the high occurrence rates of large mainshocks that serve to frequently disrupt blockages in fracture systems, thereby preventing a buildup of differential pore pressures between hydrologic units that might otherwise respond to dynamic stresses. Exceptions appear to be limited to volcanic systems on Kyushu in southern Japan, which is an area of extensional tectonics. Miyazawa *et al.* (2005) describe evidence for multiple increases in isolated tremor and microearthquakes at Aso volcano in central Kyushu, Japan, following large, distant earthquakes including the $M_W = 7.7$ Chi-Chi, Taiwan earthquake of 20 September 1999. Harrington and Brodsky (2006) show evidence for increased seismicity on Kyushu following the 26 December 2004 $M_W = 9.0$ Sumatra earthquake. Two particularly intriguing contributions from Japan describe evidence for dynamic triggering of deep, low-frequency earthquakes along the Nankai subduction zone beneath Honshu and bursts of shallow microearthquakes on Iwo Jima in the Volcano Islands group situated between the Marianas and Izu island arcs some 1250 km south of Tokyo.

The island of Iwo Jima represents the resurgent dome of a large submarine caldera. (Iwo Jima, or Sulfur Island, is best known as the site of a major World War II battle.) It has a well-developed geothermal system and has produced a Holocene volcanic eruption and several recent phreatic explosions. Ukawa *et al.* (2002) examined Iwo Jima seismograms for 21 $M > 7$ earthquakes located within 3000 km of the island and found four instances when local seismicity increased during passage of the surface waves from the distant earthquakes. In each case, the triggered response was characterized by an abrupt increase in local microearthquake activity that died out over the next 6–15 min. Iwo Jima seismograms showed a similar triggered response to the surface waves from the great $M = 9.1$ Sumatra–Andaman Islands earthquake of 26 December 2004 (Ukawa, personal communication, 2005).

The evidence for dynamic triggering of deep low-frequency tremor (DLF) beneath Japan is notable in that it, together with the Tonga example described in the next paragraph, is the only documented instance of dynamic triggering not confined to the upper, brittle crust. DLF tremor is associated with subduction of the Philippine Sea plate beneath southwestern Japan (Obara, 2002). This nonvolcanic tremor consists of clusters of low-frequency (1–10 Hz) earthquakes near the base of the crust at depths from 20 to 30 km and appears to be associated with the upward flow of hydrous fluids liberated by dehydration of the subducting slab. Obara (2002) suggests that fluctuations in the DLF activity in the spring of 2001 may have been modulated by local $M \sim 6$ earthquakes at distance of 40–50 km. Miyazawa and Mori (2005) applied a β-statistic analysis to the DLF activity and found convincing evidence of a rate increase as the surface waves from the $M_W = 8.1$ Tokachi-oki earthquake of 25 September 2003, some 1000 km to the northwest, propagated through the area. They found a similar response to the surface waves from an $M_W = 7.3$ earthquake in Siberia 40 h later (27 September) located ~4000 km to the west. Subsequently, Miyazawa and Mori (2006) document that the Rayleigh waves from the 2004 Sumatra–Andaman Islands earthquake modulated DLF along a 500-km-long zone beneath southwestern Japan suggesting that dilatational stresses may be important in the triggering process.

An $M_W = 7.6$ earthquake on 19 August 2002 at a depth of ~665 km beneath the Tonga subduction zone was followed by $M_b = 5.9$ and $M_W = 7.7$ earthquakes in the same depth range with delays of 2.2 and 7.4 min, respectively. As described by Tibi *et al.* (2003), the latter two earthquakes occurred in a previously aseismic volume located ~290 km southwest of the initial $M_W = 7.6$ event. They find that the static Coulomb stress change at this distance from the initial $M_W = 7.6$ earthquake is small and conclude

that these earthquakes were most likely dynamically triggered by the body waves from the $M_W = 7.6$ earthquake. A similar sequence of deep triggering may be represented by an $M_W = 7.1$ earthquake in a previously aseismic volume that followed an $M_W = 6.8$ slab earthquake at a distance of 257 km by 25 min (Tibi *et al.*, 2003).

4.09.2.3.3 Aftershocks and stress triggering at near to intermediate distances

Stress triggering at near and intermediate distances from a large earthquake is tied to the long-standing question of aftershock generation (Nur and Booker, 1972; Gomberg *et al.*, 2003; Kilb *et al.*, 2002). The most intense aftershock activity generally occurs within a source dimension of the mainshock rupture (the near field), where static and dynamic stresses have comparable amplitudes and minimal temporal separation (**Figure 1**). Distal portions of an aftershock zone, however, commonly extend one to two source dimensions beyond the rupture zone in the range where static and dynamic stresses begin to assume separate identities. Efforts to distinguish between static and dynamic stress triggering in aftershock zones have sought evidence in spatial distribution of aftershocks with respect to stress shadows predicted by static stress transfer models (Felzer and Brodsky, 2005; Kilb *et al.*, 2002; also *see* Chapter 4.08) as well as for the influence of rupture directivity (Gomberg *et al.*, 2003) and the variation in peak dynamic stress amplitudes with distance (Felzer and Brodsky, 2006). Pollitz and Johnston (2006) point to the relative absence of aftershocks following slow earthquakes along the central section of the San Andreas Fault compared with typical earthquakes of similar location and seismic moment as evidence for the importance of dynamic stresses in aftershock production. Based on these studies, it seems clear that dynamic stresses are important in aftershock generation at all distances, particularly for short times (hours) following the mainshock.

Hough (2005) explores the dynamic triggering potential of moderate earthquakes ($M < 7$) in the transition between the distal aftershock zone and remote distances by stacking β-statistic maps for 15 moderate ($M = 5.3$–7.1) earthquakes in central and southern California. She found evidence for slightly elevated seismicity levels in the 140–200 km distance range (more than two source dimensions for $M < 6$ earthquakes) during the month following the respective mainshocks and suggests that this may represent seismicity dynamically triggered by the large-amplitude S-waves critically reflected from the base of the crust (SmS). Some of these elevated seismicity levels are located within the transpressive stress regimes adjacent to the San Andreas Fault system. Because the elevated β-statistics in these examples are modest ($\beta \sim 2$; see Appendix 1), this intriguing result remains ripe for further testing.

This is currently a rapidly evolving research topic with the relative merits of static versus dynamic aftershock triggering the subject of energetic exchanges between advocates of one or the other viewpoint (Voisin *et al.*, 2004; Harris and Day, 1993; Harris *et al.*, 2002; Stein, 1999). Accumulating evidence suggests that dynamic triggering does indeed have a role in aftershock generation as well as in complex rupture processes in which major earthquakes consist of multiple large subevents (e.g., Aagaard *et al.*, 2004; Rybicki *et al.*, 1985; Antonioli *et al.*, 2002; Voisin *et al.*, 2000).

4.09.2.3.4 Triggering by solid Earth tides

One of the long-standing puzzles in seismology has been why the periodic stressing of the Earth's crust by solid Earth tides appears to have little or no influence on temporal patterns of earthquake occurrence – particularly, given the evidence that broad sections of the crust are in a state of incipient failure. The generally acknowledged exceptions appeared to be limited to shallow earthquake swarm activity in volcanic and geothermal areas (Klein, 1976; Tolstoy *et al.*, 2002). Recently, however, Tanaka *et al.* (2004) and Cochran *et al.* (2004) present convincing evidence that the solid Earth tides in combination with ocean loading tides do indeed modulate the occurrence of crustal thrust earthquakes in convergent margins around the Pacific basin at the 10–20% level. The key appears to lie in correlating well-constrained focal mechanisms from a massive number of earthquakes (tens of thousands) with the phase of the tidal cycle that augments the local tectonic stress directions. These results are consistent with inferences based on laboratory measurements that cyclical stressing at the 0.001–0.004 MPa level should modulate the occurrence of background seismicity (Beeler and Lockner, 2003; Lockner and Beeler, 1999). The evidence for tidal triggering on reverse faults in Japan (Tanaka *et al.*, 2004) presents an intriguing contrast to the apparent insensitivity of the many volcanic and geothermal centers in Japan to remote dynamic triggering by large earthquakes (Harrington and Brodsky, 2006). While strictly

speaking solid Earth tides represent a quasi-static process (the inertial component in the equations of motion are negligible for solid Earth tides), they suggest a lower bound for amplitudes (~0.001 MPa or ~0.01 bar) that low-frequency dynamic stresses must reach to have the potential for dynamic triggering.

4.09.3 Triggered Response Characteristics

Unraveling the physics behind dynamic triggering requires identifying key characteristics of the triggered response and their variations from site to site for a given earthquake as well as their variations at a given site for different source earthquakes. In this section, we highlight some of the more obvious of these characteristics based on published descriptions of dynamically triggered seismicity.

4.09.3.1 Tectonic Setting

The tectonic setting and stress regime have a first-order effect on the style of earthquake activity in any given area. Accumulating evidence indicates that these environmental factors also have an influence on patterns of remotely triggered seismicity.

4.09.3.1.1 Crustal triggering

Most well-documented, instrumentally recorded instances of dynamic triggering within the brittle, seismogenic crust (depth less than 15–20 km) are associated with extensional or transtensional tectonic regimes. In many of these cases, the triggering coincides with areas of elevated background seismicity and commonly (but not exclusively) with geothermal areas and areas of Quaternary to Recent volcanism.

Extensional and transtensional stress regimes have two closely related properties that might enhance their susceptibility to dynamic triggering. First, the least principal stress has a subhorizontal orientation. Because extension cracks tend to form in planes normal to the least principal stress, extensional stress regimes facilitate the upward migration of crustal fluids from warmer, higher-pressure conditions at depth toward the surface through vertically oriented cracks. One consequence of this property is the propensity for geothermal areas (and volcanism) to be concentrated in extensional regimes. Second, as illustrated in **Figure 5(a)**, faults in an extensional regime are intrinsically weaker than those in compressional regimes given a uniform coefficient of friction and Andersonian faulting (Sibson, 1982). Hough and Kanamori (2002), for example, suggest that elevated pore pressures and temperatures associated with geothermal areas should enhance this difference (illustrated by the dashed Mohr's circle in **Figure 5(a)**).

There are, however, exceptions to this pattern. One involves the evidence for remote triggering based on pre-instrumental descriptions of felt shaking in the transpressional regimes of Ohio River Valley area and Atlantic coast states following the 1811–1812 New Madrid, Missouri, earthquakes and the Charleston, South Carolina earthquake of 1887 (Hough, 2001; Hough *et al.*, 2003). Another involves Hough's (2005) evidence for apparent seismicity increases at distances of 70–200 km triggered by dynamic stresses associated with super-critical SmS waves from 15 moderate earthquakes ($M \sim 5$) in coastal California. In this case, several areas of slightly elevated β-statistic were located within the transpressional Coast Ranges. Hough *et al.* (2005) also suggest that SmS waves may have triggered an $M > 7$ earthquake a few minutes after the $M_W = 7.8$ earthquake of 1905 in the compressional Kangra region of India. However, we have yet to see evidence for remote dynamic triggering in well-instrumented compressional or stable intraplate regimes on a scale comparable to that observed in tectonically active extensional or transtensional regimes.

Subduction-related volcanic centers are located on convergent plate boundaries. Locally, however, they commonly coincide with a zone of crustal extension and transtensional deformation associated with upward flexure of the overriding plate. This is the case for the Katmai volcanic field in Alaska (Moran, 2003), for example, as well as most of the Cascade volcanoes in the Pacific Northwest states. Mt. Rainer volcano in Washington State is exceptional with its setting in a transpressional stress regime. Even here, however, focal mechanisms for earthquakes occurring within the edifice of the volcano or shallow crust immediately below the volcano show dominantly normal faulting (Moran *et al.*, 2000).

This raises a question regarding the meaning of the crust being in a state of incipient failure nearly everywhere (Zoback and Zoback, 2002). Within the context of **Figure 5(a)**, a weak, extensional crust offers no particular advantage over a strong compressional crust if incipient failure means that, on average, the stress level everywhere hovers below the failure strength by some common increment,

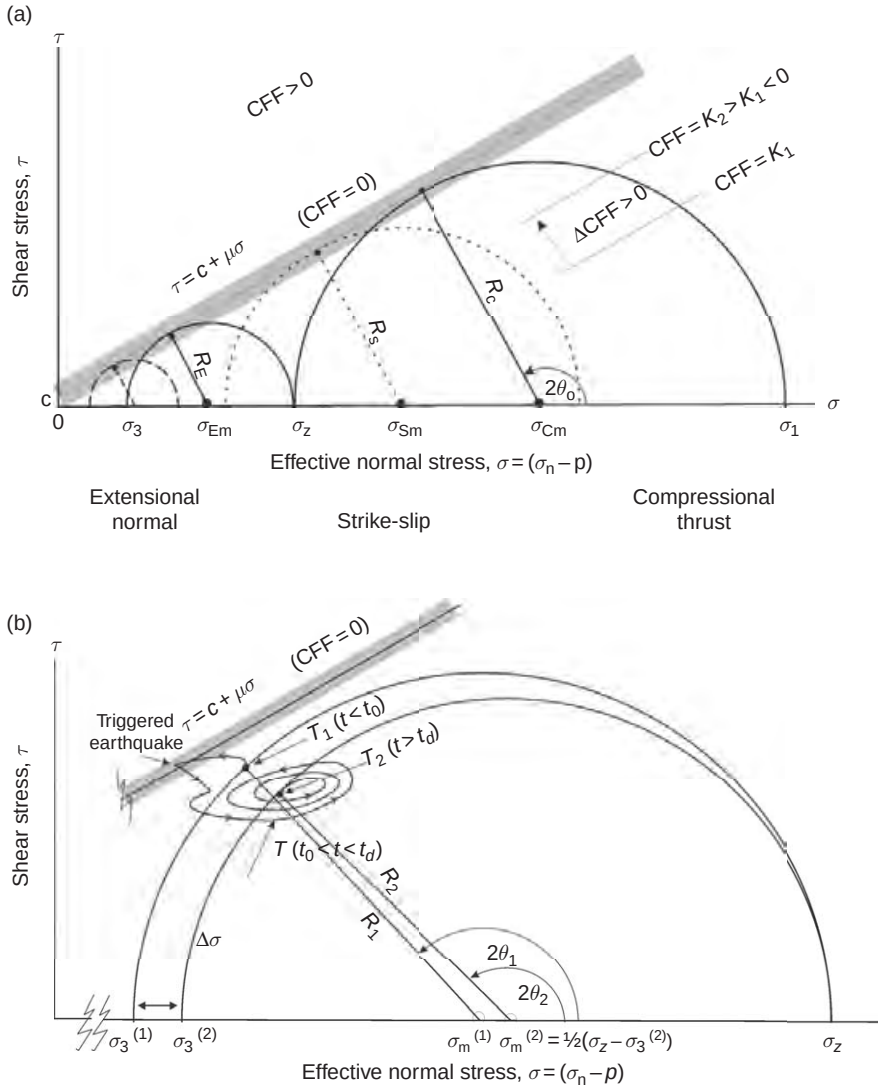


Figure 5 (a) Mohr's diagram comparing maximum stress levels for extensional (normal faulting) and compressional (thrust faulting) regimes in a pervasively faulted crust with a common (e.g., hydrostatic) pore pressure and with frictional strength limited by the Coulomb failure criteria (CFF = 0). τ and σ are the shear stress and effective normal stress components, respectively, acting on a fault plane at an angle, θ, with respect to the least principal stress, σ_3. The coefficient of static friction for rocks is $\mu \sim 0.6$ and C is the cohesive strength with the gray band representing scatter in data. The effective normal stress $\sigma = (\sigma_n - P)$, or the rock matrix normal stress σ_n reduced by the pore pressure, P. The vertical (lithostatic) stress, σ_z, is the greatest principal stress, σ_1, and least principal stress, σ_3, for the extensional (normal faulting) and compressional (thrust faulting) regimes, respectively. The mean stress is $\sigma_m = (1/2)(\sigma_1 + \sigma_3)$ and the radius, R, is half the stress difference, or $R = (1/2)(\sigma_1 - \sigma_3)$. For a crust pervasively fractured with faults of all orientations, $C \sim 0$, and failure will occur when the Mohr's circles for optimally oriented faults ($\theta = \theta_o$) first touch the Coulomb failure line, CFF = 0. Intermediate stress regimes (dotted circle) ranging from transtensional through pure strike-slip to transpressional are represented by Mohr's circles with an intermediate mean stress, σ_{Sm}, with $\sigma_{Em} < \sigma_{Sm} < \sigma_{Cm}$. The small dashed circle illustrates the case for super-hydrostatic pore pressure in an extensional regime. Hydrofracturing will result along vertically oriented cracks in the extensional regime if the pore pressure exceeds the least principal stress such that $\sigma_3 \leq 0$. (b) Schematic illustration of dynamic triggering in the Coulomb failure model in an extensional regime. The point $T_1 = T(R_1, \theta_1)$ represents a stable stress state slightly below the failure threshold on a fault at an angle θ_1 with respect to the local least principal stress direction, σ_3. At $t = t_0$ dynamic stresses from a distant earthquake arrive producing an oscillatory perturbation $T(R(t), \theta(t))$ in both the orientation and magnitude of the local stress field. Slip occurs whenever $T(R(t), \theta(t))$ enters the Coulomb failure zone, CFF $\sim$ 0 (gray band) confining $T(R(t), \theta(t))$ excursions to within or below this zone (CFF $\geq$ 0). The dynamic stresses die away at time $t = t_d > t_0$, and the stress state on the fault settles to $T_2 = T(R_2, \theta_2)$ for $t > t_d$ reflecting the stress drop associated with the triggered seismicity. The greatest principal stress, σ_z, remains pinned to the lithostatic load in an extensional regime.

$\Delta\tau$. (This corresponds to a stress state defined by a line parallel with but incrementally below the Coulomb failure curve, CFF = 0 in **Figure 5(a)**.) A weak crust would be more susceptible to dynamic triggering, however, if average stress levels were some large fraction of the failure stress (represented by a line with a smaller slope than that for CFF = 0 in **Figure 5(a)**). Currently available data seem ambiguous on this point. Any diminished propensity for triggering in stable intraplate regions due to a stronger crust may be at least partially offset, however, by the fact that seismic-wave amplitudes generally decay more rapidly with distance in tectonically active regimes than in stable intraplate regions because of higher intrinsic attenuation and the presence of velocity inversions in the lower crust and upper mantle that partially channel both body-wave and normal-mode (surface wave) energy within depth intervals below the brittle crust (Bakun and McGarr, 2002).

4.09.3.1.2 Subcrustal triggering

The few documented instances of subcrustal dynamic triggering are tied to seismogenic subducting slabs beneath Japan and the Tonga-Fiji arc. Remotely triggered increases in low-frequency (DLP), tremor-like seismicity at depths of 30–40 km beneath western Honshu and Shikoku appear to be related to dehydration of the Philippine Sea plate as it slides beneath Japan along the Nankai subduction zone (Miyazawa and Mori, 2005). Magnitude $M \sim 7$ earthquakes at depths of 500–600 km within the seismogenic subducting slab beneath the Tonga-Fiji arc have apparently triggered large ($M \geq 5.9$) earthquakes at comparable depths but at offset distances of ~300 km (Tibi *et al.*, 2003). Green (2003) speculates that dynamic stresses likely triggered a phase change in metastable olivine within the slab resulting in relatively weak inclusions that deform in the ambient stress field leading to adiabatic shear heating and an earthquake.

4.09.3.2 Mainshock Source Mechanisms and Directivity

Both the most common and most compelling displays of remote dynamic triggering are associated with $M > 6.5$ earthquakes on strike slip faults with unilateral rupture and pronounced directivity as exemplified by the Landers and Denali Fault earthquakes (and to a lesser extent by the Hector Mine earthquake). Large subduction-zone earthquakes, however, also produce remote dynamic triggering as, for example, in the case of the Valley of Mexico (Singh *et al.*, 1998), the subcrustal seismicity beneath Japan and the Tonga-Fiji arc (Miyazawa and Mori, 2005; Tibi *et al.*, 2003), and most recently, Iwo Jima (Ukawa *et al.*, 2002) and Mt. Wrangell, Alaska (West *et al.*, 2005) from the great ($M_W = 9.1$) Sumatra-Andaman earthquake (Lay *et al.*, 2005). In the latter, the sparse seismic coverage in the area of peak directivity may account for the fact that only two reports of remote triggering have yet to emerge following the largest earthquake in 40 years. The lack of published descriptions of remote dynamic triggering attributable to large normal fault earthquakes may be largely a matter of sampling. Few $M > 6.5$ normal earthquakes have occurred in well-instrumented continental environments in the last 25 years.

4.09.3.3 Triggered Onsets and Delay Times

In most cases the onset of dynamically triggered seismicity at a given site begins not with the first arriving P-wave from a distant mainshock but during or some time, Δt, after the arrival of the large-amplitude Love or Rayleigh waves. This observation suggests that (1) the reference time, t_0, for dynamic triggering at remote distances should generally be taken to coincide with the local Love-wave arrival time and (2) that dynamic stresses with periods much below 20–30 s are not efficient at inducing widespread remote triggering. Coincidence of a small, isolated earthquake in Long Valley caldera with the S-wave arrival from the Denali Fault earthquake (Prejean *et al.*, 2004) may represent a rare exception to surface-wave triggering at remote distances. Clearly, this does not apply to dynamic triggering in the aftershock zone at near to intermediate distances where long-period surface waves have yet to emerge as a separate phase. Hough's (2005) inference from stacked β-statistic maps that SmS waves with frequencies in the 1 Hz range may induce dynamic triggering at intermediate distances, for example, stands as a possible example of short-period body-wave triggering.

Reported delay times, Δt, between arrival of the dynamic waves, t_0, and the apparent onset of locally triggered seismicity vary from seconds ($\Delta t \sim 0$) to weeks or more. Delay times for the well-recorded instances of dynamic triggering from the Landers, Hector Mine, and Denali Fault earthquakes ranged from a few seconds to between 24 and 33 h. Delay times reported for suspected dynamic triggering in

the Taiwan region approach 15 days (Beresnev and Wen, 1995; Wen *et al.*, 1996) and, in the Valley of Mexico, ~30 days (Singh *et al.*, 1998). The longer the delay time, of course, the more tenuous the case for a causal link between the dynamic stresses and a local seismicity rate increase. Statistically, instances of suspected triggering may be considered 'instantaneous' if the first event in the sequence has a delay time $\Delta t < 3/\lambda$, where λ is the mean rate of the first n events in the sequence (see Appendix 1). In this sense, triggered seismicity mimics an aftershock sequence to dynamic stressing (Brodsky, 2006). Establishing the significance of longer delay times ($\Delta t > 3/\lambda$) ultimately will depend on the credibility of a well-tested physical model that accounts for the delay between arrival of dynamic stresses and the local onset of brittle failure (Section 4.09.4.2).

It is important to bear in mind that reported delay times based on recording from standard, short-period seismic networks may overestimate actual delay times. In particular, evidence for short delay times is not likely to appear in standard earthquake catalogs because local earthquakes are typically masked by the large-amplitude surface waves from large earthquakes. The development of broadband, high-dynamic-range instrumentation and digital processing has greatly enhanced our ability to recognize the early onset of local earthquake activity within coda waves from large, distant earthquakes through effective filtering. Prejean *et al.* (2004), for example, found this to be the case in their search for dynamic triggering along the west coast of the United States following the $M = 7.9$ Denali Fault earthquake of 2002 (**Figure 6**). Applying a β-statistic test to the Pacific Northwest and California earthquake catalogs produced no clear evidence for dynamic triggering. By applying a high-pass filter to seismograms from broadband and strong motion instruments, however, they discovered a plethora of small earthquakes embedded in the strong surface waves from the Denali earthquake at four sites in Washington and California from Mount Ranier ($\Delta \sim 3108$ km) south to the Coso volcanic field in southern California ($\Delta \sim 3660$ km). (**Figure 6** illustrates the value of spectrograms in identifying small, local earthquakes in the surface wave coda from large earthquakes.) Small delay times were the rule for most instances of dynamic triggering from the Denali Fault earthquake.

This underscores two points: (1) dynamic triggering with small delay times ($\Delta\tau \sim 0$) with respect to peak dynamic stresses is probably considerably more common than generally realized, and (2) our ability to detect dynamic triggering remains extremely uneven depending critically on the distribution of dense, high-quality seismic networks that include broadband, high-dynamic-range digital instruments.

4.09.3.4 Repeat Triggering and Recharge Times

Several sites with documented remote triggering have responded more than once to the dynamic waves from distant earthquakes (see **Table 2**). The most notable example is the Geysers geothermal field in northern California, which has responded to at least 11 different earthquakes. Other repeating sites include the Coso geothermal field in southeastern California; the Katmai volcanic field, Alaska; and in Japan, Iwo Jima, Aso volcano, and the Nankai trough subduction zone. Each responded to at least four different earthquakes. In each of these cases, the response was much the same from one earthquake to the next with the triggered seismicity beginning during the surface wave train and persisting for at most an hour or so. With the exception of the deep, long-period earthquakes along the Nankai trough (Miyazawa and Mori, 2006), the triggered seismicity appeared as a rapid-fire series of overlapping brittle-failure earthquakes (spasmodic bursts) and involved much of the seismogenic volume of the respective systems.

A number of other sites listed with multiple responses in **Table 2** do not strictly involve repetitive sequences. Long Valley caldera, for example, is commonly referenced as a site that has responded to at least three distant earthquakes. Each instance, however, has involved spatially distinct crustal volumes within this large, distributed volcanic field. The Landers response was the most energetic involving the entire seismogenic section of south moat of the caldera (Hill *et al.*, 1995) while response to the Hector Mine and Denali Fault earthquakes involved limited, nonoverlapping volumes beneath the north and south side of Mammoth Mountain, respectively (Johnston *et al.*, 2004a). The delayed response ($\Delta t \sim 24$ h) to the Denali Fault earthquake in the south moat of Long Valley caldera was limited to a relatively shallow volume that had not been previously active (Prejean *et al.*, 2004). Even when the same crustal volume appears to show repetitive triggered responses, successive responses may involve different fault sets with differing orientations.

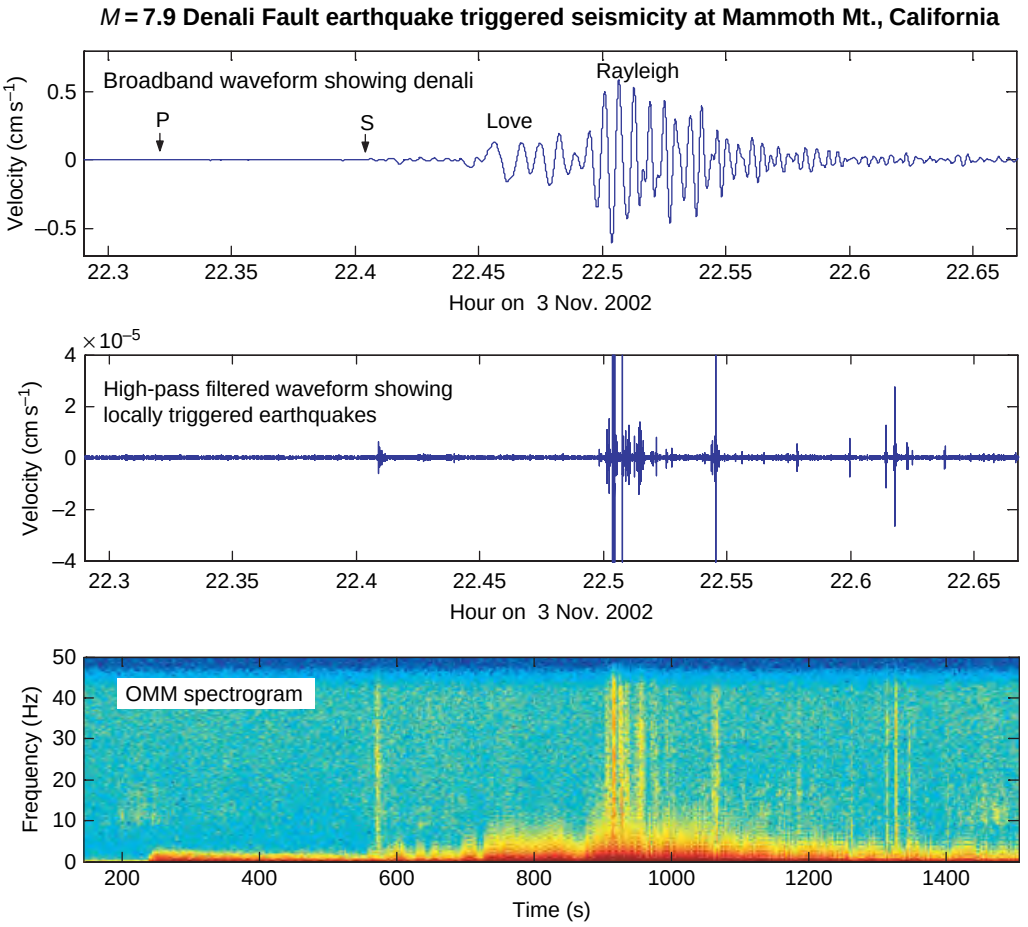


Figure 6 Examples of triggered seismicity detected within the surface wave coda from the Denali Fault earthquake at Mammoth Mountain. The top trace is a broadband seismogram, and the middle trace is a high-pass filtered version of the broadband record revealing the locally triggered activity as a spasmodic burst consisting of rapid-fire earthquakes. The bottom panel is a spectrogram of the broadband seismogram from station OMM located 4 km southeast of Mammoth Mountain. Note the utility of the spectrogram display in revealing local earthquakes within the low-frequency coda as narrow vertical bands rich in high-frequency energy. Modified from Prejean SG., Hill DP, Brodsky EE, *et al.* (2004) Remotely triggered seismicity on the United States West Coast following the M_w 7.9 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S348–S359, courtesy of the Seismological Society of America.

Identification of strictly repeating sites raises the question of recharge times. Each instance of remotely triggered seismicity releases locally stored energy, thus moving the responding site incrementally away from the near-critical state that existed just prior to arrival of the dynamic stresses. The time required for the site to return to a near-critical state (the recharge time) will depend on a host of factors including (1) the energy released during the most recent episode of seismic activity whether remotely triggered or not, and (2) the background rate of active processes feeding energy into the local crustal volume (Herrington and Brodsky, 2006; Miyazawa *et al.*, 2005). As an example of (1), Rabaul caldera (Papua New Guinea) responded to a $M = 7$ earthquake at a distance of 180 km with a pronounced earthquake swarm but produced no detectable activity in response to a second $M = 7$ earthquake 2 months later at a distance of only 60 km (Mori *et al.*, 1989). Recharge times for the Geysers geothermal area appear to be short – a matter of months or less (Gomberg and Davis, 1996). They are apparently much longer (but ill-constrained) for Long Valley caldera where none of the triggered sites are strictly repetitive. Recharge times and their spatial–temporal fluctuations complicate efforts to establish triggering thresholds for dynamic stresses (e.g., Gomberg, 1996; Gomberg and Johnson, 2005). At the same time, they

Table 2 Reported instances of remote dynamic triggering

Site	Responses		Regime	Triggering mainshocks		References
	Number	Mmax		M min–max	Distance (km) min–max	
Mt. Wrangell, AK	1	$M<1$	V	9.1	~11 000	West *et al.*, 2005
Katmai, AK	4	$M=2.3$	G, V	7.9	115–740	Moran *et al.*, 2004
South B.C., Canada	1	N/A	C	7.9	1800–2200	Gomberg *et al.*, 2004
Mt. Rainer, WA	1	$M<1$	V	7.9	3108	Prejean *et al.*, 2004
Geysers, CA	~11?	$M<3$	E, G, V	6.5–7.9	202–3120	Gomberg, 1996, Prejean *et al.*, 2004
Coso, CA	>4	$M=3.2$	E, G, V	to 7.9	165–660	Prejean *et al.*, 2004
Long Valley, CA	2	$M=3.4$	E, G, V	7.4–7.9	414–3454	Gomberg *et al.*, 2001, Prejean *et al.*, 2004
Mammoth Mtn, CA	2	$M<2$	E, G, V	7.2–7.9	420–3454	Prejean *et al.*, 2004; Johnston *et al.*, 2004
Lassen Peak, CA	1	$M=2.8$	E, V	7.4	840	Hill *et al.*, 1995
Burney, CA	1	$M=2.8$	E	7.4	900	Hill *et al.*, 1995
Salton Sea area, CA	1	$M=4.7$	E, V, G	7.1	120–150	Hough and Kanamori, 2002
Central and South CA	>5	$M=5$?	E & C	5.8–6.1	70–120	Hough, 2005
Offshore S. CA	1	$M=2.5$	E	7.9	4003	Prejean *et al.*, 2004
Western Nevada	1	$M\sim4$	E, G	7.4	450–650	Anderson *et al.*, 1994
Little Skull Mtn, NV	1	$M=5.6$	E	7.4	240	Anderson *et al.*, 1994
Yellowstone, WY	2	$M=3.0$	E, G, V	7.4–7.9	1250–3100	Husen *et al.*, 2004b
Wasatch front, UT	2	$M=3.2$	E, G	7.4–7.9	3000–3500	Pankow *et al.*, 2004
Cascade, ID	2	$M=1.7$	E, G	7.4	1100	Husker and Brodsky, 2004
Eastern US (1811-12)	1	$M\sim5$?	C	M>7	~1000	Hough, 2005
Cerro Prieto, Mexico	1	$M=4.1$	E, V, G	7.1	260	Glowacka *et al.*, 2002
Valley of Mexico	~7	$M\sim4$	E, G, V	7.6–8.0	303–588	Singh *et al.*, 1998
Aso, Japan	5	$M\sim2$	E, V	7.1–7.7	900–2213	Miyazawa *et al.*, 2006
Iwo Jima, Japan	4	$M<2$	IA, G, V	7.1–8.0	1228–2002	Ukawa *et al.*, 2002
SISZ, Iceland	1	$M\sim5$	E, G	6.5	80–100	Arnadottir *et al.*, 2004
Roer Valley, Holland	1	$M=3.7$	E	5.4	40	Camelbeeck *et al.*, 1994
Greece	1	$M<3.5$	E	7.4	400–1000	Brodsky *et al.*, 2000
Syria-Lebanon border	1	$M=3.7$	C	7.3	500	Mohamad *et al.*, 2000
Tiawan region	9	$M>4$	?	6.5–7.1	138–2959	Wan *et al.*, 1996
Nanki Trough, Japan	2	N/A	S	7.3–8.1	900–4000	Miyazawa and Mori, 2005
Tonga trench	2	$M=5.9$–7.7	S	7.1–7.6	260–290	Tibi *et al.*, 2002

Tectonic regimes: E, extensional, transtensional; C, convergent, transpressional; G, geothermal; V, volcanic; IA, island arc; S, Subcrustal subduction zone; SISZ, South Iceland Seismic Zone.

offer important clues to the processes behind dynamic triggering at any given site.

4.09.3.5 Peak Dynamic Stresses, Triggered Magnitudes, and Durations

Seismic-wave amplitudes responsible for dynamic triggering are variously reported as peak dynamic stress, T_p, or peak dynamic strain, ε_p. In the plane-wave approximation, peak dynamic strain is proportional to peak particle velocity, $\mathring{u}$ divided by the phase velocity, or $\varepsilon_p \sim \mathring{u}\ v_s^{-1}$ for shear waves, and $T_p \sim G(\mathring{u}\ v_s^{-1})$ where G is the shear modulus (commonly taken as $G \sim 3 \times 10^4$ MPa) and v_s is the shear-wave velocity (e.g., see note 20 in Hill *et al.*, 1993). Although peak dynamic stresses at depth will in general differ from those based on seismograms recorded on the Earth's surface, they can be estimated given a reasonable model for the physical properties of the underlying crust (Gomberg, 1996). In computations of stress amplitudes at depth, for example, the tendency for surface-wave displacement amplitudes and strains to decrease with depth will be offset to one degree or another by the tendency of elastic moduli to increase with depth.

Reported peak dynamic stresses associated with remotely triggered seismicity range from 0.01 MPa ($\varepsilon_p \sim 0.3$ microstrain) at the Coso volcanic field 3660 km from the $M_W = 7.9$ Denali Fault earthquake (Prejean *et al.*, 2004) to ~1 MPa ($\varepsilon_p \sim 3$ microstrain) or more for the Little Skull Mountain earthquake 240 km north of the $M = 7.4$ Landers earthquake (Hill *et al.*, 1993; Anderson *et al.*, 1994). Peak dynamic stresses can easily exceed 4 MPa within the transitional region to the aftershock zone and near field of a large earthquake (Kilb *et al.*, 2002). The large range in peak dynamic stresses (or strains) that have resulted in remote triggering together with variations in intrinsic site characteristics and recharge times indicate that the triggering process does not depend on a simple minimum amplitude threshold for dynamic stresses to be effective. Thus, although most instances of dynamic triggering involve dynamic strains, $\varepsilon_p \geq 1 \times 10^{-6}$, or dynamic stresses, $T_p \geq 0.03$ MPa (Gomberg and Johnson, 2005), this is neither a necessary nor sufficient threshold for dynamic triggering (most areas with $T_p \geq 0.03$ MPa are not triggered and some areas with $T_p < 0.01$ MPa are triggered). The weight of evidence at this point suggests that for a given peak amplitude, dynamic stresses in the periods range 20–30 s are more effective at inducing a triggered response than those at higher frequencies and that a lower bound on the peak dynamic stress capable of inducing a triggered response may be at the level of tidal stresses, or ~0.001 MPa at periods of 12–24 h (Beeler and Lockner, 2003; also see Section 4.09.2.3.4).

Durations of dynamically triggered seismicity sequences range from a few minutes to several weeks. Short-lived episodes of triggered seismicity often appear as a rapid-fire sequence of overlapping, brittle-failure earthquakes (spasmodic bursts) within the surface wave train. Longer-lived episodes more commonly evolve as earthquake swarms or foreshock–aftershock sequences with a temporal decay well described by a modified Omori's law of the form

$$n(t) = k/(t + c)^p \qquad [3]$$

where $n(t)$ is the number of earthquakes per unit time t, and k, c, and p are parameters (Kisslinger and Jones, 1991). Pankow *et al.* (2004), for example, use this relation to estimate durations of ~25 days for the seismicity triggered in central Utah by the 2002 Denali Fault earthquake. Hainzl and Ogata (2005) describe a promising application of the epidemic-type aftershock sequence (ETAS) model that can be used to recognize the modulating role of pore-pressure diffusion in swarm-like triggered sequences. Generally speaking, longer durations tend to be associated with more energetic episodes of triggered seismicity. Indeed, Brodsky (2006) argues that the duration of most triggered sequences can be, in principle, predicted from the cumulative seismic moment of events triggered during the passage of dynamic stressing. The validity of this conjecture will be tested with recordings of future instances of remote triggering on broadband, high-dynamic-range instrumentation such that triggered events within the coda of large-amplitude dynamic waves can be clearly resolved.

4.09.3.6 Dynamically Triggered Deformation

Only three sites that have responded to dynamic stresses with remotely triggered seismicity are equipped with continuous, high-resolution deformation monitoring instrumentation: Long Valley caldera, California; the Reykjanes Peninsula, Iceland; and the Cerro Prieto geothermal field, Mexico. At each of these sites, the triggered seismicity was accompanied by a deformation transient – an order of magnitude larger than that is attributed to the summed slip of the triggered earthquakes.

Deformation transients in Long Valley caldera triggered by the 1992 Landers earthquake were recorded by a borehole strainmeter (dilatometer) and a long-base tiltmeter. The deformation transients beneath Mammoth Mountain triggered by the 1999 Hector Mine and the 2003 Denali Fault earthquakes were both recorded by three borehole strainmeters. In each case, the deformation transient began during the surface wave train. As illustrated in **Figure 7**, the Landers deformation transient grew to a peak amplitude of ~0.3 microstrain over a period of 5 days in parallel with the cumulative seismicity (Hill *et al.*, 1995; Johnston *et al.*, 1995) while the Hector Mine and Denali Fault transients more closely resembled strain steps reaching peak amplitudes of ~0.1 microstrain over a period of roughly 10 min (Johnston *et al.*, 2004b).

The deformation transient across the Reykjanes Peninsula triggered by the 17 June 2000 $M = 6.5$ earthquake was captured by a pair of continuous GPS stations located on either side of the Peninsula. Displacements of campaign-mode GPS stations occupied in July 2000 with respect to a 1998 survey indicate that most of the coseismic deformation can be attributed to aseismic slip (a slow earthquake with a geodetic moment of $\sim 7 \times 10^{17}$ N m) on a fault 78 km (approximately four source dimensions) west of the $M = 6.5$ mainshock (Arnadottir *et al.*, 2004).

Ground deformation in the Mexicali Valley area, Mexico, a pull-apart basin formed by the right-stepping offset between the dextral Cerro Prieto fault zone and the southern end of the dextral Imperial fault, is monitored by both crack meters (creep meters) and tiltmeters. The crack meters recorded subsidence of ~10 cm along the southern end of the Imperial fault accompanied by the onset of local microearthquake ($M < 2.5$) activity that developed as the surface waves from the 1999 Hector Mine earthquake 260 km to the north passed through the area. Some 30 h later, tiltmeters recorded the onset of pronounced deformation within the Cerro Prieto geothermal area within the Mexicali Valley followed by the onset in a surge in $M > 2.5$ seismicity beneath the geothermal field that gradually slowed following an $M = 4.1$ earthquake beneath the geothermal field a day later (Glowacka *et al.*, 2002).

4.09.4 Proposed Models

The convincing display of dynamic triggering by the 1992 Landers earthquake spawned a host of physical models to explain how relatively low-amplitude oscillatory seismic waves can trigger earthquakes at distances of hundreds to thousands of kilometers. Here we describe the more widely cited of these models and the conditions and locations under which each is viable.

4.09.4.1 Triggering by Frictional Failure

This class of models involves direct triggering with the dynamic stresses providing the stress increment necessary to exceed the frictional strength of faults, thus leading to unstable slip and local earthquakes. Frictional models are commonly discussed within the context of a steady loading rate (far-field plate motion or the steady extension of a spring in the case of a slider block model) and a 'clock change'. A 'clock advance', for example, results if a small dynamic perturbation in the applied stress triggers a slip event that would not have otherwise occurred until the failure threshold was reached under the steady, far-field loading rate (e.g., Gomberg *et al.*, 2005, 1997; Perfettini *et al.*, 2003). One consequence of a clock advance in this context is that the time to the next slip event under far-field loading is inversely proportional to the far-field loading rate. Frictional models involve fluids only indirectly through the effect of pore pressure on the effective normal stress (or equivalently, an effective coefficient of friction) on fault planes. For dynamic stresses with periods of minutes or less, the rock matrix will behave as an undrained poroelastic medium with fluid transport having a negligible influence on pore-pressure fluctuations (Cocco and Rice, 2002). Frictional models for dynamic triggering will be limited to areas where the stress state on pre-existing crustal faults hovers below the Coulomb failure stress by less than the peak dynamic stress amplitude.

4.09.4.1.1 Coulomb failure under dynamic stresses

Under the Coulomb model for friction, brittle failure occurs when the combination of shear (τ) and normal stress (σ) components on a fault plane exceed the cohesive strength, C, and the static coefficient of friction, μ_s, such that the Coulomb failure function (Eqn [2]) $\mathrm{CFF} \geq 0$. In a Mohr's diagram (**Figure 5(a)**), $\mathrm{CFF} = K_i$, a constant, forms a family of straight lines of with slope, μ with $\mathrm{CFF} = 0$ representing Byerlees' law for frictional failure. $\Delta\mathrm{CFF} = \Delta\tau - \mu\Delta\sigma$ (eqn [1]) is thus the interval between two lines, K_{i+1} and K_i. In the stable domain ($\mathrm{CFF} < 0$), a

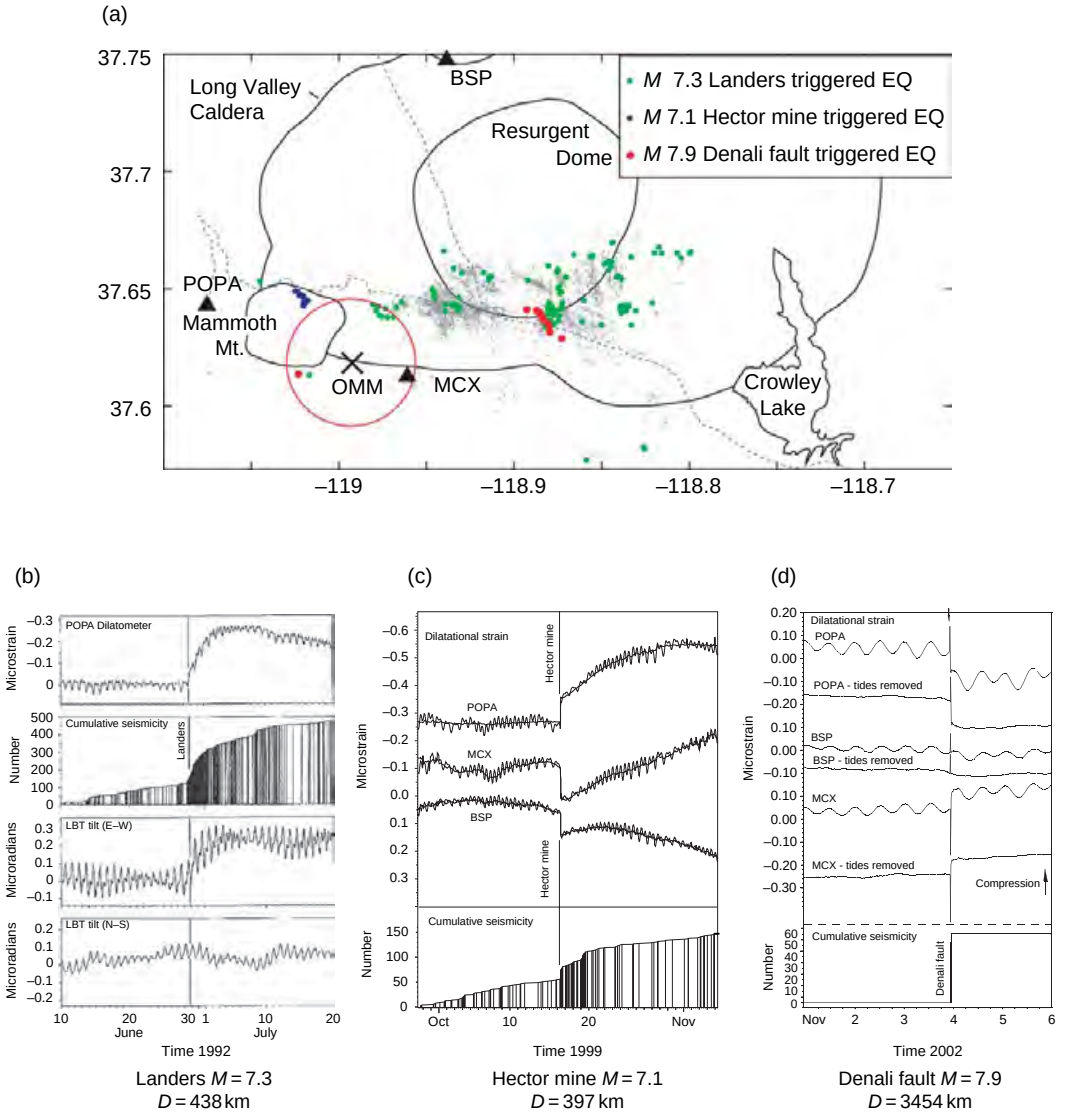


Figure 7 (a) Map of triggered seismicity beneath Long Valley caldera and Mammoth Mountain for the Landers (green), Hector Mine (blue), and Denali Fault (red) earthquakes. Gray dots show seismicity for the period 1997–1998. The red circle centered on station OMM indicates area within which the Denali Fault-triggered earthquakes must be located based on S-P times. The single red dot is the epicenter for the only of these earthquakes large enough for a multistation location (Prejean *et al.*, 2004). (b) Deformation transient triggered by the Landers earthquake as recorded on the POPA borehole dilatometer (top) and the long base tiltmeter, LBT. The bottom two panels show E–W and N–S tilt components, respectively. Cumulative number of earthquakes are plotted in the second panel. (c) Deformation transient triggered by the Hector Mine earthquake as recorded on borehole dilatometers POPA, MCX, and BSP (Johnston *et al.*, 2000). Cumulative number of earthquakes plotted in bottom panel. (d) Deformation transient triggered by the Denali Fault earthquake as recorded on borehole dilatometers POPA, MCX, and BSP. Strain records for each dilatometer are shown with solid Earth tides (top) and with tides filtered out (bottom). Cumulative number of earthquakes plotted in bottom panel. (a) Modified from Prejean SG., Hill DP, Brodsky EE, *et al.* (2004) Remotely triggered seismicity on the United States West Coast following the M_w 7.9 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S348–S359, courtesy of the Seismological Society of America. (b) Adapted from Johnston MJS, Hill DP, Linde AT, Langbein J, and Bilham R (1995) Transient deformation during triggered seismicity from the 28 June 1992 $M_w = 7.3$ Landers earthquake at Long Valley volcanic caldera, California. *Bulletin of the Seismological Society of America* 85: 787–795, courtesy of the Seismological Society of America. (d) Adapted from Johnston MJS, Prejean SG, and Hill DP (2004a) Triggered deformation and seismic activity under Mammoth Mountain in Long Valley Caldera by the 3 November 2002 M_w 7.9 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S360–S369, courtesy of the Seismological Society of America.

positive ΔCFF indicates that the stress state has moved incrementally toward failure. Under the Coulomb failure model for dynamic triggering (**Figure 5(b)**), the traction $T[R(t), \theta(t)]$ on a fault at angle θ with respect to the initial stress state, T_1, begins an oscillatory trajectory as the seismic waves arrive at time t_0. A triggered earthquake results if the trajectory enters the Coulomb failure zone CFF ~ 0 (gray band in **Figure 5**). The stress drop associated with the triggered earthquake alters the stress state (its orientation, θ, and the stress difference, $R_2 < R_1$) in the vicinity of the fault such that the traction on the fault settles to a new value, T_2, after the seismic waves have passed ($t > t_d$). Multiple earthquakes may occur during a single dynamic stress cycle if the initial triggered earthquake produces its own aftershock sequence or if the stress cycle triggers earthquakes on multiple nearby faults. A particularly compelling example of this behavior is illustrated in **Figure 8** where bursts of small earthquakes at Mt. Wrangell, Alaska, occur in phase with peaks in the horizontal extensional stresses during the

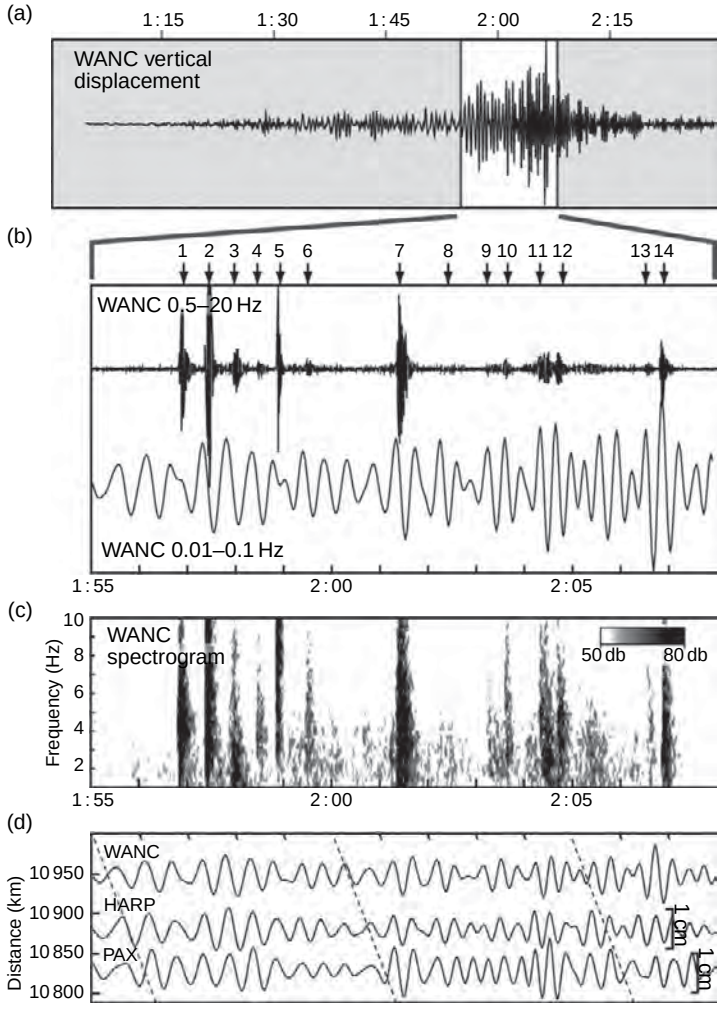


Figure 8 Seismic records showing the triggered response of Mt. Wrangell, Alaska, to the dynamic waves from the 26 December 2004 $M_W = 9.1$ Sumatra–Andaman Islands earthquake from (West *et al.*, 2005). (a) Short-period vertical displacement seismogram of the Sumatra–Andaman Islands wave train at the summit station WANC at a distance of ~11 000 km from the epicenter. Time at top in hours, universal time. (b) Expanded view of highlighted section in (a) showing signal filtered with a band pass 0.5–20 Hz (top) and 0.01–0.1 Hz (bottom). (c) Spectrogram for the same time interval in (b). (d) comparison of WANC displacement seismograms with those from nearby stations. Dashed lines indicate a reference horizontal phase velocity of 3.1 km s^{-1}. Reproduced from *Nature*.

Rayleigh wave train from the $M_W = 9.1$ Sumatra–Andaman Islands earthquake of 26 December 2004 (West *et al.*, 2005).

In principle, this model should be viable in any region where faults are critically stressed and favorably oriented with respect to the polarity of the dynamic stresses from a distant earthquake. If most of the seismogenic crust is pervasively fractured and critically stressed as suggested by Zoback and Zoback (2002), earthquakes triggered by this process have an equal likelihood of occurring nearly everywhere. If this were the dominant triggering mechanism, one would expect that, given some minimum dynamic stress level, the distribution of observed triggered seismicity would correlate directly with density of seismic networks and sensitivity of earthquake detection techniques. As Rivera and Kanamori (2002) argue, however, heterogeneity of both the stress field and frictional strength seems to be an essential property of the crust. In a similar vein, Lapusta and Rice (2003) argue for heterogeneous strength along a given fault such that earthquake nucleation is confined to weak spots while the average fault strength remains well above the Coulomb failure threshold.

4.09.4.1.2 Nonlinear friction

Extensive laboratory experiments on rock friction reveal that frictional failure involves a much richer range of (nonlinear) behaviors than predicted by the familiar first-order model involving a constant static μ_s and dynamic μ_d coefficients of friction with $\mu_s > \mu_d$. The Dieterich–Ruina rate-state friction laws (Dieterich, 1979) have been widely used in studies of both static and dynamic triggering. These rate-state laws, in which static friction retains a memory of sliding history along the fault surface and dynamic friction depends on the sliding velocity and an evolving state parameter, predict a range of behaviors depending on temporal variations in the applied stress and the parameter regime for the constituative relation (Dieterich, 1979; Scholz, 1998). In the unstable, velocity-weakening regime, for example, rate-state friction appears to be most effective under conditions of low effective normal stress (high pore pressure and weak faults) together with high-frequency vibrations (high stressing rates) (Gomberg *et al.*, 2003; Perfettini *et al.*, 2003; Scholz, 1998). Brodsky and Prejean (2005), however, conclude that near-lithostatic pore pressures would be required for the rate-state friction model to be consistent with the dynamic triggering observed in Long Valley caldera. To first order, then, rate-state friction offers no significant advantage over the simple Coulomb failure model for the onset of stick-slip (seismicity) under dynamic triggering. Because the frictional state is changed by slip, however, the rate-state model admits the possibility of an Omori-like decay of triggered seismicity after shaking has stopped (Gomberg *et al.*, 2005).

One interesting implication of rate-state friction in the conditionally stable regime is that dynamic stresses can temporarily convert stable sliding (steady fault creep) to a stick-slip mode. As discussed by Gomberg *et al.* (1997), this represents 'new seismicity' in the sense that the earthquakes generated during the stick-slip mode would not have occurred in the absence of dynamic triggering. This model may apply to Stierman's (1977) recordings of acoustic emissions in a borehole within creeping section of the San Andreas fault near Parkfield as the seismic waves from the underground nuclear explosion KASSERI passed through the area on 28 October 1975.

Experiments by Johnson and Jia (2005) on the response of granular media to imposed dynamic stresses offer insight on another possible mode of nonlinear frictional failure. They find that seismic waves with peak strain amplitudes greater than $\sim 10^{-6}$ (or $> \sim 0.03$ MPa) incident on a weak fault with a core composed of fault gouge (the granular medium) can result in an abrupt decrease in the modulus of the fault gouge. Their experiments show that for faults with effective normal stresses less than ~ 1 bar (e.g., near-lithostatic pore pressures) and a near-critical shear stress, the result can be catastrophic slip and an earthquake.

4.09.4.1.3 Subcritical crack growth

Subcritical crack growth or stress corrosion is a process that has long been known in materials science to make cracks unstable (Anderson and Grew, 1977; Atkinson, 1984; Gomberg *et al.*, 2001). When a crack experiences a change in its environment, such as a sudden increase in loading in the presence of particularly high temperatures and fluids, it may grow due to weakening at the crack tip by chemical corrosion. The crack may then rupture catastrophically. Thus, a sudden increase in loading or oscillatory loading (dynamic stressing) may shorten the time to earthquake failure. The equations governing subcritical crack growth are mathematically identical to those governing failure by rate and state friction (Kanamori and Brodsky, 2004), and thus, as Brodsky and Prejean

(2005) point out, subcritical crack growth requires near-lithostatic pore pressures to be effective as a dynamic triggering process.

Because each of the above nonlinear models (rate-state friction, granular media, and subcritical crack growth) apparently requires near-lithostatic pore pressures to be effective in dynamic triggering, it seems unlikely that they offer a universal explanation for dynamic triggering. Any of the three, however, may be important in limited volumes of the crust where pore pressures approache the lithostatic limit.

4.09.4.2 Triggering through Excitation of Crustal Fluids

Crustal fluids play a critical role in a wide range of tectonic and magmatic processes (*see* Chapter 4.10), and there is no reason to suppose that dynamic triggering should be an exception. Large earthquakes have long been recognized as capable of disturbing hydrologic regimes at distances of thousands of kilometers beginning with the $M_w = 9.2$ Alaska earthquake of 1964 (Vohis, 1967; Roeloffs, 1996), and fluids have long been known to play a role in earthquake rupture (e.g., Nur and Booker, 1972). These observations have inspired a class of physical models in which dynamic strains from the distant earthquake trigger local seismicity through fluid transport and pore-pressure changes. This process may modify the Coulomb failure function (Section 4.09.4.1.1) such that the effective normal stress is decreased sufficiently to trigger failure (Cocco and Rice, 2002) or that quasi-static (aseismic) strains associated with local, fluid-driven deformation are sufficient to trigger earthquakes. Fluids are active agents in geothermal and volcanic areas, which appear to be particularly susceptible to dynamic triggering, and a number of explanations for triggered seismicity involving the movement of fluids have been proposed in the literature (e.g., Brodsky *et al.*, 1998; Linde *et al.*, 1994; Moran *et al.*, 2004; Johnston *et al.*, 1995; Hill *et al.*, 1995).

In geothermal and volcanic areas, the simple linear version of 'clock advance' may be compromised because, in addition to steady loading from far-field tectonic stresses, advection of fluids and heat from lower-crustal or upper-mantle source supply local energy into a recently triggered volume of the brittle crust. As is evident from the behavior of the Long Valley caldera volcanic-geothermal system over the past 25 years, these processes may operate episodically over timescales ranging from weeks to years (Hill, 2006).

4.09.4.2.1 Hydrous fluid transport: changes in permeability and pore pressure

In this class of models for earthquake triggering, the dynamic strains from a distant earthquake modify crustal permeability by disrupting clogged fractures and hydraulic fracturing, thus leading to a redistribution of pore pressure. These models should be viable in any area where isolated pockets of high-pressure fluids develop. Brodsky *et al.* (2003), for example, proposed that coseismic water-level changes observed in wells at Grant's Pass in southern Oregon might result from dynamic stresses from $M = 7.2$ and $M = 7.4$ earthquakes at ~300 km, and ~3850 km, respectively, opening permeable fractures clogged by accumulating detritus over time.

The pore-pressure redistribution mechanism may be particularly relevant in active geothermal areas, such as the Geysers and Coso geothermal fields, as fractures are sealed and high-pressure compartments form over relatively short timescales as minerals are precipitated from hot brines. The mechanism may also be important in the crust above a magma body in which hydrous magmatic fluids may reach near-lithostatic pressures in the low-permeability plastic zone between the roof of the magma body and the brittle crust, where hydrothermal fluids circulate under hydrostatic conditions. This is the hydraulic surge model proposed by Fournier (1999) in which dynamic stresses from a large distant earthquake may rupture this impermeable zone, permitting high-pressure fluids to surge into the overlying brittle crust. This mechanism may explain both earthquakes triggered during and after the wave train of the large distant earthquake has passed as pore fluids move by diffusion through the crust.

Hough and Kanamori (2002) looked for evidence of fluid movement in the failure process of earthquakes triggered in the Salton Sea by the 1999 Hector Mine earthquake by examining earthquake source parameters. They found that the triggered earthquakes had essentially normal to slightly low stress drops and spectral contents typical of brittle shear-failure with no resemblance to the long-period earthquakes or tremor-like sequences often seen in volcanic areas (Chouet, 1992; Julian, 1994). As pointed out by Brodsky and Prejean (2005), earthquakes triggered by pore-pressure increases would not necessarily have a signature of active fluid

involvement in the source process, as increased pressure may simply decrease normal stress, leading to standard shear failure.

4.09.4.2.2 Magmatic fluids

Growing evidence indicates that dynamic stresses are capable of perturbing the state of crustal or upper-mantle magma bodies triggering internal pressure changes and, in some cases, eruptions (Hill *et al.*, 2002; Manga and Brodsky, 2005; Linde and Sacks, 1998; Marzocchi, 2002). Under these models, triggered seismicity is a secondary result of locally triggered deformation associated with the change in state of a nearby magma body. Because these models predict distinctive patterns of ground deformation associated with triggered seismicity, continuous high-resolution deformation monitoring networks will be needed to test them.

Bubble excitation. One intriguing class of models proposed to explain earthquake–magmatic interactions centers on the role of bubbles in magma. Bubbles play a central role in the source mechanisms for long-period and very-long-period volcanic earthquakes (Chouet, 1992) as well as in eruption dynamics (Mangan and Sisson, 2000). Two highly idealized physical models widely mentioned as possible explanations for remotely triggered seismicity appeal to changes in bubble pressurization in a two-phase fluid. In the rectified diffusion model, volatiles in a saturated fluid are selectively pumped into bubbles during the dilatational phase of each strain cycle as the wave train of the distant earthquake passes (Sturtevant *et al.*, 1996; Brodsky *et al.*, 1998). Increasing pressure in the bubbles is transmitted to the surrounding interstitial fluid, thereby increasing pore pressure and triggering earthquakes. In the advective overpressure model, bubbles adhering to the rigid walls of a magma body are shaken loose by the dynamic strains from the remote earthquake. As these bubbles rise a distance h through the (incompressible) magma, pressure in the magma chamber increases by $\rho g h$ (where ρ is magma density and g is the acceleration of gravity), thereby deforming the surrounding elastic crust and increasing pore pressure in the surrounding brittle rock (Linde *et al.*, 1994; Sahagian and Proussevitch, 1992). In principle, both models could apply to either hydrous fluids or magma, although both have been criticized on the basis that they depend on restrictive assumptions that are unrealistic under natural conditions in the Earth (Pyle and Pyle, 1995; Brodsky *et al.*, 1998; Bagdassarov, 1994; Ichihara and Brodsky, 2006).

Manga and Brodsky (2006) explore what appears to be a more realistic model for the role of bubbles in dynamic triggering based on the creation of new bubbles (bubble nucleation) in a supersaturated magma. They conclude that, under proper conditions, small pressure changes associated with dynamic stresses in a crystallizing magma that is close to critical supersaturation should be capable of triggering a significant excess in bubble nucleation leading to a marked pressure increase. Numerical modeling by Shimomura *et al.* (2006) and Chouet *et al.* (2006) indicates that a densely packed matrix of tiny bubbles in a magma-filled crack when subject to a small, externally imposed pressure drop can lead to rapid diffusion-driven bubble growth and volumetric expansion of the crack. These models have yet to be evaluated for their response to realistic dynamic stresses, but they keep open the possibility that bubbles may have an important role in the triggered response of magmatic systems to dynamic stresses.

Magmatic intrusions. Dilatational strain meters at Mammoth Mountain recorded deformation transients coincident with the seismicity triggering in the Long Valley caldera area following the Landers, Hector Mine, and Denali Fault earthquakes. Based on these data, Johnston *et al.* (2000) and Johnston *et al.* (2004a) have proposed that in the case of Mammoth Mountain and Long Valley, the dynamic waves from distant earthquakes trigger magmatic intrusions into the shallow crust. These intrusions, which may involve a massive hydrofracture by aqueous magmatic fluids or magma itself, then trigger earthquakes by changing the static stress field or locally increasing the pore pressure as fluids are exsolved. Thus, the local seismicity is a secondary response to the initiating ground motion.

Relaxing magma body. Dynamic waves from a large distant earthquake may disrupt the stability of a partially crystallized magma body releasing accumulated deviatoric stress supported by the interconnected crystal structure (Hill *et al.*, 1995). As in the magmatic intrusion model, the local seismicity triggered by a relaxing magma body is responding to a local, quasi-static strain source rather than the initiating dynamic stresses. The recharge time for this model is presumably rather long depending on both the crystallization rate in cooling magma body and the far-field strain rate. If such a model has any merit, it raises a question regarding the response of Long Valley Caldera to the $M = 7.6$ Owens Valley earthquake of 1872, which ruptured to

within 50 km of the caldera based on this model. Did the static stress change from the Owens Valley earthquake load strain energy into the crystalline matrix of crystallizing magma body that was subsequently released by dynamic stresses from the 1992 Landers earthquake? If not, the validity of the model as an explanation for the Landers deformation transient is in doubt. Indeed, it seems unlikely that a crystalline matrix could deform sufficiently rapidly to accumulate significant elastic strain in the intervening 120 years if shaking from the Owens Valley earthquake disrupted the magma body in 1872.

Sinking crystal plumes. Another possible response of a magma body to dynamic stresses suggests that passing seismic waves may dislodge dense crystals from the ceiling and walls of a magma chamber. As these crystals then sink due to gravity, they may stimulate convection in the magma chamber. Bubbles would presumably nucleate and grow in the ascending volatile-rich magma, thereby increasing pressure in the magma body and deforming the overlying crust (Hill *et al*., 2002). Manga and Brodsky (2005) show that this process may be capable of creating sufficient overpressure to trigger a volcanic eruption on the timescale of days. Thus, it is likely that the process could trigger earthquakes on a similar timescale, although such calculations have not yet been performed.

4.09.5 Challenges for the Future

The study of remotely triggered seismicity gives us a unique window into the earthquake nucleation process, as it is the only naturally occurring example where we know precisely what perturbation to the Earth's stress field triggered an earthquake. Although this is a fertile field of study with great potential for illuminating the earthquake initiation process, many fundamental questions remain regarding the detection, temporal and spatial distribution, and physical processes behind dynamically triggered seismicity.

4.09.5.1 Challenges in Detecting Triggered Seismicity

Although identifying triggered earthquakes can be straightforward in the case of an abrupt increase in seismicity (e.g., **Figures 2**, **6**, and **7**), it is likely that many earthquakes triggered by dynamic waves remain undetected. Swarms of earthquakes are far easier to link to the remote earthquake than isolated earthquakes. Whether the two earthquakes offshore of CA event following the Denali Fault Earthquake (Prejean *et al*., 2004) are a triggered event or not remains a vexing question of statistical significance. Also, as the time separation between the onset of the dynamic wave train and the triggered earthquake increases, it becomes increasingly difficult to show statistically that the two are linked.

When triggered earthquakes occur in swarms, we are faced with the problem of distinguishing between triggered earthquakes and aftershocks to the triggered earthquakes (Ziv, 2006). One promising approach to this problem involves the ETAS model described by Hainzl and Ogata (2005) to distinguish earthquakes driven by external forcing (such as fluid transport) from those resulting from an Omori-type aftershock sequence. High-resolution imaging of spatial-temporal patterns of swarm evolution can also provide evidence for fluid forcing, particularly when the seismicity front propagates as $r/\sqrt{t}$ characteristic of fluid diffusion (Shapiro *et al*., 1997). This approach should be particularly useful in tests for Fournier's (1999) hydraulic surge model, which implies an upward propagating seismicity front driven by fluid diffusion.

Another challenge we face is discriminating dynamically triggered earthquakes from those triggered by static and quasi-static process. To avoid this uncertainty many researchers have limited their data sets to events triggered at distances greater than $\sim$2 fault lengths from the mainshock where static stress changes become too small to trigger earthquakes (Hill *et al*., 1993). However, as Gomberg *et al*. (2003) suggest, we may be severely limiting our data set by overlooking dynamically triggered earthquakes within the aftershock zone even though static stress triggering also plays a role in this realm (King *et al*., 1994). By comparing dynamic stress amplitudes with the aftershock distribution, Kilb *et al*. (2002) have shown that many aftershocks are probably dynamically triggered. More recently, Feltzer and Brodsky (2006) show that the density of aftershocks with distance from a mainshock is directly proportional to the dynamic stress amplitudes from the mainshock at those distances. Pollitz and Johnston (2006) find that slow (quasi-static) earthquakes in central California have many fewer aftershocks than typical brittle-failure earthquakes of comparable moment. These studies indicate that dynamic waves may play a much larger role in triggering early aftershocks in the near field than was previously suspected.

Finally, automated earthquake-detection algorithms are ineffective at identifying small triggered events recorded by a single, low-dynamic-range station. Reliable identification of triggered seismicity buried in the surface wave coda from a large earthquake requires either visual scanning of high-pass filtered signal or a spectrogram display as illustrated in **Figure 6** (e.g., Prejean *et al.*, 2004). This capability is unevenly distributed in existing seismic networks.

4.09.5.2 Challenges in Mapping the Distribution of Triggered Seismicity

Detailed investigations of triggered seismicity following the Denali Fault earthquake showed conclusively that triggering by dynamic seismic waves is probably far more common and widespread than was previously recognized. Prior to the Denali Fault earthquake, triggered seismicity was primarily identified through abrupt seismicity rate increases in earthquake catalogs following a large distant earthquake (e.g., **Figure 2** and **7**). In the case of the Denali Fault earthquake, however, hundreds to thousands of triggered earthquakes did not appear in earthquake catalogs, either because they were masked by the mainshock coda when many instruments are saturated and automatic earthquake-detection algorithms do not work, or because they were too small to be seen at more than one station (Gomberg *et al.*, 2004; Prejean *et al.*, 2004; Pankow *et al.*, 2004; Husker and Brodsky, 2004). The installation of many broadband, high-dynamic-range instruments in the last decade has allowed scientists to identify high-frequency local earthquakes in the relatively low-frequency wave train of the Denali Fault earthquake at many new sites (**Figure 6**). In many places, however, limitations in dynamic range continue to limit our ability to search for triggered events during the mainshock's wave train.

Many lines of evidence suggest that areas with higher background seismicity rates are more likely to experience dynamic triggering than areas with lower background seismicity rates because these areas are more frequently closer to failure (see Section 4.09.3.1.1). Because areas with high seismicity rates tend to be densely instrumented however, we are more likely to catch triggering in those areas. This highlights the question: to what degree is our current picture of the distribution of remotely triggered seismicity real or the result of a bias due to variations in network density? Spatial variations in network density compound the potential lack of conformity in earthquake triggering maps.

Many more cases of remote dynamic triggering have been observed in geothermally and/or volcanically active areas and areas undergoing active crustal extension than in other tectonic regimes (see Section 4.09.3.1.1 and **Table 2**). Whether these observations are an artifact of high background seismicity rates, high network densities, or high levels of scrunity in these areas has not been thoroughly addressed in the literature and remains an open question. Based on case studies, Spudich *et al.* (1995) suggest that remote triggering is less common in compressional and transtensional nongeothermal areas (specifically along the San Andreas fault) than in geothermally active areas, even when the background seismicity rates, network densities, and levels of scrutiny are compatible. A systematic study of the occurrence of triggering over large areas (such as the tectonically active western United Sates or the more stable midcontinent) that includes more recent and complete observations is yet to be done.

4.09.5.3 Challenges in Determining the Triggering Processes

In the last two decades many intriguing physical models have been proposed to explain remotely triggered seismicity. Some of these models are highly quantitative and testable in the right circumstances; others have not left the realm of thought-provoking but untested speculation. For this field to advance, we must find ways to quantify and constrain these models.

To determine the physical processes responsible for a remotely triggered earthquake, we need thorough knowledge of the triggering wave field in space and time, precise hypocentral locations for the triggered seismicity, focal mechanisms, and other source properties as well as the physical environment where triggering occurs (fault orientations, frictional properties of faults, stress field, and hydrologic regime). Earthquakes that are triggered during the mainshock wave train (**Figure 6**) provide a promising opportunity to investigate failure processes because we can extract information on the spatial-temporal properties of the dynamic stresses at the time triggering is initiated.

The idea that remote triggering depends on a simple amplitude threshold or a distance-magnitude threshold has yet to be proved useful (e.g., Gomberg *et al.*, 2001; Moran *et al.*, 2004). Anderson *et al.* (1994) suggested that large, distant earthquakes trigger seismicity more readily than smaller more local

earthquakes. The $M_W = 6.4$ Chalfant Valley earthquake of June 1986, for example, did not trigger seismicity in Long Valley caldera or Mammoth Mountain at a distance of just 20–30 km despite peak amplitudes much higher than more distant events that did lead to triggering. This suggests that the propensity for earthquake triggering may depend on frequency of shaking in addition to amplitude. Brodsky and Prejean (2005) show that in the case of Long Valley and Mammoth Mountain, earthquakes are generally triggered at remote distances more easily by low-frequency surface waves (periods 15–30 s) than high-frequency surface waves. In contrast, Gomberg (1996) argued that triggering at the Geysers, CA, depended on high-frequency waves. This conclusion in their 1996 paper, however, was partially based on lack of evidence at the time for tidal triggering at periods of 12–24 h (see Section 4.09.2.3.4) compared with triggering by surface waves at periods of 15–30 s (Brodsky and Prejean, 2005). As of mid-2006, the Geysers, CA; Katmai, AK; and Long Valley caldera, Mammoth Mountain, CA, are the only sites that have multiple examples of triggering and nontriggering dynamic stresses to test these hypotheses. Of these, only the Long Valley caldera area has broadband instrumentation capable of faithfully recording ground motion spanning a wide dynamic range and over a wide frequency spectrum. As more triggering observations are made and more broadband, high-dynamic-range data become available, it is likely that we will make progress in addressing this question.

West *et al.* (2005) show that in the Sumatra–Andaman Islands earthquake of 2004, earthquakes were triggered at Mt. Wrangell, Alaska only during the large positive vertical displacements of the Rayleigh wave train (**Figure 8**). Thus, they were able to calculate the specific strain at the general location of the earthquakes at the time of triggering and conclude that these observations would be consistent with triggering by simple shear failure on normal faults. With the limited data available, however, focal mechanisms of the triggered earthquakes could not be resolved.

Estimating the character of the dynamic wave field at the hypocenter of a triggered earthquake requires extrapolation in three dimensions from recordings at nearby stations (e.g., Gomberg and Davis, 1996). Many triggered earthquakes are small and lack reliable focal mechanisms, so even if we understand the dynamic ground motion at the site of triggering, we do not know how that strain field affected the particular failure plane as in the case of Wrangell triggering (West *et al.*, 2005).

A number of informative studies using existing data could be conducted to test individual models for the physical process leading to dynamic triggering. Specific parameters for models based on rate and state friction might be tested by analyzing the decay rate of triggered events and searching for changes in rates of background seismicity after episodes of triggering.

Although strain transients have been observed in every episode of triggering when continuous, high-resolution strain data are available, whether these strain changes are an intrinsic part of the triggering process remains unknown. Inferences about magmatic intrusion and relaxing magma bodies are based on strain data (e.g., Linde *et al.*, 1994; Sturtevant *et al.*, 1996; Johnston *et al.*, 2004a). To test these models it is imperative that we have strain data in more locations and higher densities of strain networks, so that we can adequately model the deformation source. Strain data are key in testing models that involve fluid transport. However, it is difficult to tell at this time whether strain signals are reflecting a compact or a distributed source, such as a regional pore-pressure change (Segall *et al.*, 2003). High-frequency water well data will also be helpful in testing models that involve crustal fluid flow and pore-pressure changes (Roeloffs, 1996), as these data can help determine how the crust responds hydrologically to large distant earthquakes.

Finally, in considering these models we must also recognize that there may be no one unifying causative process; rather, a range of triggering mechanisms may be valid in different locations, over different timescales. In the case of Long Valley for example, a swarm of earthquakes $M < 1$ that occurred under Mammoth Mountain by the Denali Fault dynamic stresses were likely triggered by a different process than a swarm of $M <= 3$ earthquakes that began 24 h later, 10 km to the east in the caldera's south moat. Thus, testing individual models against large data sets that cover a variety of physical environments, distances, and timescales may not prove to be successful.

4.09.6 Conclusions

The question is no longer whether seismic waves (dynamic stresses) from large earthquakes are capable of triggering local earthquake activity over

distances ranging from tens to thousands of kilometers (the evidence in support of dynamic triggering as a common phenomenon is compelling); rather, key questions have become: what conditions favor dynamic triggering, and what physical processes do dynamic stresses induce in given crustal volumes that trigger local earthquakes and, at least in some cases, local deformation? To a large degree, emerging answers will come into focus as we obtain more complete sampling of the triggered response to future large earthquakes in terms of both spatial distribution and variations in response characteristics over a wide spectrum of timescales afforded by more widely distributed broadband seismic and continuous deformation instrumentation.

Based on the still woefully uneven sampling available through mid-2006, it appears as though tectonically active extensional regimes (such as the Basin and Range province in the western United States with strain rates of $\sim 10^{-8}\,\mathrm{yr}^{-1}$) are more susceptible to dynamic triggering than stable cratonic regimes (such as the eastern United States with strain rates generally $<10^{-9}\,\mathrm{yr}^{-1}$). Higher tectonic strain rates mean shorter recovery (recharge) times between episodes of strain release (whether dynamically triggered or not). In volcanic and geothermal areas, recovery times may be further shortened by episodic advection of thermal fluids into the crust from the lower crust or upper mantle that serve as local stress sources. Shorter recovery times mean that, at any give time, more areas are likely to be hovering in a near-critical state and thus susceptible to triggering by small dynamic stresses than in areas with low strain rates and extended recovery times. As Hough *et al.* (2003) argue, however, nonelastic deformation in mid-continental low strain-rate environments should maintain stresses at a near-critical state failure for longer periods than in high strain-rate environments such that a dynamic stress change can lead to a relatively large clock advance. We have yet, however, to see clear examples of remote dynamic triggering in the central or eastern sections of the United States recorded on modern seismic networks.

In any case, elevated strain rates alone are not sufficient. The central section of the San Andreas Fault through the California Coast Ranges with its high strain rate ($\sim 5 \times 10^{-7}\,\mathrm{yr}^{-1}$) and frequent seismicity appears not to be particularly susceptible to dynamic triggering (Spudich *et al.*, 1995). One explanation for this apparent difference in triggering susceptibility appeals to the importance of vertical mobility of crustal fluids in the dynamic triggering process. Fluid-filled cracks open in the direction of the least compressive stress and thus tend to have vertical orientations in extensional regimes and horizontal orientations in transpressional regimes such as that along the San Andreas Fault. Whatever the case, currently available data are not adequate to clearly distinguish between the many competing models for the triggering process, which range from various formulations of friction to hydrous fluid transport or the excitation of magma bodies. Models involving the activation of fluids or the transition to a frictional domain of stable sliding admit the possibility that triggered seismicity is a response to some form of local, aseismic ground deformation, which underscores the importance of obtaining continuous deformation data in areas susceptible to dynamic triggering.

Models proposed to explain dynamic triggering fall into two broad categories: (1) frictional models tied to the Coulomb failure function and direct triggering of local seismicity, and (2) models that appeal to the excitation of local crustal fluids or aseismic creep with seismicity developing as an indirect (and possibly delayed) response. For dynamic triggering under the frictional models, the stress state in the crust must differ from the Coulomb failure stress by less than the peak amplitudes of the dynamic stresses (typically <0.1 MPa for remote triggering). These models are generally consistent with the onset of triggered seismicity during the dynamic stresses followed by an Omori-like decay depending on the parameters for the specific friction law (rate-state, granular media, etc.).

Fluid excitation models involve fluid transport or a phase change driven by the triggered release of locally stored gravitational or chemical potential energy. These fluid-based models, whether hydrous or magmatic, together with models based on aseismic creep, involve some degree of crustal deformation through, for example, intrusion, pressure changes in the case of bubble excitation, advection in a magma body, or poroelasticity in the case of fluid diffusion. Triggered seismicity for many of these models can be a secondary response to quasi-static stresses generated by local deformation, and thus they are somewhat less dependent on a near-critical stress state than the friction models. These models admit the possibility of delayed onsets of triggered seismicity and increasing or sustained activity rates for extended periods following passage of the dynamic stresses. Confidence in delay times $\Delta t \gg 3/\lambda$

(Appendix 1), however, will depend on the existence of a compelling physical model. In general, fluid-based models are more likely to be more applicable to extensional stress regimes that host geothermal and volcanic systems than to stable regimes. Of course, nothing precludes a triggering episode from being the result of some combination of frictional and fluid-based models.

Finally, because the three-dimensional dynamic stress fields for seismic waves from large earthquakes or solid Earth tides can be calculated with reasonable accuracy, they serve as potentially powerful tools for investigating the state of stress in the crust and what it means to be in a near-critical state as well as insight on complex rupture processes through near-field dynamic triggering of nearby large earthquakes.

Appendix 1: Statistical Significance

The β-statistic (Matthews and Reasenberg, 1988; Reasenberg and Simpson, 1992) is widely used in the literature as a test for statistical significance in instances of suspected dynamic triggering. In this appendix, we focus on the β-statistic as a means of illustrating some important issues associated with evaluating the statistical significance of dynamic triggering and we note some alternative approaches where appropriate. This appendix is not intended to be an exhaustive survey of statistical tests for dynamic triggering.

Matthews and Reasenberg (1988) developed the β-statistic as a test for the statistical significance of an offset in average background seismicity rates associated with static stress changes due to a nearby earthquake, under the assumption that background seismicity is a Poisson process. The value of β produced by this test represents the number of standard deviations by which the seismicity rate, $r_a = n_a/t_a$, in a given area after a mainshock of interest differs from an estimate of the background seismicity rate, $r_b = n_b/t_b$, prior to the mainshock. Here, n_a and n_b are the number of earthquakes during the time intervals t_a and t_b after and before the mainshock, respectively. Reasenberg and Simpson (1992) express the β-statistic as

$$\beta(n_a, n_b, t_a, t_b) = [n_a - E(n_a)]/[\mathrm{var}(n_a)]^{1/2} \qquad [4]$$

where $E(n_a) = r_b t_a = n_b(t_a/t_b)$ is the expected number of earthquakes during t_a based on a sample of the background seismicity during t_b, and $\mathrm{var}(n_a)$ is the variance. For a Poisson process, $\mathrm{var}(n_a) = r_b t_a$. The β-statistic is commonly evaluated over an array of spatial bins to create a β-statistic map as in **Figure 3**. Results with $|\beta| > 2$ (approximately two standard deviations) are generally considered to be statistically significant (Reasenberg and Simpson, 1992).

As with any statistical test, however, it is important to recognize the assumptions behind the test and to insure that the same assumptions are consistent with the data at hand. It is equally important to have a clear idea of the hypothesis being tested. The β-statistic was developed to test the hypothesis of a change (either increase or decrease) in average seismicity rate at time t_0 against the null hypothesis of no significant rate change, where t_0 marks the transition from t_b to t_a. In the case of dynamic triggering, we are more interested in the significance of a transient seismicity cluster beginning within some time interval $\Delta t < t_a$ following the onset of dynamic stressing at t_0. Accordingly, applying the β-statistic to putative cases of dynamically triggered seismicity requires some compromises.

One such compromise involves the choice made for the duration of post-mainshock interval, t_a. An overly long interval may contaminate the test by including background seismicity fluctuations not related to dynamic triggering, while too short an interval may introduce problems associated with the statistics of small samples. This choice is unimportant to first order when applied to the detection of a step change in seismicity rate as originally intended because the rate is assumed to be constant after the mainshock. When applied to triggered seismicity, however, the choice of t_a directly affects the calculation of $r_a = n_a/t_a$ and thus the value of β.

Strictly speaking, application of the original formulation of the β-statistic to the evaluation of triggered seismicity violates the assumption that seismicity in both intervals, t_a and t_b, reflects a Poisson process. The clustered seismicity typical of a triggered response might better be represented using the epidemic-type aftershock sequence (ETAS) model of Ogata (1993). Using short intervals for t_a in a β-statistic test for the significance of questionable instances of dynamic triggering may be particularly susceptible to unstable results, considering that the Poisson distribution is valid for a large number of events, n, each with a small probability, p. In an effort to circumvent this problem, Pankow *et al.* (2004) proposed a test based on the binomial distribution in their analysis of dynamically triggered

earthquakes in Utah by the 2002 Denali Fault earthquake, and Gomberg *et al.* (2001), Kilb *et al.* (2002), and Hough (2005) use a modified form of the β-statistic incorporating the binomial distribution. In the end, however, results based on a small number of earthquakes should be viewed with caution.

Viable estimates of background seismicity rates, r_b and r_a, for the β-statistic as originally intended depend on the effective removal of aftershock sequences and earthquake swarms for the entire time interval ($t_a + t_b$) by running the earthquake catalog through a de-clustering algorithm such as that developed by Reasenberg (1985) or Hainzl *et al.* (2006). Viable estimates of background rates also require a stable earthquake catalog with a well-established completeness threshold (the minimum magnitude above which all earthquakes in the area are detected by the seismic network and included in the catalog) and reasonable assurance that any change in data processing has not introduced an offset in earthquake parameters such as magnitude and completeness threshold (Wiemer, 2001; Wiemer and Wyss, 2000). De-clustering the post-dynamic stress interval, t_a, in a test for dynamic triggering, however, is likely to compromise the signal of interest (the triggered seismicity cluster).

One of the more useful applications of the β-statistic to dynamic triggering involves developing a map showing the spatial distribution of likely instance of dynamic triggering as in **Figure 3** (Gomberg *et al.*, 2001). Choice of an appropriate bin size over which to evaluate the β-statistic is analogous to the choice of an appropriate post-mainshock time interval, t_a, both of which are typically under-constrained. The bin area should be sufficiently large to include a 'reasonably large' number of earthquakes. As the area increases, however, spatial resolution decreases and the odds of including random fluctuations in background seismicity increase – particularly in regions with elevated background seismicity rates. Optimal bin size is likely to vary from one region to another depending on background seismicity patterns. One approach to this problem is to evaluate the sensitivity of β to bin size in a given region by calculating β over a range of bin sizes. Because the β-statistic is an imperfect test for dynamic triggering, however, prudence dictates that areas highlighted as possible sites of dynamic triggering be examined for independent evidence of a seismicity cluster beginning within some 'reasonable' time interval Δt, following the dynamic stresses.

The delay time, Δt, between arrival of dynamic stresses from a distant mainshock and the onset of triggered seismicity is one of the more important parameters in dynamic triggering in that it provides clues on the triggering process. Although the onset of a suspected triggered response is often marked by a relatively clear increase in seismicity rate, the question of a reasonable upper bound on the time delay, Δt, remains. Clearly, delay times that approach the mean inter-event time for fluctuations in background seismicity rates in a given area beg the question of statistical significance. The β-statistic is of little use on this front. A conservative approach to this question is that applied by Paul Reasenberg to the triggered response to the 1992 Landers earthquake (see note 15 in Hill *et al.*, 1993). He assumed that seismicity at a given site following a large, distant earthquake can be represented by a Poisson process with a rate, λ, and that the onset of the triggering process coincides with the origin time of the distant earthquake. Because inter-event times in a Poisson process are expected to exceed $3/\lambda$ just 5% of the time, the corresponding upper bound on the delay time for statistically significant triggering becomes $\Delta t < 3/\lambda$. In the case of the Landers earthquake, Reasenberg estimated λ from the first 10 earthquakes, $t_1 \ldots t_{10}$ following the mainshock origin time, t_o. He found $\Delta t = (t_1 - t_o) < 3/\lambda$ in 13 out of 14 instances with Δt ranging from 3 min (corresponding to the surface wave propagation time) to 23 h. It is important to note that (1) this approach does not rely on the β-statistic, (2) Reasenberg's choice of the first 10 events to define the rate is rather arbitrary, and (3) the triggered clusters were initially visually identified by their abrupt onset and high rate (**Figure 2**).

For additional statistical tests relevant to dynamic triggering, see for example, Marsan (2003), Ziv (2006), Harrington and Brodsky (2006), and references therein. Carefully applied, any of these tests offer an important aid in judging the validity of an inferred causal link between dynamic stresses from a distant mainshock and local seismicity rate increases. Blindly applied, of course, they can be seriously misleading. Clearly, any of these tests depend critically on the quality, duration, and completeness of local earthquake catalogs as a basis for establishing a reliable estimate of the background seismicity for the area of interest.

References

Aagaard BT, Anderson G, and Hudnut KW (2004) Dynamic rupture modeling of the transition from thrust to strike-slip motion in the 2002 Denali Fault earthquake, Alaska. *Bulletin of the Seismological Society of America* 94: S190–S201.

Anderson JG, Brune JN, Louie JN, *et al.* (1994) Seismicity in the western Great Basin apparently triggered by the Landers, California, earthquake, 28 June 1992. *Bulletin of the Seismological Society of America* 84: 863–891.

Anderson OL and Grew PC (1977) Stress corrosion theory of crack propagation with applications to geophysics. *Reviews of Geophysics* 15: 77–104.

Antonioli A, Cocco M, Das S, and Henery C (2002) Dynamic stress triggering during the great 25 March 1998 Antarctic Plate earthquake. *Bulletin of the Seismological Society of America* 92: 896–903.

Armstrong BH and Stierman DJ (1989) Acoustic emissions from foreshocks and secular strain changes prior to earthquakes. In: Reginald HJ (ed.) *Acoustic Emission/ Microseismic Activity in Geologic Structures and Materials*, vol. 17, pp. 309–326. Pennsylvania: Trans Tech Publications, Pennsylvania State University.

Arnadottir T, Geirsson H, and Einarsson P (2004) Coseismic stress changes and crustal deformation on the Reykjanes Peninsula due to triggered earthquakes on 17 June 2000. *Journal of Geophysical Research* 109: 12 (doi:10.1029/2004JB003130).

Atkinson BK (1984) Subcritical crack growth in geological materials. *Journal of Geophysical Research* 89: 4077–4114.

Bagdassarov N (1994) Pressure and volume changes in magmatic systems due to the vertical displacement of compressible materials. *Journal of Volcanology and Geothermal Research* 63: 95–100.

Bakun WH and McGarr A (2002) Differences in attenuation among stable contentenal regions. *Geophysical Research Letters* 29: 36-1–36-4.

Beeler NM, Hickman SH, and Wong T-F (2001) Earthquake stress drop and laboratory-inferred interseismic strength recovery. *Journal of Geophysical Research* 106: 30701–30713.

Beeler NM and Lockner DA (2003) Why earthquakes correlate weakly with the solid Earth tides: Effects of periodic stress on the rate and probability of earthquake occurrence. *Journal of Geophysical Research* 108: 17 (doi:1029/2001JB001518).

Beresnev IA and Wen K-L (1995) Remotely triggered seismicity inferred from Taiwan regional catalog. *Geophysical Research Letters* 22: 3155–3158.

Brodsky EE (2006) Long-range triggered earthquakes that continue after the wave train passes. *Geophysical Research Letters* 33: 4 (doi: 10.1029/2006GL026605).

Brodsky E, Sturtevant B, and Kanamori H (1998) Earthquakes, volcanones, and rectified diffusion. *Journal of Geophysical Research* 103: 23827–23838.

Brodsky EE, Karakostas V, and Kanamori H (2000) A new observation of dynamically triggered regional seismicity: Earthquakes in Greece following the August, 1999, Izmit, Turkey earthquake. *Geophysical Research Letters* 27: 2741–2744.

Brodsky EE and Prejean SG (2005) New constraints on mechanisms of remotely triggered seismicity at Long Valley Caldera. *Journal of Geophysical Research* 110: 14 (doi:10.1029/2004JB003211).

Brodsky EE, Roeloffs E, Woodcock D, Gall I, and Manga M (2003) A mechanism for sustained groundwater pressure changes induced by distant earthquakes. *Journal of Geophysical Research* 108: 10 (doi:10.1029/2002JB002321).

Byerlee JD (1980) Friction of rocks. *Pure and Applied Geophysics* 116: 615–626.

Camelbeeck T, van Eck T, Pelzing R, *et al.* (1994) The 1992 Roermond earthquake, The Netherlands, and its aftershocks. *Geologie en Mijnbouw* 73: 181–197.

Chouet B (1992) A seismic model for the source of long-period events and harmonic tremor. In: Gasparini P, Scarpa R, and Aki K (eds.) *Volcanic Seismology*, pp. 133–156. Berlin, Heidelberg, New York: Springer-Verlag.

Chouet B, Dawson P, and Nakano M (2006) Dynamics of diffusive bubble growth and pressure recovery in a bubbly rhyolitic melt embedded in an elastic solid. *Journal of Geophysical Research* 20.

Christiansen L, Hurwitz S, Saar MO, Ingebritsen SE, and Hsieh P (2005) Seasonal seismicity at western United States volcanic centers. *Earth and Planetary Science Letters* 240: 307–321.

Cocco M and Rice JR (2002) Pore pressure and poroelasticity effects in Coulomb stress analysis of earthquake interactions. *Journal of Geophysical Research* 107: 17 (doi:10.1029/2002JB002319).

Cochran ES, Vidale JE, and Tanaka S (2004) Earth tides can trigger shallow thrust fault earthquakes. *Science* 306: 1164–1166.

Das S and Scholz C (1981) Off-fault aftershock clusters caused by shear stress increase? *Journal of Geophysical Research* 71: 1669–1675.

Dieterich JH (1979) Modeling of rock friction. 1. Experimental results and constitutive equations. *Journal of Geophysical Research* 84: 2161–2168.

Eberhart-Phillips D, Haeussler PJ, Freymueller JT, *et al.* (2003) The 2002 Denali Fault earthquake, Alaska: A large magnitude, slip-partitioned event. *Nature* 300: 1113–1118.

Emiliani C, Harrison CGA, and Swanson M (1969) Underground nuclear explosions and the control of earthquakes. *Science* 165: 1255–1256.

Felzer KR and Brodsky EE (2005) Testing the stress shadow hypothesis. *Journal of Geophysical Research* 110: 13 (doi:10.1029/2004JB003277).

Felzer KR and Brodsky EE (2006) Evidence for dynamic aftershock triggering from earthquake densities. *Nature* 441: 735–738.

Fournier RO (1999) Hydrothermal processes related to movement of fluid from plastic to brittle rock in the magmatic–epithermal environment. *Economic Geology* 94: 1193–1211.

Freed AM (2005) Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annual Reviews of Earth and Planetary Science* 33: 335–367.

Galperin EI, Petersen NV, Sitnikov AV, and Vinnik LP (1990) On the properties of short-period seismic noise. *Physics of the Earth and Planetary Interiors* 63: 163–171.

Glowacka E, Nava AF, Cossio Dd, Wong V, and Farfan F (2002) Fault slip, seismicity, and deformation in the Mexicali Valley, Baja California, Mexico, after the M 7.1 Hector Mine earthquake. *Bulletin of the Seismological Society of America* 92: 1290–1299.

Gomberg J (1996) Stress/strain changes and triggered seismicity following the Mw 7.3 Landers, California, earthquake. *Jounral of Geophysical Research* 101: 751–764.

Gomberg J, Blanpied ML, and Beeler NM (1997) Transient triggering of near and distant earthquakes. *Bulletin of the Seismological Society of America* 87: 294–309.

Gomberg J, Bodin P, Larson K, and Dragert H (2004) Earthquakes nucleated by transient deformations caused by the $M = 7.9$ Denali, Alaska, earthquake. *Nature* 427: 621–624.

Gomberg J, Bodin P, and Reasenberg PA (2003) Observing earthquakes triggered in the near field by dynamic deformations. *Bulletin of the Seismological Society of America* 93: 118–138.

Gomberg J and Davis S (1996) Stress/strain changes and triggered seismicity at The Geysers, California. *Journal of Geophysical Research* 101: 733–749.

Gomberg J and Johnson P (2005) Dynamic triggering of earthquakes. *Nature* 437: 830.

Gomberg J, Reasenberg PA, Bodin P, and Harris R (2001) Earthquakes triggering by seismic waves following the Landera and Hector Mine earthquakes. *Nature* 411: 462–465.

Gomberg J, Reasenberg PA, Cocco M, and Belardinelli ME (2005) A frictional population model of seismcity rate change. *Journal of Geophysical Research* 110: 10 (doi:10.1029/2004JB003404).

Green HWI (2003) Tiny triggers deep down. *Nature* 424: 893–894.

Hainzl S and Ogata Y (2005) Detecting fluid signals in seismicity data through statistical earthquake modeling. *Journal of Geophysical Research* 110: 10 (doi:10.1029/2004JBOO3247).

Hainzl S, Sherbaum F, and Beauval C (2006) Estimating background activity based on intervent-time distributions. *Bulletin of the Seismological Society of America* 96: 313–320.

Harrington RM and Brodsky EE (2006) The absence of remotely triggered seismicity in Japan. *Bulletin of the Seismological Society of America* 96: 871–878.

Harris RA and Day SM (1993) Dynamics of fault interaction: Parallel strike-slip faults. *Journal of Geophysical Research* 98: 4461–4472.

Harris RA, Dolan JF, Hartlib R, and Day SM (2002) The 1999 Izmit, Turkey, earthquake: A 3-D dynamic stress transfer model of interearthquake triggering. *Bulletin of the Seismological Society of America* 92: 245–255.

Harris RA (1998) Introduction to a special section: Stress triggers, stress shadows, and implications for seismic hazards. *Journal of Geophysical Research* 103: 24347–24358.

Hill DP (1977) A model for earthquake swarms. *Journal of Geophysical Research* 82: 1347–1352.

Hill DP (2006) Unrest in Long Valley Caldera, California: 1978–2004. In: De Natale G, Troise C, and Kilburn CRJ (eds.) *Geological Society of London Special Publication*, Vol. 269, *Mechanisms of Activity and Unrest at Large Calderas*, pp. 1–24. London: Geological Society of London.

Hill DP, Johnston MJS, Langbein JO, and Bilham R (1995) Response of Long Valley Caldera to the $M_w = 7.3$ Landers, California, earthquake. *Journal of Geophysical Research* 100: 12985–13005.

Hill DP, Pollitz F, and Newhall C (2002) Earthquake–volcano interactions. *Physics Today* 55: 41–47.

Hill DP, Reasenberg PA, Michael A, *et al.* (1993) Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* 260: 1617–1623.

Hough SE (2001) Triggered earthquakes and the 1811–1812 New Madrid, central United States, earthquake sequence. *Bulletin of the Seismological Society of America* 91: 1547–1581.

Hough SE (2005) Remotely triggered earthquakes following moderate mainshocks (or why California is not falling into the ocean). *Seismological Research Letters* 76: 58–66.

Hough SE, Billham R, Ambraseys N, and Feldl N (2005) Revisiting the 1897 Shillong and 1905 Kangra earthquakes in northern India: Site response, Moho reflections and a triggered earthquake. *Current Science* 88: 1632–1638.

Hough SE and Kanamori H (2002) Source properties of earthquakes near the Salton Sea triggered by the 16 October 1999 M_w 7.1 Hector Mine, California, earthquake. *Bulletin of the Seismological Society of America* 92: 1281–1289.

Hough SE, Seeber L, and Armbruster JG (2003) Intraplate triggered earthquakes: Observations and interpretation. *Bulletin of the Seismological Society of America* 93: 2212–2221.

Hudnut KWL, Seeber L, and Pacheco J (1989) Cross-fault triggering in the November 1978 Superstition Hill earthquake sequence, southern California. *Geophysical Research Letters* 16: 199–202.

Husen S, Taylor R, Smith RB, and Healser H (2004a) Changes in geyser eruption behavior and remotely triggered seismicity in Yellowstone National Park produced by the 2002 M 7.9 Delani fault earthquake, Alaska. *Geology* 32: 537–540.

Husen S, Wiemer S, and Smith RB (2004b) Remotely triggered seismicity in the Yellowstone National Park region by the 2002 Mw 7.9 Denali Fault earthquake, Alaska. *Bulletin of the Seismological Society of America* 94: S317–S331.

Husker AL and Brodsky EE (2004) Seismcity in Idaho and Montana triggered by the Denali Fault earthquake: A window into the geologic context for seismic triggering. *Bulletin of the Seismological Society of America* 94: S310–S316.

Ichihara M and Brodsky EE (2006) A limitation on the effect of rectified diffusion in volocanic systems. *Geophysical Research Letters* 33: 4 (doi:10.1029/2005GL024733).

Johnson P and Jia X (2005) Nonlinear dynamic, granular media and dynamic earthquake triggering. *Nature* 437: 871–874.

Johnson P, Jai X, and Gomberg J (submitted) The role of nonlinear dynamics in dynamic earthquake triggering. *Journal of Geophysical Research*.

Johnston MJS, Hill DP, Linde AT, Langbein J, and Bilham R (1995) Transient deformation during triggered seismicity from the 28 June 1992 $M_w = 7.3$ Landers earthquake at Long Valley volcanic caldera, California. *Bulletin of the Seismological Society of America* 85: 787–795.

Johnston MJS, Hill DP, and Pitt AM (2000) Strain transient recorded in the Long Valley Caldera during triggered seismicity from the October 16, 1999 M7.1 Hector Mine, California, earthquake. *Transactions of the American Geophysical Union, EOS* 81, WP1384.

Johnston MJS, Prejean SG, and Hill DP (2004a) Triggered deformation and seismic activity under Mammoth Mountain in Long Valley Caldera by the 3 November 2002 M_w 7.9 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S360–S369.

Johnston MJS, Prejean S, and Hill DP (2004b) Triggered deformation and seismic activity in Long Valley Caldera by the November 3, 2002, M 7.9 Denali Fault Earthquake. *Bulletin of the Seismological Society of America* 94: S360–S369.

Julian BR (1994) Volcanic tremor: Nonlinear excitation by fluid flow. *Journal of Geophysical Research* 99: 11859–11877.

Kanamori H and Brodsky EE (2004) The physics of earthquakes. *Reports in Progress in Physics* 67: 1429–1496.

Kilb D, Gomberg J, and Bodin P (2000) Triggering of earthquake aftershocks by dynamic stresses. *Nature* 408: 570–574.

Kilb D, Gomberg J, and Bodin P (2002) Aftershock triggering by complete Coulomb stress changes. *Journal of Geophysical Research* 107: 14 (doi:10.1029/2001JB000202).

King GCP and Cocco M (2001) Fault interactions by elastic stress changes: New clues from earthquake sequences. *Advances in Geophysics* 44: 1–38.

King GCP, Stein RS, and Lin J (1994) Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* 84: 935–953.

Kisslinger C and Jones LM (1991) Properties of aftershock sequences in southern California. *Journal of Geophysical Research* 96: 11947–11958.

Klein FW (1976) Earthquake swarms and the semidiurnal solid earth tide. *Geophysical Journal of the Royal Astronomical Society* 45: 245–295.

Lapusta N and Rice JR (2003) Nucleation and early seismic propagation of small and large events in a crustal earthquake model. *Journal of Geophysical Research* 108: 18 (doi:10.1029/2001JB000793).

Lay T, Kanamori H, Ammon CJ, *et al.* (2005) The great Sumatra–Andaman earthquake of 26 December 2004. *Science* 308: 1127–1132.

Linde AT and Sacks IS (1998) Triggering of volcanic eruptions. *Nature* 395: 888–890.

Linde AT, Sacks IS, Johnston MJS, Hill DP, and Billham RG (1994) Increased pressure from rising bubbles as a

mechanism for remotely triggered seismicity. *Nature* 371: 408–410.
Lockner DA and Beeler NM (1999) Premonitory slip and tidal triggering of earthquakes. *Journal of Geophysical Research* 104: 20133–20151.
Lockner DA and Beeler NM (2002) Rock failure and earthquakes. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology Part A*, pp. 505–537. Amsterdam: Academic Press.
Manga M and Brodsky EE (2006) Seismic triggering of eruptions in the far field: Volcanoes and geysers. *Annual Reviews of Earth and Planetary Science* 34: 263–291.
Mangan M and Sisson T (2000) Delayed, disequilibrium degassing in rhyolite magma: Decompression experiments and implications for explosive volcanism. *Earth and Planetary Science Letters* 183: 441–455.
Marsan D (2003) Triggering of seismicity at short distances following California earthquakes. *Journal of Geophysical Research* 108: 14 (doi:10.1029/2002JB001946).
Marzocchi W (2002) Remote seismic influence on large explosive eruptions. *Journal of Geophysical Research* 107: 7 (doi:10.1029/2001JB000307).
Matthews MV and Reasenberg PA (1988) Statistical methods for investigating quiescence and other temporal seismicity patterns. *Pure and Applied Geophysics* 126: 357–372.
McGarr A, Simpson D, and Seeber L (2002) Case histories of induced and triggered seismicity. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology, Part A*, pp. 647–664. Amsterdam: Academic Press.
Meltzner AJ and Wald DJ (2003) Aftershocks and triggered events of the great 1906 California earthquake. *Bulletin of the Seismological Society of America* 93: 2160–2186.
Miyazawa M and Mori J (2005) Detection of triggered deep low-frequency events from the 2003 Takachi-oki earthquake. *Geophysical Research Letters* 32: L10307.
Miyazawa M and Mori J (2006) Evidence suggesting fluid flow beneath Japan due to periodic seismic triggering from the 2004 Sumatra–Andaman earthquake. *Geophysical Research Letters* 33: 4 (doi:10.1029/2005GL025087), L05303-L.
Miyazawa M, Nakanishi I, Sudo Y, and Ohkura T (2005) Dynamic response of frequent tremors at Aso volcano to teleseismic waves from the 1999 Chi-Chi, Taiwan earthquake. *Journal of Volcanaology and Geothermal Research* 147: 173–186.
Mohamad R, Darkal AN, Seber D, Sandoval E, Gomez F, and Barazangi M (2000) Remote earthquake triggering along the Dead Sea Fault in Syria following the 1995 Gulf of Aqaba earthquake (Ms = 7.3). *Seismolological Research Letters* 71: 47–52.
Moran SC (2003) Multiple seismogenic processes for high-frequency earthquakes at Katmai National Park, Alaska: Evidence from stress tensor inversions of fault plane solutions. *Bulletin of the Seismological Society of America* 93: 94–108.
Moran SC, Power JA, Stihler SD, Sanchez JJ, and Caplin-Auerback J (2004) Earthquake triggering at Alaskan volcanoes following the 3 November 2002 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S300–S309.
Moran SC, Zimbelman DR, and Malone SD (2000) A model for the magmatic–hydrothermal system at Mount Rainier, Washington, from seismic and geochemical observations. *Bulletin of Volcanology* 61: 425–436.
Nicolaev AV and Troitskii PA (1987) Lithospheric studies based on array analysis of P-waves and microseisms. *Tectonophysics* 140: 103–113.
Nur A and Booker JR (1972) Aftershocks caused by pore fluid flow?. *Science* 175: 885–887.
Obara K (2002) Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* 296: 1679–1681.
Ogata Y (1993) Fast likelihood computation of epicemic type aftershock sequence model. *Geophysical Research Letters* 20: 2143–2146.
Omori F (1894) On the aftershocks of earthquakes. *Journal of the College of Science Imperial University of Tokyo* 7: 111–120.
Oppenheimer DH, Reasenberg PA, and Simpson RW (1988) Fault plane solutions for the 1984 Morgan Hill, California, earthquake sequence: Evidence for the state of stress on the Calavaras fault. *Journal of Geophysical Research* 93: 9007–9026.
Pankow KL, Arabasz WJ, Pechmann JC, and Nava SJ (2004) Triggered seismicity in Utah from the 3 November 2002 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S332–S347.
Perfettini HJ, Schmittbuhl J, and Cochard A (2003) Shear and normal load perturbations on a two-dimensional continuous fault. 2. dynamic triggering. *Journal of Geophysical Research* 108: 16 (doi:10.1029/2002JB001805).
Plafker G (1969) Tectonics of the March 27, 1964 Alaska earthquake. *US Geological Survey Professional Paper I* 543: I1–I74.
Pollitz FF and Johnston MJS (2006) Direct test of static stress versus dynamic stress triggering of aftershocks. *Geophysical Research Letters* 33: 4 (doi: 10:1029GL026764).
Pollitz FF and Sacks IS (2002) Stress triggering of the 1999 Hector Mine earthquake by transient deformation following the 1992 Landers earthquake. *Bulletin of the Seismological Society of America* 92: 1487–1496.
Power JA, Moran SC, McNutt SR, Stihler SD, and Sanchez JJ (2001) Seismic response of the Katmai volcanoes to the 6 December 1999 magnitude 7.0 Karlluk Lake earthquake, Alaska. *Bulletin of the Seismological Society of America* 91: 57–63.
Prejean SG, Hill DP, Brodsky EE, *et al.* (2004) Remotely triggered seismicity on the United States West Coast following the M_w 7.9 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S348–S359.
Pyle DM and Pyle DL (1995) Bubble migration and the initiation of volcanic eruptions. *Journal of Volcanology and Geothermal Research* 67: 227–232.
Reasenberg PA (1985) Second-order moment of central California seismicity, 1969–1982. *Journal of Geophysical Research* 90: 5479–5495.
Reasenberg PA and Simpson RW (1992) Response of regional seismicity to the static stress change produced by the Loma-Prieta earthquake. *Science* 255: 1687–1690.
Reinecker J, Heidbach O, Tingay M, Connolly P, and Muller B (2004) The 2004 release of the World Stress Map. www.world-stress-map.org (accessed Oct 2006).
Rivera L and Kanamori H (2002) Spatial heterogeneity of tectonic stress in the crust. *Geophysical Research Letters* 39: 4 (doi:10.1029/2001GL013803).
Roeloffs E (1996) Poroelastic techniques in the study of earthquake-related hydrologic phenomena. *Advances in Geophysics* 37: 135–195.
Rowe C, Christensen D, and Carver GA (2004) Preface to the issue dedicated to the 2002 Denali Fault earthquake sequence. *Bulletin of the Seismological Society of America* 94: S1–S4.
Rybicki K, Kato T, and Kasahara K (1985) Mechanical interaction between neighboring active faults – static and dynamic stress field induced by faulting. *Bulletin of the Earthquake Research Institute, Univeristy of Tokyo* 60: 1–21.
Sahagian DL and Proussevitch AA (1992) Bubbles in volcanic systems. *Nature* 359: 485.

Sanchez JJ and McNutt SR (2004) Intermediate-term declines in seismicity at Mt. Wrangell and Mt. Veniamin of volcanoes, Alaska, following the 3 November 2002 Mw 7.9 Denali Fault earthquake. *Bulletin of the Seismological Society of America* 94: S370–S383.

Savage JC and Clark MM (1982) Magmatic resurgence in Long Valley Caldera, eastern California. *Science* 217: 531–533.

Scholz CH (1998) Earthquakes and friction laws. *Nature* 391: 37–42.

Segall P, Jonsson S, and Agustsson K (2003) When is the strain in the rock the same as the strain in the rock? *Geophysical Research Letters* 30: 4 (doi:1029/2003GL017995).

Shapiro SA, Huenges E, and Borm G (1997) Estimating the crust permeability from fluid-injection-induced seismic emission at the KTB site. *Geophysical Journal International* 131: F15–F18.

Shimomura Y, Nishimura T, and Sato H (2006) Bubble growth processes in magma surrounded by an elastic medium. *Journal of Volcanology and Geothermal Research* 155: 307–322.

Sibson RH (1982) Fault zone models, heat flow, and the depth distribution of earthquakes in the continental crust of the United States. *Bulletin of the Seismological Society of America* 72: 151–163.

Singh SK, Anderson JG, and Rodriguez M (1998) Triggered seismicity in the Valley of Mexico from major Mexican earthquakes. *Geofiscia International* 37: 3–15.

Song S-R, Tsao S, and Lo H-J (2000) Characteristics of the Tatun volcanic eruptions, northern Taiwan: Implications for a caldron formation and volcanic evolution. *Journal of the Geological Society of China* 43: 361–378.

Spudich P, Steck LK, Hellweg M, Fletcher JB, and Baker LM (1995) Transient stresses at Parkfield, California, produced by the M 7.4 Landers earthquake of June 28, 1992: Observations from the UPSAR dense seismograph array. *Journal of Geophysical Research* 100: 675–690.

Stark MA and Davis SD (1995) Remotely triggered microearthquakes at The Geysers geothermal field, California. *Geophysical Research Letters* 23: 945–948.

Steacy S, Gomberg J, and Cocco M (2005) Introduction to special section: Stress transfer, earthquake triggering, and time-dependent seismic hazard. *Journal of Geophysical Research* 110: 12 (doi:10.1029/2005JB003692).

Steeples DW and Steeples DD (1996) Far field aftershocks of the 1906 earthquake. *Bulletin of the Seismological Society of America* 86: 921–924.

Stein RS (1999) The role of stress transfer in earthquake occurrence. *Nature* 402: 605–609.

Stein RS and Lisowski M (1983) The 1979 Homestead Valley earthquake sequence, California: Control of aftershocks and postseismic deformation. *Journal of Geophysical Research* 88: 6477–6490.

Stierman DJ (1977) A study of stress-dependent velocity variations *in situ*. *Department of Geophysics*, 199 pp. Stanford University, Stanford, CA.

Sturtevant B, Kanamori H, and Brodsky E (1996) Seismic triggering by rectified diffusion in geothermal systems. *Journal of Geophysical Research* 101: 25269–25282.

Tanaka S, Ohtake M, and Sato H (2004) Tidal triggering of earthquakes in Japan related to the regional tectonic stress. *Earth, Planets, and Space* 56: 511–515.

Tibi R, Wiens DA, and Inoue H (2003) Remote triggering of deep earthquakes in the 2002 Tonga sequence. *Nature* 424: 921–925.

Tolstoy M, Vernon FL, Orcutt JA, and Wyatt FK (2002) Breathing of the seafloor: Tidal correlations of seismcity at Axial volcano. *Geology* 30: 503–506.

Ukawa M, Fujita E, and Kumagai T (2002) Remote triggering of microearthquakes at the Iwo-Jima volcano. *Journal of Geography* 111: 277–286.

Vohis RC (1967) *Hydrologic Effects of the Earthquake of March 27, 1964, outside Alaska*. Washington, DC: US Geological Survey.

Voisin C, Campillo M, Ionescu IR, Cotton F, and Scotti O (2000) Dynamic versus static stress triggering and friction parameters: Inferences from the November 23, 1980, Irpinia earthquake. *Journal of Geophysical Research* 105: 21647–21658.

Voisin C, Cotton F, and Di Carli S (2004) A unified model for dynamic and static stress triggering of aftershocks, antishocks, remote seismicity, creep events, and multisegmented rupture. *Journal of Geophysical Research* 109: 12 (doi:10.1029/2003JB002886).

Wen K-L, Beresnev IA, and Cheng S (1996) Moderate-magnitude seismicity remotely triggered in the Taiwan Region by large earthquakes around the Philippine Sea Plate. *Bulletin of the Seismological Society of America* 86: 843–847.

West M, Sanchez JJ, and McNutt SR (2005) Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake. *Science* 308: 1144–1146.

Wiemer S (2001) A software package to analyze seismicity: ZMAP. *Seismological Research Letters* 72: 373–382.

Wiemer S and Wyss M (2000) Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America* 90: 859–869.

Zoback MD and Zoback ML (2002) State of stress in the Earth's lithosphere. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology, Part A*, pp. 559–568. Amsterdam: Academic Press.

Ziv A (2006) On the role of multiple interactions in remote aftershock triggering: The Landers and the Hector Mine case studies. *Bulletin of the Seismological Society of America* 96: 80–89.

4.10 Earthquake Hydrology

M. Manga and C.-Y. Wang, University of California Berkeley, Berkeley, CA, USA

4.10.1 Introduction

For thousands of years, changes in the amount, direction, and rate of water flow in streams and in fluid pressure in the subsurface have been documented following large earthquakes. Examples include the formation of new springs, disappearance of previously active springs, increased discharge in streams by an order of magnitude that persists for weeks or even months, fluctuations in water levels in wells by several meters, and permanent changes in water levels in wells coincident with earthquakes located thousands of kilometers away. Such observations, although often simply assigned to the category of curious phenomena, provide unique insight into hydrogeologic processes and, in particular, their correlation with tectonic processes at spatial and temporal scales that otherwise could not be studied.

Hydrologic responses are not unexpected: earthquakes cause strain, and strain changes fluid pressure and alters hydrogeologic properties such as permeability, which controls the rate of fluid flow. What is more surprising about many hydrologic responses, however, is their large amplitude and the great distances over which they occur.

This review is divided into two main parts. First we describe the processes by which fluid pressure and hydrologic properties can change following an earthquake. Next we review hydrologic phenomena attributed to earthquakes and summarize our current state of understanding about the processes that cause these phenomena. We then conclude with a brief discussion of hydrologic precursors.

4.10.2 Hydrologic Response to Stress

Earthquakes change the static stress (i.e., the offset of the fault generates a static change in stress in the crust) and also cause dynamic stresses (from the

Table 1 Amplitude and period of stress changes of earthquakes compared with other sources of stress

		Stress (MPa)		Period
Solid earth tides		10^{-3}		12 h
Ocean tides		10^{-2}		12 h
Hydrological loading		10^{-3} to 10^{-1}		Days to years
Glacier loading		10^{1} to 10^{2}		10^{3} years
	10^{2} km	10^{3} km	10^{4} km	
Static stress changes, M8	10^{-1}	10^{-4}	10^{-7}	NA
Dynamic stress changes, M8	3	0.06	0.001	20 s

(from Manga M, and Brodsky EE (2006) Seismic triggering of eruptions in the far field: Volcanoes and geysers. *Annual Review of Earth and Planetary Sciences* 34: 263–291.)

seismic waves). Both stresses increase as the seismic moment of the earthquake, but they decay very differently with distance r. Static stresses decrease as $1/r^3$. By comparison, dynamic stresses, which are proportional to the seismic wave amplitude, decrease more gradually. A standard empirical relationship between surface-wave amplitude and magnitude has dynamic stress decreasing as $1/r^{1.66}$ (e.g., Lay and Wallace, 1995). Other factors such as directivity, radiation pattern, and crustal structure will also influence the amplitude of shaking. However, the significant difference in the dependence on distance is a robust feature distinguishing static and dynamic stresses.

Table 1, reproduced from Manga and Brodsky (2006), lists and compares the magnitude of earthquake stresses with other external stresses of hydrologic interest. Both static and dynamic stresses may be significant in the near and intermediate field (within up to a few fault lengths), but only dynamic stresses are larger than tidal stress in the far-field (many fault lengths away).

4.10.2.1 Effect of Static Stress

4.10.2.1.1 Poroelastic flow and deformation

In response to stresses, porous solids deform, the pressure of fluids within pores changes, and pore fluids can flow. The basic theory of poroelasticity, first developed by Biot (1941, 1956a, 1956b), describes the coupling between changes in stress σ, strain ϵ, pore pressure p, and the material properties that relate these three variables. A complete discussion and derivation of the theory, even its simplest forms, is beyond the scope of this review, but several key relationships are useful for understanding the hydrological response to earthquakes and interpreting observations. Wang's (2000) text on linear poroelasticity, and review papers by Rice and Cleary (1976), Roeloffs (1996), and Neuzil (2003), provide thorough and pedagogic reviews.

Consider an isothermal, linear, isotropic poroelastic material, subjected to a stress, σ_{ij}. The material will respond by deforming with the total strain ϵ_{ij}, given by

$$\epsilon_{ij} = \frac{1}{2G}\left[\sigma_{ij} - \frac{\nu}{1+\nu}\sigma_{kk}\delta_{ij}\right] + \frac{\alpha}{3K}p\delta_{ij} \qquad [1]$$

where G is the shear modulus, K is the bulk modulus, ν is the Poisson's ratio, and α is the Biot–Willis coefficient, which is the ratio of the increment in fluid content to the volumetric strain at constant pore pressure. The increment in fluid content, denoted ψ, is the change in fluid mass per unit volume divided by the density of the fluid at the reference state and is given by

$$\psi = \frac{\alpha}{3K}\sigma_{kk} + \frac{\alpha}{KB}p \qquad [2]$$

B is called Skempton's coefficient and has a value between 0 and 1, with 'hard rocks' such as sandstone and granite having values between about 0.5 and 0.9 and unconsolidated materials having a value close to 1. The variable B is related to the porosity ϕ and to the compressibilities of the pore fluid, solid grains, and saturated rock. The sign convention is chosen so that negative σ_{kk} indicates compression.

First we consider the so-called 'undrained' limit in which there is no flow of pore fluid, that is, $\psi = 0$ (the opposite extreme, termed 'drained', corresponds to the limit in which there is no change in fluid pressure, that is, $p = 0$). The change in pore pressure p caused by a change in mean stress σ_{kk} can be found from eqn [2]:

$$p = -\frac{B}{3}\sigma_{kk} \qquad [3]$$

Equation [3] can be expressed in terms of the volumetric strain ϵ_{kk} using constitutive relationship [1] and physical relationship [2]:

$$p = -\frac{B}{3}\sigma_{kk} = -BK_u\epsilon_{kk} = -\frac{2GB}{3}\frac{1+\nu_u}{1-2\nu_u}\epsilon_{kk} \qquad [4]$$

where the subscript u indicates a value under undrained conditions. The variable G should have the same value in drained and undrained conditions for a linear poroelastic material.

If there are gradients of fluid pressure, there will be a flux of fluid $\mathbf{q}$ (volume/area per unit time) governed by the Darcy's equation

$$\mathbf{q} = -\frac{k}{\mu}\nabla p \qquad [5]$$

where the pressure p is the excess pressure, that is, the pressure difference from hydrostatic, k is permeability, and μ is fluid viscosity; p, as defined, is equivalent to the hydraulic head multiplied by ρg, where ρ is the density of the fluid and g the gravitational acceleration. Darcy's equation, coupled with the continuity equation governing conservation of mass, if the permeability k and μ are constant, leads to pore pressure being governed by an inhomogeneous diffusion equation

$$\frac{\alpha}{KB}\left[\frac{B}{3}\frac{\partial\sigma_{kk}}{\partial t}+\frac{\partial p}{\partial t}\right] = \frac{k}{\mu}\nabla^2 p \qquad [6]$$

The term involving the time derivative of stress couples the mechanical deformation and the fluid flow. The time derivative of stress will not appear in the case of uniaxial strain and constant vertical stress, both reasonable approximations for near-surface aquifers (but not rigorously correct if there is two- or three-dimensional flow), and can be neglected for very compressible fluids such as gases (Wang, 2000). In these cases, the evolution of fluid pressure resulting from the flow is uncoupled from that for the stress and can be solved independently. Then eqn [6] reduces to the form commonly used to model groundwater flow:

$$S\frac{\partial p}{\partial t} = \frac{k\rho g}{\mu}\nabla^2 p \quad or \quad \frac{\partial p}{\partial t} = D\nabla^2 p \qquad [7]$$

where S is standard hydrogeologic specific storage and is equal to volume of water released per unit change in head while maintaining no lateral strain and constant vertical stress, and D is called the hydraulic diffusivity.

Assuming that eqn [7] is a good approximation, the timescale τ for changes in pore pressure to diffuse or relax can be calculated by $\tau \sim L^2/D$, where L is the length scale over which the pressure diffuses. Typical values of D are 10^{-10}–$10^4\,m^2\,s^{-1}$, with small values characterizing low-permeability rocks (e.g., shale), and large values characterizing coarse, unconsolidated sediments or highly permeable rocks (e.g., gravel and karst, respectively). A value of $10^0\,m^2\,s^{-1}$ would be typical of sandstone aquifers. For a 100 m thick aquifer, then, pore pressure changes will diffuse on timescales that range from minutes to many decades, and with a timescale of hours for a sandstone aquifer.

On examining hydrological observations, we see that many of the most dramatic hydrologic phenomena mentioned in Section 4.10.1, and in particular those that involve water discharge at the Earth's surface, cannot be explained by traditional linear poroelasticity theory. The reason is quite straightforward. Consider the case in which a stress change σ_{kk} causes an undrained pressure change p given by eqn [3]. For a (very large) static stress change of 1 MPa (see **Table 1**), eqn [3] implies pressure changes of 0.1–0.3 MPa. This would cause a 10–30 m change in water level in a well. To bring the pore pressure back to its original value requires the removal of an increment of fluid content

$$\psi = \frac{3(\nu_u-\nu)}{2GB(1+\nu)(1+\nu_u)}\sigma_{kk} \qquad [8]$$

Using values for Berea sandstone (Wang, 2000) of $\nu = 0.17$, $\nu_u = 0.33$, $G = 5.6$ GPa, and $B = 0.75$, we obtain $\psi = 3.5 \times 10^{-5}$. For a 100 m thick aquifer with porosity $\phi = 0.4$, this is equivalent to about 1 mm of water per unit area discharged at the surface–an amount that is small compared to that needed to explain the excess discharge in stream following large earthquakes (see table 1 in Manga (2001)).

Equations [1]–[3] apply for linear, isotropic porous materials. In this limit, shear stresses cause no change in pore pressure. In natural rocks, however, shear stresses can induce changes in pore pressure because of anisotropy or the nonlinear elastic behavior of rocks (e.g., Hamiel *et al.*, 2005).

4.10.2.1.2 Permanent deformation

At shear strains from 10^{-4} to 10^{-3}, brittle rocks, sediments, and soils show permanent deformation. At even greater shear strains ($\sim 10^{-2}$), failure occurs. The changes in porosity of sediments and soils

during permanent deformation may be many orders of magnitude greater than that during elastic deformation. Thus, under undrained conditions, permanent deformation can be associated with significant changes in pore pressure and groundwater flow.

During permanent deformation, the cohesion among the grains of sediments and soils is disrupted and the grains tend to move to reach a new state of equilibrium. The basic characteristics of deformation, however, depend on the consolidation state of sediments and soils. For loose deposits, shear deformation causes grains to move into pre-existing pores. This reduces the original porosity, and thus the volume of sediments (soils) decreases–a process commonly known as consolidation. Shear deformation of dense deposits, in contrast, will cause sediment grains to roll over each other and, in so doing, create new porosity; thus, the volume of sediments (soils) increases – a process commonly known as dilatancy. In both cases, a critical state is eventually reached in undrained deformation, beyond which porosity no longer changes with continued shearing.

The deformation behavior of brittle rocks is distinctly different from that of sediments and soils, leading to different predictions for the hydrologic and seismic consequences under applied stresses. Laboratory measurements show that beyond the elastic limit, shearing of consolidated rocks causes microcracks to open and the volume of rock to increase – that is, they become dilatant (e.g., Brace *et al.*, 1966). At still higher deviatoric stresses, microcracks may coalesce and localize into a shear zone, leading to eventual rupture (Lockner and Beeler, 2002). Repeated rupturing of the brittle crust along a fault zone can produce pulverized gouge material with significant porosity. Geologic investigations of exhumed fault zones in California (e.g., Chester *et al.*, 1993, 2004) and in Japan (Wibberley and Shimamoto, 2003; Forster *et al.*, 2003), and studies of drilled cores from fault zones (e.g., Ohtani *et al.*, 2001) have demonstrated that some fault zones are composed of fluid-saturated, porous fault gouge several hundred meters thick. Attempts to image the structure of fault zones by using conventional seismic reflection and refraction methods (e.g., Thurber *et al.*, 1997) suffered from not having high enough station and source densities to resolve deep fault-zone structures on the scale of hundreds of meters (Mizuno, 2003). More recent seismic imaging of the fault-zone structures relies on the use of fault-zone trapped waves (e.g., Li *et al.*, 2000, 2006; Mizuno, 2003) and high-frequency body waves (Li and Zhu, 2006). These studies have delineated a 100–250 m wide gouge zone along active faults in California (Li *et al.*, 2000; Li and Zhu, 2006) and in Japan (e.g., Mizuno, 2003), in which the seismic velocities are reduced by 30–50% from the wall-rock velocities. Direct drilling of the San Andreas Fault Zone to a depth of ~3 km near Parkfield, CA, also revealed a process zone 250 m in width, in which the seismic velocities are 20–30% below the wall-rock velocities (Hickman *et al.*, 2005).

Using high-resolution electromagnetic imaging, Unsworth *et al.*, (2000) showed that the San Andreas Fault Zone contains a wedge of low-resistivity material extending to several kilometers in depth; they further suggested a range of saturated porosity to account for the observed electrical resistivity. Using experimentally measured seismic velocity and electrical conductivity of fault gouge at elevated pressures to interpret the then-available seismic velocity and electrical resistivity of the San Andreas Fault Zone in central California, Wang (1984) suggested that the fault zone may consist of saturated fault gouge. Inverting the gravity anomaly across the San Andreas Fault Zone in central California and constraining the inversion by using existing seismic reflection profiles, Wang *et al.*, (1986) further showed that the fault zone may be characterized by relatively low density with a corresponding porosity greater than 10% extending to seismogenic depths.

That fault zones may be porous and saturated at seismogenic depths has important implications on a host of issues related to the dynamics of faulting (e.g., Sleep and Blanpied, 1992; Sibson, 1973; Lachenbruch, 1980; Mase and Smith, 1987; Brodsky and Kanamori, 2001; Andrew, 2003; Wibberley and Shimamoto, 2005) and the frictional strength (or weakness) of faults (e.g., Lachenbruch and Sass, 1977; Mount and Suppe, 1987; Zoback *et al.*, 1987; Rice, 1992). These topics, while beyond the objective of the present chapter, highlight the importance of understanding better the interaction between hydrology and earthquakes.

4.10.2.2 Effect of Dynamic Strain

4.10.2.2.1 Poroelastic deformation and fluid flow

Seismic waves cause spatial variations in strain and hence spatial variations in pore pressure. The resulting pore pressure gradients cause fluid flow. Seismic wave speeds in saturated rocks will thus differ from those in unsaturated rocks and will be frequency

dependent. As the fluid moves, energy is also dissipated by the fluid flow, which in turn is governed by the permeability of the porous material, the geometry of the pores, and the viscosity of the fluid. This contributes to seismic attenuation. Wave speed and attenuation can be calculated theoretically as a function of frequency for porous materials with different pore structures by solving the poroelastic equations coupled with Darcy's equation for fluid flow at the macroscopic scale. As frequency increases, pore scale (or 'squirt') flows become more important (e.g., Mavko *et al.*, 1998). These ideas were first quantified in a series of papers by Biot (e.g., Biot, 1941, 1956a, 1956b, 1962) and extended by many authors since, for example, by accounting for squirt flows (e.g., Dvorkin *et al.*, 1994). Additional important extensions to actual geological settings include accounting for heterogeneity with dual porosity models (e.g., Berryman and Wang, 2000) or fractal geometries (e.g., Masson and Pride, 2006) and partial or multiphase saturation (e.g., Pride *et al.*, 2004).

Changes in pore pressure and stress also change fracture and pore geometry, which in turn alters wave speeds and anisotropy. The frequency dependence of attenuation and wave speeds also depends on hydrogeologic properties such permeability and the nature of permeability heterogeneity. The relationship between seismic properties and hydrogeologic properties and fluid pressure involves so many parameters that inferring properties of the subsurface from seismic measurements is challenging. However, measurements of seismic attenuation, seismic anisotropy, and wave speeds can be used as a noninvasive probe to monitor changes in the subsurface (e.g., Liu *et al.*, 2004).

4.10.2.2.2 Permanent deformation

The dynamic deformation of sediments and soils, in contrast to the static deformation, is affected by inertial forces that depend on loading rates and number of loading cycles. An often-used measure of the total amount of seismic energy imparted to an engineered structure is the Arias intensity, which is proportional to the square of acceleration (Arias, 1970). Therefore, even if the magnitude of stresses is small, inertial forces may significantly affect deformation at high frequencies. For this reason, it is necessary in earthquake engineering to examine the dynamic deformation of sediments and soils for dynamic strains as small as 10^{-6}. Permanent volumetric deformation, however, does not occur until the magnitude of shear strain reaches some threshold. Dobry (1989) summarized undrained experimental results in which excess pore pressure developed during the application of 10 cycles of sinusoidal loads. He showed that, for different sands with a wide range of dry densities subjected to a wide range of effective confining pressures, pore pressure did not increase until the cyclic shear strain amplitude was increased to 10^{-4}. The increase in pore pressure indicates a decrease in pore volume, and hence consolidation. It is interesting to note that the strain threshold is comparable in magnitude to that for permanent deformation to occur in static deformation (Section 4.10.2.1).

The number of loading cycles during dynamic loading is another important factor in determining the mechanical responses of sediments and soils. Results of laboratory experiments show that during cyclic loading at constant strain amplitude, pore pressure in sediments and soils increases with the number of loading cycles. At decreasing strain amplitudes, a greater number of loading cycles is required to build up excess pore pressure to a given magnitude (Ishihara, 1996). During large earthquakes, about 10–20 cycles of major ground shaking occur, though the amplitude of each cycle will vary.

4.10.2.2.3 Liquefaction

When pore pressures approach the overburden pressure, soils lose their rigidity and become fluid-like – a phenomenon widely known as liquefaction (Seed and Lee, 1966; Terzaghi *et al.*, 1996), which is a major source of seismic hazard for engineered structures. Liquefied sediments and soils in the subsurface are associated with substantial pore pressure, as indicated by the height of the smear on the wall in **Figure 1**, left by ejected sands during the 1999 M 7.5 Chi-Chi earthquake in Taiwan. Most liquefaction occurs near the ruptured fault, but liquefaction can also occur at large distances from the earthquake epicenter; for example, ejection of fluidized sediments was observed during the 1964 M 9.2 Alaska earthquake at distances more than 400 km from the epicenter (Waller, 1966).

Field and laboratory studies show that the occurrence of liquefaction depends on many factors, such as earthquake magnitude, shaking duration, peak ground motion, depth to the groundwater table, basin structures, site effects, and liquefaction susceptibility of sediments (e.g., Youd, 2003). Thus, the occurrence of liquefaction is difficult to predict on a theoretical basis, and empirical approaches are, as a rule, adopted in assessing the liquefaction potential of an area. Various ground penetration tests (e.g.,

Figure 1 Smear left on a building wall created by ejected sand during the 1999 *M* 7.5 Chi-Chi earthquake in Taiwan. Motorcycle at lower left shows scale. Reproduced from Su T-C, Chiang K-W, Lin S-J, Wang F-G, and Duann S-W (2000) Field reconnaissance and preliminary assessment of liquefaction in Yuan-Lin area. *Sino-Geotechnics* 77: 29–38.

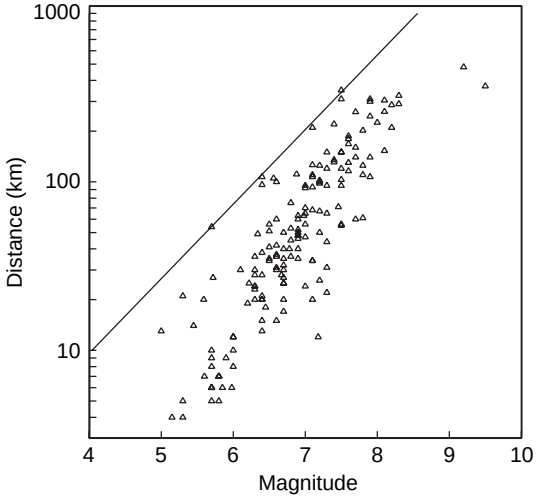


Figure 2 Relationship between earthquake moment magnitude and maximum distance between the epicenter and occurrences of liquefaction. The solid line is an empirical upper bound for the maximum distance over which shallow liquefaction has been observed (Wang *et al*., 2005a): $M = -5.0 + 2.26 \log R_{max}$, where R_{max} is in meters. Data shown based on the compilation in Wang *et al*. (2005a).

Bardet, 2003) are routinely applied at sites of engineering importance and in some metropolitan areas (e.g., Los Angeles and Memphis in the United States). On large scales, empirical relations between the magnitude of ground shaking and liquefaction (e.g., Wang *et al*., 2003) may be combined with numerical simulations of ground shaking during an earthquake (e.g., Stidham *et al*., 1999) to evaluate the regional liquefaction potential. Such applications, however, require extensive information about sediment properties and subsurface structures of the area of interest. In areas where such data are not available, a simpler approach may be applied to set some limits to the expected extent of liquefaction during potential earthquakes. Field observations show that, for earthquakes of a given magnitude M, the occurrence of liquefaction is confined within a particular distance from the earthquake epicenter, that is, the liquefaction limit R_{max} beyond which liquefaction may not be expected. **Figure 2** shows a compilation of data for the occurrence of liquefaction (Kuribayashi and Tatsuoka, 1975; Ambraseys, 1988; Papadopoulos and Lefkopulos, 1993; Galli, 2000; Wang *et al*., 2005a) and an empirical bound on the relationship between R_{max} and M. The liquefied sites at the liquefaction limit are likely to be those with the optimal conditions for liquefaction, that is, saturated soils with high liquefaction susceptibility. Liquefaction at closer distances may include less optimal conditions but are exposed to greater seismic input.

Because liquefaction damages engineered structures, there has been a concerted research effort, both in the laboratory and in the field, to understand the processes of liquefaction and to predict its occurrences (for an overview of these efforts, see National Research Council (1985), Bardet (2003), and Youd (2003)). In laboratory studies, sediment and soil samples are often obtained from the field and subjected to cyclic loading. At constant amplitude of deviatoric stress (stress minus the mean normal stress), deformation may be stable up to some number of cycles of shearing, but sudden onset of large shear deformation (onset of liquefaction) may occur if the cycles of shearing exceed some threshold number (e.g., National Research Council, 1985). This onset of liquefaction marks the moment when the stress path between the deviatoric stress and the effective stress (defined later in eqn [15], which decreases with increasing pore pressure) eventually intersects the failure envelope of sediments and soils.

These laboratory studies have helped to advance our understanding of the processes of liquefaction and to identify the major physical factors that affect liquefaction. However, it is difficult to duplicate natural conditions in the laboratory or to obtain undisturbed samples from the field. Thus laboratory studies may have limited practical application. In the field, various penetration tests, such as the 'standard' penetration test, are employed to assess the liquefaction potential of specific sites. Despite their wide

application, these tests are empirical. Thus, much progress can still be made toward a better physical understanding of the liquefaction process.

4.10.3 Observations and Their Explanations

Here we review hydrologic phenomena that accompany earthquakes. We will see that the magnitude of hydrological changes caused by earthquakes are likely to be of little human consequence, though depletion of groundwater induced by seismic activity has been invoked to explain the abandonment of regions of Crete, Greece, during the Late Minoan period (Gorokhovich, 2005).

We categorize observations by the type of observation (wells, streamflow, mud volcanoes, and geysers). It will also be helpful when discussing responses to distinguish hydrologic responses by the spatial relationship between the observation and the earthquake. We will use the terms near-field, intermediate-field, and far-field, for distances within one fault length, up to several fault lengths, and many fault lengths, respectively. In the near-field, changes in the properties of the fault zone itself may be responsible for hydrologic responses. Examples include the formation of new springs along ruptured faults (e.g., Lawson, 1908) or changes in groundwater flow paths because of changes in the permeability of or near the fault zone (e.g., Gudmundsson, 2000). In both the near- and intermediate-fields, dynamic and static strains are large enough that they might cause a measurable hydrologic response. In the far-field, however, only dynamic strains are sufficiently large to cause responses. In this chapter, we focus on hydrologic responses outside the fault zone.

Hydrologic changes following earthquakes modify pore pressures, which in turn can influence local seismicity. The topic of triggered seimicity is covered in the chapter by Hill and Prejean, and we simply note here that the mechanisms responsible for the observed hydrological phenomena discussed next may be connected to those responsible for distant, triggered seismicity.

4.10.3.1 Wells

The water level in wells measures the fluid pressure at depths the well is open to the surrounding formations. Several types of coseismic and postseismic responses are observed: water-level oscillations, coseismic changes in water level, and delayed changes in water level. **Figure 3** shows a particularly dramatic example of a response in a well in which water erupted to a height of 60 m above the surface following the 2004 *M* 9.2 Sumatra earthquake 3200 km away.

Figure 3 Well in China responding to the 2004 *M* 9.2 Sumatra earthquake 3200 km away. The picture was taken by Hou Banghua, Earthquake Office of Meizhou County, Guangdong, 2 days after the Sumatra earthquake. The fountain was 50–60 m high when it was first sighted, 1 day after the earthquake.

4.10.3.1.1 Water-level oscillations

Water wells can act like seismometers by amplifying ground motions, in particular long-period Rayleigh waves. Water-level fluctuations as large as 6 m (peak-to-peak amplitude) were recorded in Florida, thousands of kilometers away from the epicenter, during the 1964 Alaska earthquake (Cooper *et al.*, 1965). Hydroseismograms have been recorded since the early days of seismometer use (Blanchard and Byerly, 1935).

Figure 4 (from Brodsky *et al.*, 2003) shows water level in a well in Grants Pass, OR, following the 2002 *M* 7.9 Denali earthquake 3100 km away. Also shown is the vertical component of ground velocity measured on a broadband seismometer located adjacent to the well. Large water-level fluctuations, such as those shown in **Figure 4**, are unusual and occur only for the right combination of geometric (water depth in the well, well radius, aquifer thickness) and hydrologic properties (transmissivity). The frequency-dependent response of the water level in a well to dynamic strain in an aquifer can be determined theoretically by solving the coupled equations for pressure change in the aquifer, flow toward/away from the well, and flow within the well (e.g.,

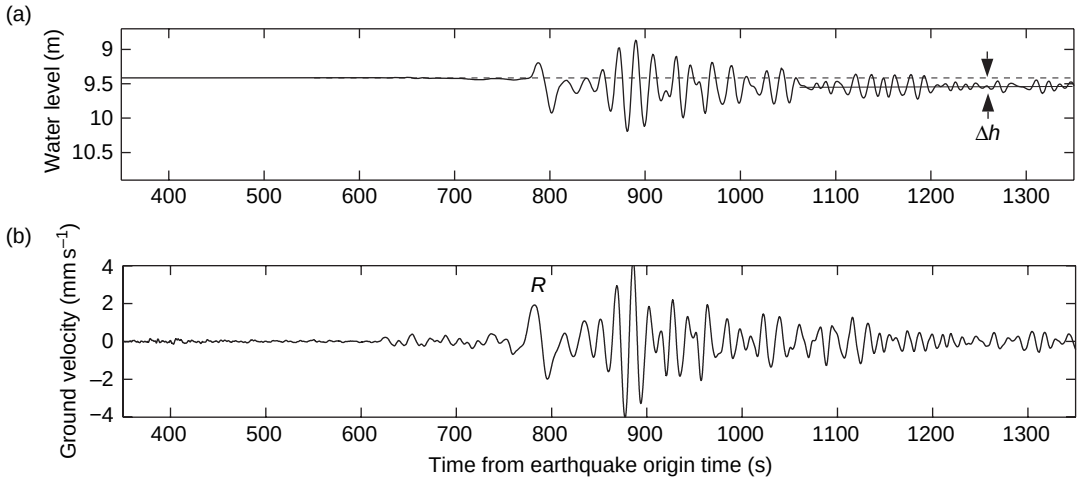


Figure 4 (a) Hydroseismograph recorded at 1 s intervals and (b) vertical component of ground velocity measured on an broadband seismometer at a well in Grants Pass, OR, following the 2002 *M* 7.9 Denali earthquake located 3100 km away. Δh shows the 12 cm permanent change in water level that followed the passage of the seismic waves. Figure based on figure 7 in Brodsky EE, Roeloffs E, Woodcock D, Gall I, and Manga M. (2003) A mechanism for sustained ground water pressure changes induced by distant earthquakes. *Journal of Geophysical Research* 108: doi:10.1029/2002JB002321.

Cooper *et al.*, 1965; Liu *et al.*, 1989). In general, high transmissivity favors large amplitudes.

Kono and Yanagidini (2006) found that in closed (as opposed to open) borehole wells, pore pressure variations are consistent with an undrained poroelastic response (see Section 4.10.2.1) at seismic frequencies. In addition, they found no pore pressure response to shear deformation.

4.10.3.1.2 Persistent and delayed postseismic changes in water level in the near and intermediate field

Persistent changes in water level in wells are probably the best-documented hydrologic responses to earthquakes. Significant advances in quantitative analysis of water-level changes during earthquakes have been made during the last decade (e.g., Roeloffs, 1998; King *et al.*, 1999; Roeloffs *et al.*, 2003; Brodsky *et al.*, 2003; Matsumoto and Roeloffs, 2003; Montgomery and Manga, 2003; Wang *et al.*, 2003; Sato *et al.*, 2004; Kitagawa *et al.*, 2006; Sil and Freymueller, 2006).

Sustained changes in water level in the near- and, possibly, intermediate-field can in some cases be explained by the coseismic static strain created by the earthquake. In this case, the water level will rise in zones of contraction, and fall in regions of dilation.

Figure 5 (from Jonsson *et al.*, 2003) shows a pattern of water-level change that mimics the pattern of coseismic volumetric strain after a strike-slip event in Iceland. An analogous correlation of volumetric strain and water-level changes was found after the 2003 *M* 8.0 Tokachi-oki thrust event in Japan (Akita and Matsumoto, 2004). Similar conclusions are reported by others (e.g., Wakita, 1975; Quilty and Roeloffs, 1997). This pattern of water-level change is not, however, universal. Following the 1999 *M* 7.5 Chi-Chi earthquake in Taiwan, the water level rose in most near-field wells where the coseismic strain should have caused aquifer dilation (Koizumi *et al.*, 2004). Following the 2004 *M* 9 Sumatra earthquake, more than 5000 km away in Japan, Kitagawa *et al.*, (2006) found that only about half the monitoring wells equipped with strain instruments recorded water-level changes consistent with the measured coseismic strain, implying that dynamic strains can also change water levels in the intermediate field.

Both the pattern and magnitude provide insight into the origin of water-level changes. The magnitude of expected water-level changes caused by coseismic strain can be determined from models of coseismic strain and calibrated models for well sensitivity that are calibrated on the basis of their response to barometric and tidal strains (e.g., Roeloffs, 1996). In some instances, recorded changes are similar to those predicted (e.g., Igarashi and Wakita, 1991), but changes can also be much larger than predicted, even in the near-field (e.g., Igarashi and Wakita, 1995).

The time evolution of water-level changes provides additional constraints on the location and magnitude of pore pressure changes. **Figure 6** (from

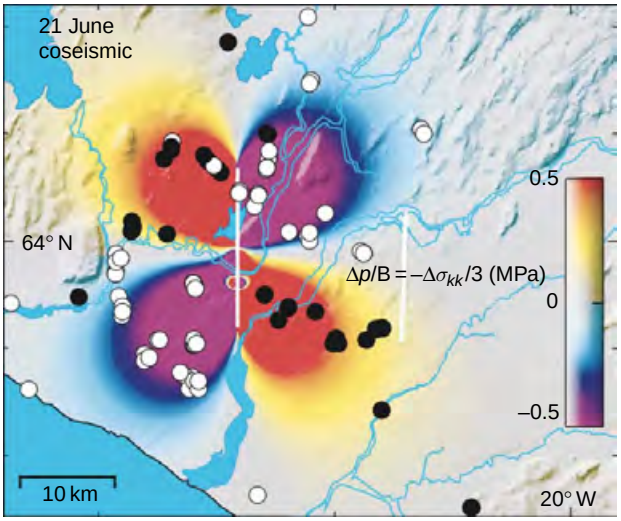


Figure 5 Predicted (color map) and observed (circles) coseismic water-level changes following a 2000 *M* 6.5 strike-slip earthquake in Iceland. Black and white circles indicate water-level increases and decreases, respectively. The white line shows the mapped surface rupture. Reproduced form Jonsson S, Segall P, Pedersen R, and Bjornsson G., (2003) Post-earthquake ground movements correlated to pore-pressure transients. *Nature* 424: 179–183.

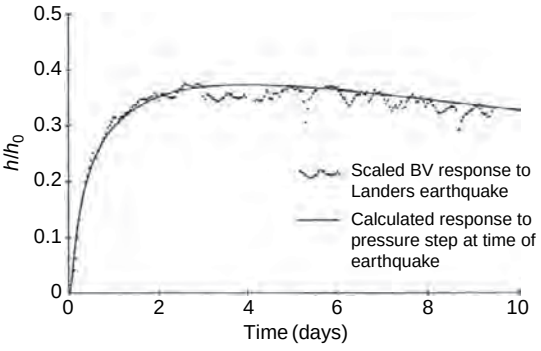


Figure 6 Water level as a function of time after the 1992 *M* 7.2 Landers, CA, earthquake. The epicenter is 433 km away from the well (named BV). The smooth curve shows the predicted evolution of water level for a coseismic, localized pressure change at some distance from the well. Water level is normalized by the magnitude of the head change where pore pressure changed. Reproduced from Roeloffs EA (1998) Persistent water level changes in a well near Parkfield, California, due to local and distant earthquakes. *Journal of Geophysical Research* 103: 869–889.

Roeloffs, 1998) shows water level for a period of 10 days following the 1992 *M* 7.3 Landers earthquake 433 km from the well. The observed water-level changes, shown with data points, can be modeled by a coseismic, localized pore pressure change at some distance from the well, as shown by the solid curve. The origin, however, of the hypothesized pressure changes cannot be determined from measurements at the well (Roeloffs, 1998). Other processes that can cause delayed or gradual changes in water level are localized changes in porosity and/or permeability (e.g., Gavrilenko *et al.*, 2000).

4.10.3.1.3 Near-field response of unconsolidated materials

At shallower depths and in unconsolidated materials, changes in water level in wells often exhibit larger amplitudes and different signs than predicted by the coseismic elastic strain. The 1999 Chi-Chi earthquake in Taiwan provided an excellent opportunity to examine models for postseismic water-level changes because of the very high density of monitored wells. We discuss these observations in much more detail than other types of well responses because the near-field hydrologic response at shallow depths and in unconsolidated materials is much less well studied.

As a fortuitous coincidence, a network of 70 evenly distributed hydrological stations (**Figure 7**), with a total of 188 wells, were installed in 1992 on a large alluvial fan (the Choshui River fan) near the Chi-Chi epicenter. The network was installed for the purpose of monitoring the groundwater resources over the Choshui River fan. At each station, one to five wells were drilled to depths ranging from 24 to 306 m to monitor groundwater level in individual aquifers. Each well was instrumented with a digital piezometer that automatically records the groundwater level hourly, with a precision of 1 mm (Hsu

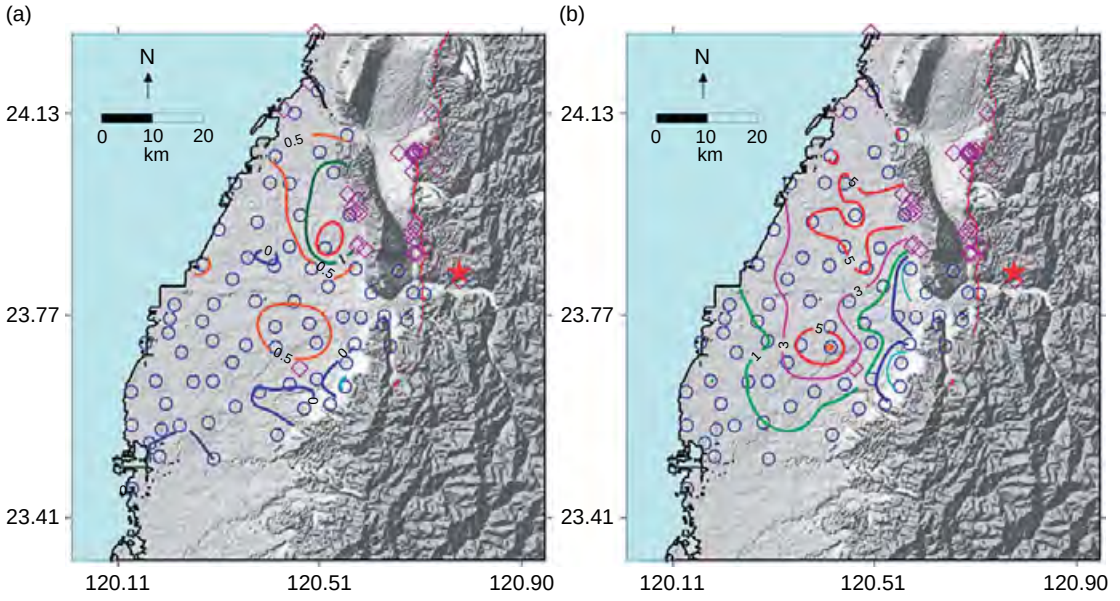


Figure 7 Spatial distribution of coseismic changes in groundwater level during the Chi-Chi earthquake. (a) Contours (m) showing coseismic groundwater-level changes in a confined aquifer (Aquifer II) in the Choshui River fan. (b) Contours (m) showing coseismic groundwater-level changes in the uppermost aquifer (Aquifer II) in the Choshui River fan. Groundwater monitoring stations are shown in open circles, liquefaction sites in open diamonds, epicenter of Chi-Chi earthquake in star, and the ruptured fault in discontinuous red traces. Reproduced from Wang C-Y Wang C-H, and Kuo C-Y (2004a) Temporal change in groundwater level following the 1999 $M_W = 7.5$ Chi-Chi earthquake, Taiwan. *Geofluids* 4: 210–220.

et al., 1999). The dense network of hydrological stations and its close proximity to a large earthquake made the groundwater records during the Chi-Chi earthquake the most comprehensive and systematic obtained so far. In addition, a dense network of broadband strong-motion seismographic stations captured the ground motion in the area, and detailed studies of liquefaction occurrences were made. Finally, the isotopic composition of the groundwater was measured before and after the earthquake (Wang *et al.*, 2005c).

The coseismic changes of groundwater level during this earthquake have been reported in several papers (Hsu *et al.*, 1999; Chia *et al.*, 2001; Wang *et al.*, 2001; Lee *et al.*, 2002). Near the ruptured fault, where consolidated sedimentary rocks occur near the surface, the groundwater level showed a coseismic, stepwise decrease (**Figures 8(a)** and **8(b)**). Away from the ruptured fault, in the fan of unconsolidated sediments, confined aquifers showed a distinctly different pattern of coseismic changes than in the uppermost aquifer. In the confined aquifers, groundwater levels showed a coseismic stepwise rise (**Figures 8(c)** and **8(d)**). The magnitude of this rise increases with distance from the ruptured fault, reaching more than 5 m at distances of 20–30 km away from the fault before decreasing at greater distances (**Figure 7(a)**; Wang *et al.*, 2001). In the uppermost aquifer, the coseismic changes in the groundwater level were much smaller in magnitude with an irregular pattern (**Figure 7(b)**). Liquefaction sites on the Choshui River fan are clearly correlated with significant coseismic rise of the groundwater level in the uppermost aquifer, but no such correlation exists for the confined aquifers (**Figure 7**). This observation strongly indicates that liquefaction occurred in the uppermost aquifer but not in the confined aquifers. In the basins nearest the ruptured fault (east of the Choshui River fan), a comparison between liquefaction sites and the coseismic change in the groundwater level cannot be made, because no wells were installed and thus no groundwater-level data were available in these basins.

Lee *et al.* (2002) claimed that the spatial distribution of the groundwater-level change during the Chi-Chi earthquake may be accounted for by a poroelastic model. The widespread occurrence of liquefaction, however, provides clear evidence that at least part of the sediment deformation was not elastic. Furthermore, for confined aquifers that have a typical thickness of 1–10 m, the required changes in

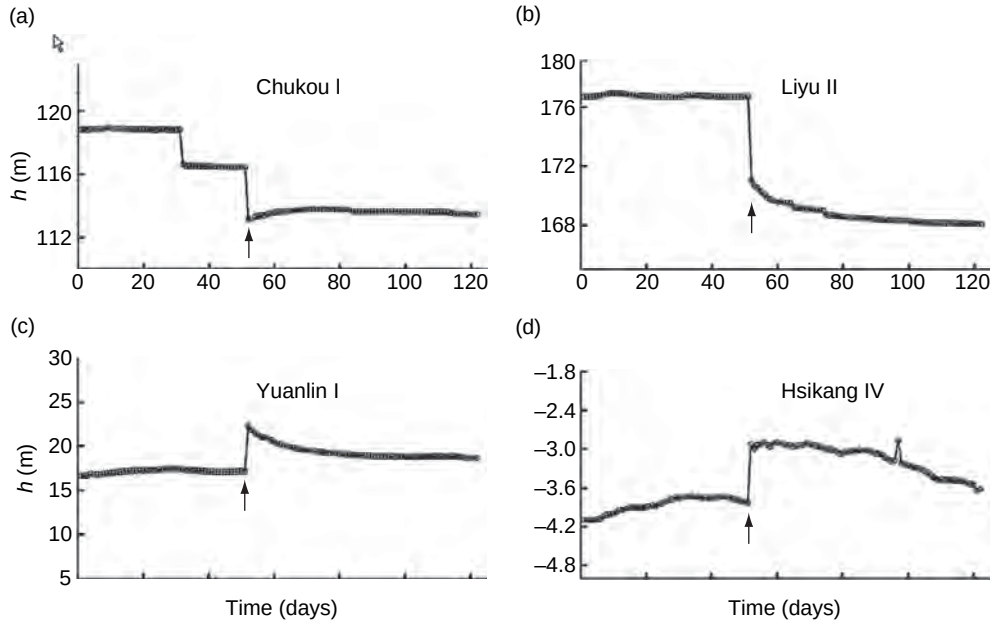


Figure 8 Four types of groundwater level responses during the Chi-Chi earthquake: (a) Groundwater level dropped during the earthquake and gradually rose afterward; (b) groundwater level dropped during the earthquake and continued to drop gradually afterward; (c) groundwater level rose during the earthquake and gradually declined afterwards; (d) groundwater level rose during the earthquake and remained at the new level for some time before slowly declining. The different groundwater-level responses may be due to differences in the proximity to the ruptured fault and different hydrogeological conditions as discussed in Wang *et al*. (2001). Figure created by Chung-Ho Wang.

porosity would have to be from 10^{-3} to 10^{-2} to account for the range of coseismic groundwater-level changes (Wang *et al*., 2001), beyond the magnitude of strains calculated from an elastic model (Lee *et al*., 2002). Finally, the sign of the changes in water level are in general the opposite of those expected from a poroelastic model (Koizumi *et al*., 2004).

As noted in Section 4.10.2.2, when loose sediments and soils are sheared at stresses between some upper and lower thresholds, grains tend to move together, leading to a decrease in porosity. If the deformation is undrained, as expected during seismic shaking, pore pressure would increase when the strain exceeds about 10^{-4}. This may explain the stepwise rise in the groundwater level in the fan (**Figures 8(c)** and **8(d)**). On the other hand, when consolidated sediments or sedimentary rocks are cyclically sheared beyond some critical threshold, microfracture and fracture may occur, leading to increased porosity and decreased pore pressure. This may explain the coseismic decline in the groundwater level in sedimentary rocks close to the ruptured fault (**Figures 8(a)** and **8(b)**).

It is possible to make an order-of-magnitude estimate of the change in porosity required to account for the coseismic water-level change. Because the coseismic water-level change does not involve fluid flow, the mass of pore water must be conserved, that is,

$$\frac{\mathrm{d}(\phi\rho)}{\mathrm{d}t} = 0 \qquad [9]$$

where ϕ again denotes the porosity. Equation (9) can be rewritten as

$$\frac{S}{b}\frac{\mathrm{d}h}{\mathrm{d}t} + \frac{\mathrm{d}\phi}{\mathrm{d}t} = 0 \qquad [10]$$

where S, as before, is the aquifer storativity, h the hydraulic head, and b the aquifer thickness. The second term in eqn [10] is the rate of change of sediment porosity at a given location resulting from ground shaking, which we assume is proportional to the intensity of ground shaking. Given that the intensity of ground shaking during the Chi-Chi earthquake decayed approximately exponentially with time (Ma *et al*., 1999), the rate of change in sediment porosity at a given location under earthquake shaking may be expressed as

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = Ae^{-\beta t} \qquad [11]$$

where β is of the order of 10^{-2}–10^{-1} s^{-1}, and A is a constant to be determined. Substituting eqn [11] into eqn [10] and integrating with respect to time, we get the coseismic change in the hydraulic head:

$$\Delta h(t) = \frac{bA}{S\beta}\left(e^{-\beta t} - 1\right) \qquad [12]$$

Pumping tests in the Choshui River fan yield $S \sim 10^{-3}$ (Lee and Wu, 1996). Given that ground shaking on the Choshui River fan during the Chi-Chi earthquake lasted between 20 and 100 s (Lee *et al.*, 2002a), for a 10 m groundwater-level change in a 1 m thick confined aquifer, $A = -10^{-4}$ to $-10^{-3}\,s^{-1}$, and for the same level of change in a 10 m aquifer, $A = -10^{-5}$ to $-10^{-4}\,s^{-1}$. Integrating eqn [10], we also obtain a coseismic change in porosity between 10^{-3} and 10^{-2}.

The uppermost aquifer on the Choshui River fan is either unconfined or poorly confined, resulting in less regular changes in pore pressure and groundwater level. Thus, qualitatively, the processes of consolidation and dilatation of sediments and sedimentary rocks, respectively, appear to provide a self-consistent interpretation of the spatial pattern of the coseismic groundwater-level changes on the Choshui River fan during the Chi-Chi earthquake (Wang *et al.*, 2001).

The availability of both well records and strong-motion records at the time of the Chi-Chi earthquake allows quantitative examination of the relationship between ground motion, groundwater-level change, and the distribution of liquefaction sites. Interpolating the strong-motion parameters at the well sites, Wang *et al.* (2003) showed that there is only a weak correlation between the horizontal component of the peak ground acceleration (PGA) and the occurrence of water-level changes and liquefaction (**Figure 9(a)**). Wong (2005) confirmed this finding by using a more extensive data set. This result is unexpected, because PGA is frequently used to predict the occurrence of liquefaction. Instead, Wong (2005) showed that there is a better correlation between the horizontal peak ground velocity (PGV) and the sites of liquefaction. Because PGV depends more on low-frequency components of ground motion than does PGA, this indicates that low-frequency ground motions are better correlated with elevated pore pressure and groundwater level than are high-frequency motions. This conclusion supports the finding that spectral accelerations and velocities at frequencies of about 1 Hz and lower were strongly correlated with the distribution of liquefaction sites (**Figure 9(b)**), whereas those above about 1 Hz were not (Wang *et al.*, 2003; Wong, 2005). A similar conclusion was reached by comparing the hydrologic changes with various measures of earthquake intensity computed from

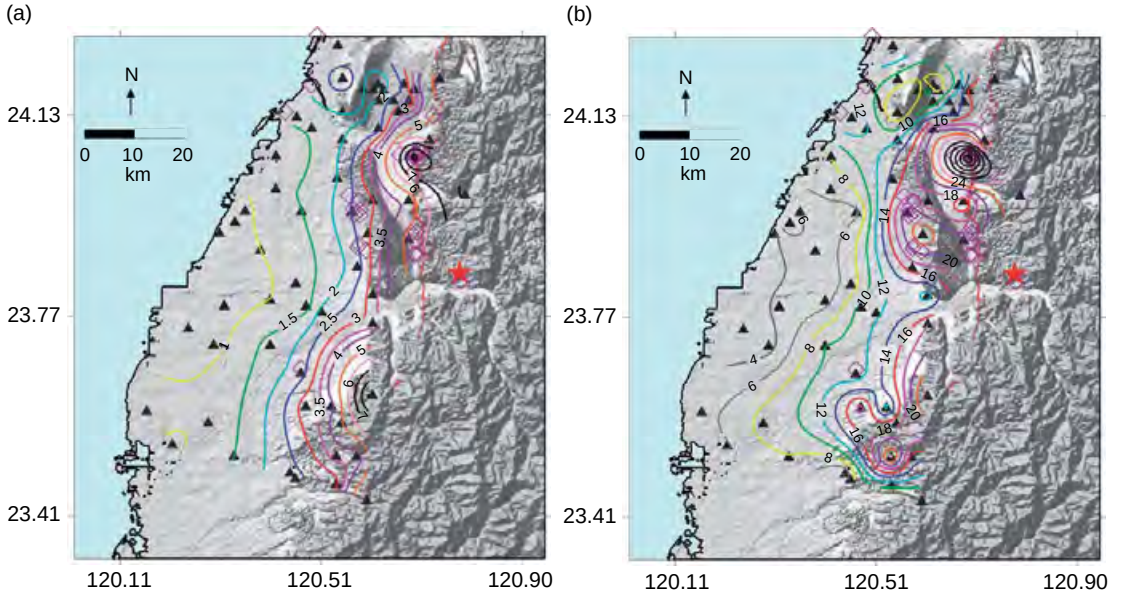


Figure 9 (a) Contours of the horizontal PGA during the Chi-Chi earthquake together with liquefaction sites in open diamonds. Note the weak correlation between PGA and liquefaction sites. (b) Contours of the spectral acceleration at 1 Hz. Note the fairly strong correlation between this and liquefaction sites.

seismograms passed through low-pass and high-pass filters (Wong, 2005). Although there is extensive evidence that low-frequency ground motions are better correlated with elevated pore pressure and groundwater level than high-frequency motions, it is unclear at this time whether the low-frequency ground motion was the cause for the coseismic groundwater-level change and liquefaction, or whether the hydrological changes, including liquefaction, caused the different spectral composition of the observed seismograms.

4.10.3.1.4 Far-field response

In the far-field (many fault lengths), the static stress due to the earthquake is nearly zero. Thus, sustained changes in water level at such distances (e.g., **Figure 4**) must be caused by the interactions between the aquifer and seismic waves. Such changes are not common, but confined to a small number of unusually sensitive sites usually located near active faults (King *et al.*, 2006). The mechanism for these interactions, however, is a matter of debate (Roeloffs, 1998; Matsumoto and Roeloffs, 2003; Brodsky *et al.*, 2003).

A permanent change in some aspect of pore structure is required, but the process by which small periodic strains change permeability or storage properties is unclear. Hydrogeochemical (e.g., Wang *et al.*, 2004d) and temperature (e.g., Mogi *et al.*, 1989) changes that accompany water-level changes confirm that pore-space connectivities change. One mechanism that has been proposed (to explain the particular postseismic change in water level shown in **Figure 4**) is that oscillatory flow back and forth in fractures caused by cyclic strain removes 'barriers' of fracture-blocking deposits, which then increases permeability and affects the final distribution of pore pressure (Brodsky *et al.*, 2003). Elkhoury *et al.* (2006) confirmed that distant earthquakes can indeed change permeability. Other proposed, but also unverified, mechanisms include pore pressure increases caused by a mechanism 'akin to liquefaction' (Roeloffs, 1998), shaking-induced dilatancy (Bower and Heaton, 1978), or increasing pore pressure through seismically induced growth of bubbles (e.g., Linde *et al.*, 1994). A feature of at least some far-field responses is that the magnitude of water-level changes scales with the amplitude of seismic waves (Sil and Freymueller, 2006).

One common feature of wells that exhibit far-field postseismic responses is that the sign of the water-level changes always appears to be the same (Matsumoto, 1992; Roeloffs, 1998; Sil and Freymueller, 2006). Any proposed mechanism should explain this latter feature, and the barrier explanation can only do so if the location of barriers is always located either upgradient or downgradient of the well. Far-field shaking-induced dilatancy (e.g., Bower and Heaton, 1978) or a process analogous to liquefaction (Roeloffs, 1998) would always cause the response to have the same sign, but none of them can be universal mechanisms because the sign of water-level changes varies from well to well.

Observations from wells that have responded to multiple earthquakes indicate that wells may have an enhanced sensitivity over a specific range of frequencies (Roeloffs, 1998). Interestingly, the threshold for triggered earthquakes also appears to have a frequency dependence (Brodsky and Prejean, 2005), though the frequency sensitivities for the two studies just cited are different. Observations of, and models that account for, the frequency-dependence response may be useful for distinguishing between processes.

4.10.3.1.5 Summary

To generalize, on the basis of the reported changes in water level in the near- and intermediate-field, it appears that deep wells in sound rock display a poroelastic response that mimicks coseismic strain (both in sign and magnitude), whereas shallow wells and wells in poorly consolidated materials tend to show larger changes and increases in water level.

4.10.3.2 Streamflow

One of the most spectacular surface hydrologic responses to earthquakes is large changes in streamflow. **Figure 10** shows two typical examples. The increased discharge is persistent, at least until rainfall obscures the changes caused by the earthquake. The peak discharge can occur from within a day to as much as several weeks after the earthquake. Increased streamflow occurs in the near- and intermediate-field (Manga, 2001). The total excess discharge (i.e., the total volume of water released in excess of that expected in the absence of the earthquake) can be large: 0.7 km^3 for the M 7.5 Chi-Chi earthquake (Wang *et al.*, 2004b), and 0.5 km^3 after the M 7.5 Hebgen Lake earthquake (Muir-Wood and King, 1993). Because streamflow responds to precipitation in the drainage basin, earthquake-induced changes are best recorded and studied during the dry season or during dry periods when there is little or no precipitation.

Explanations for changes in streamflow can be divided into five categories: expulsion of deep crustal

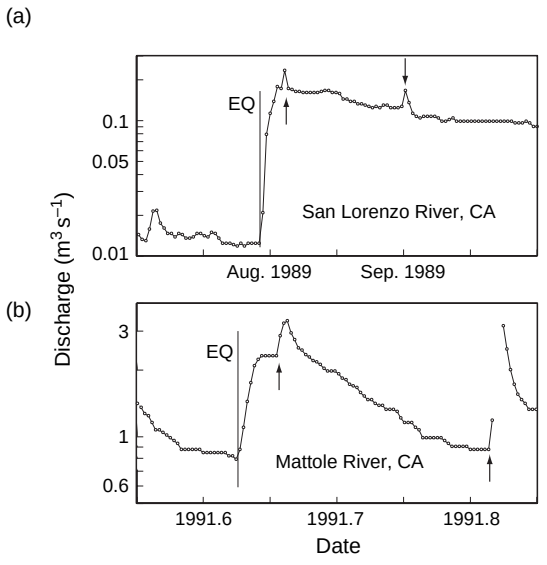


Figure 10 Hydrographs of (a) San Lorenzo River, CA, and (b) Mattole River, CA, showing postseismic response to the 1989 *M* 7.3 Loma Prieta earthquake and 1991 *M* 6 Honeydew earthquake, respectively. The vertical line labeled EQ indicates the time of earthquake. Small arrows show the time of rainfall events. Figure made with US Geological Survey stream gauge data.

fluids resulting from coseismic elastic strain (e.g., Muir-Wood and King, 1993), changes in near-surface permeability (Briggs, 1991; Rojstaczer and Wolf, 1992; Rojstaczer *et al.*, 1995; Tokunaga, 1999; Sato *et al.*, 2000), consolidation or even liquefaction of near-surface deposits (Manga, 2001; Manga *et al.*, 2003; Montgomery *et al.*, 2003), rupturing of subsurface reservoirs (Wang *et al.*, 2004c), and the release of water trapped in fault zones. The differences between these different explanations are nontrivial because of their implications for the magnitude of crustal permeability, its evolution, and the nature of groundwater flow paths. We thus summarize the basis and some problems with each explanation and then consider the streamflow response to the 1999 Chi-Chi earthquake as an illustrative example:

1. *Coseismic elastic strain.* Muir-Wood and King (1993) applied the coseismic elastic strain model, proposed by Wakita (1975) to explain coseismic groundwater-level changes, to explain the increased stream discharge after some large earthquakes, including the 1959 *M* 7.5 Hebgen Lake earthquake and the 1983 *M* 7.3 Borah Peak earthquake. Muir-Wood and King (1993) suggested that saturated microcracks in rocks opened and closed in response to the stress change during an earthquake, causing pore volume either to increase or decrease, resulting in a decrease or increase in the groundwater discharge into streams.

Following the 1989 Loma Prieta earthquake, however, chemical analysis of the stream water showed that the extra water that appeared in the streams following the earthquake had a shallow, rather than deep, origin (Rojstaczer and Wolf, 1992; Rojstaczer *et al.*, 1995). Furthermore, Rojstaczer and colleagues showed that to account for the extra water in the increased streamflow by coseismic elastic strain, a very large volume of the crust must be involved in the expulsion of groundwater, which in turn would require an unreasonably high permeability for the crust. In addition, one important constraint on models is that streamflow generally increases, and the few streams with decreased discharge show very small changes.

Manga *et al.* (2003) reported that streamflow increased at Sespe Creek, California, after several earthquakes, irrespective of whether the coseismic strain in the basin was contraction or dilatation. For this reason, the dynamic strain caused by the earthquake must be responsible for the observed changes in discharge.

2. *Enhanced permeability.* Briggs (1991) and Rojstaczer and Wolf (1992; also Rojstaczer *et al.*, 1995) proposed a model of enhanced permeability of the shallow crust resulting from seismically induced fractures and microfractures to explain the increased stream discharge in the nearby basins and the changes in the ionic concentration of stream water following the 1989 Loma Prieta earthquake in California. A shallow origin of the excess water is supported by a decrease in water temperature after discharge increased. Similar models were applied to the 1995 Kobe earthquake in Japan to explain the observed hydrological changes (Tokunaga, 1999; Sato *et al.*, 2000). Permeability enhancement was also invoked to explain the increased electrical conductivity of water discharged after an earthquake (Chamoille *et al.*, 2005). Elkhoury *et al.*, (2006) found that earthquakes in southern California caused phase shifts in the water-level response to tidal strain, and interpreted these to be due to increased permeability by seismic waves. The increase in permeability may be temporary because biogeochemical processes may act to reseal or block any newly created fresh fractures.

Manga (2001) found, however, that the rate of baseflow recession (i.e., the slow decrease in discharge during periods without precipitation) was unchanged by earthquakes that caused increases in streamflow. Assuming that the groundwater discharge Q to the

stream is governed by Darcy's equation [5], baseflow recession (a long time after recharge) is given by

$$\frac{dQ}{dt} \propto \exp[-aDt] \qquad [13]$$

where a is a constant that characterizes the geometry of the drainage basin, D is the hydraulic diffusivity, and t is time. Although this is a great simplification of the complex subsurface flow paths of water to streams, such models in general provide an excellent empirical fit to discharge records. The slope of the hydrograph in **Figure 11(a)** is proportional to aD, and **Figure 11(b)** shows that the recession constant does not change following the earthquake. Given that the basin geometry and flow pathways are unlikely to have been reorganized by the earthquake, an unchanged baseflow recession implies that aquifer permeability did not change after the earthquake, even over the time period of increased discharge. This conclusion was substantiated by later studies in other areas (Montgomery *et al.*, 2003) and during other earthquakes (Manga *et al.*, 2003).

3. *Coseimic consolidation and liquefaction.* Consolidation of loose materials is one way to increase pore pressures. Indeed, liquefaction occurred in many of the areas where streamflow increased (Montgomery and Manga, 2003). Consequently, Manga (2001; also Manga *et al.*, 2003) suggested that coseismic liquefaction of loose sediments on floodplains may provide the water for the increases in stream discharge following earthquakes.

Figure 12 shows the relationship between earthquake magnitude and distance between the epicenter and the center of the gauged basin for streams that responded to earthquakes. Also shown for reference is the maximum distance over which liquefaction has been reported (from **Figure 2**). That the limit for both responses is similar does not imply that liquefaction causes streamflow to increase. Instead, the correlation simply means that dynamic strains sufficient to induce liquefaction under optimal conditions are also sufficient for streamflow to change.

Wang *et al.*, (2004b), however, used the extensive network of stream gauges in Taiwan to identify the source of excess discharge in streams. Excess discharge did not originate on the unconsolidated

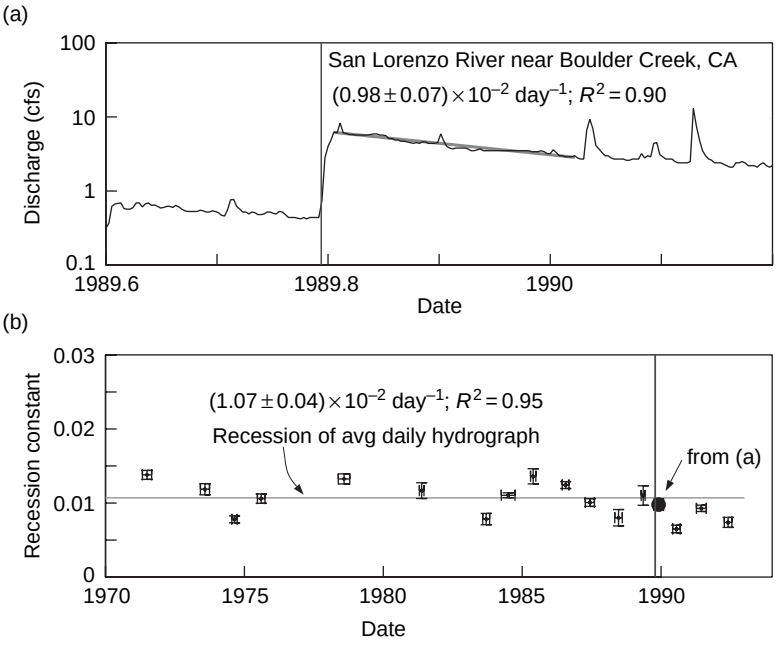


Figure 11 (a) Hydrograph of the San Lorenzo River, CA, showing postseismic response to the 1989 *M* 7.3 Loma Prieta earthquake. The vertical line labeled EQ indicates the time of the earthquake. The postseismic period of baseflow recession is shown by the bold sloping line. (b) The baseflow recession constant for periods of baseflow before and after the earthquake shows that even though discharge increased by an order of magnitude there was no significant change in baseflow recession. Figure made with US Geological Survey stream gauge data. Modified from Manga M (2001) Origin of postseismic streamflow changes inferred from baseflow recession and magnitude-distance relations. *Geophysical Research Letters* 28: 2133–2136.

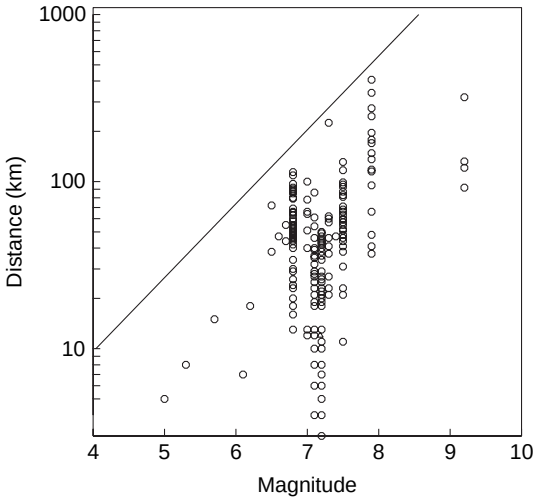


Figure 12 Relationship between earthquake magnitude and distance from the epicenter of streams that exhibited clear and persistent increases in discharge. The solid line is an empirical upper bound for the maximum distance over which shallow liquefaction has been observed (Wang *et al*., 2005a): $M = -5.0 + 2.26 \log R_{max}$, where R_{max} is in meters. Data shown based on the compilation in Montgomery and Manga (2003) with additional data from Wang *et al*. (2005a).

alluvial fan where there was, in fact, widespread liquefaction. Rather, the excess discharge originated in the mountains where there is little alluvial material and groundwater originates in bedrock. In addition, after the 2001 M 6.8 Nisqually, WA, earthquake, Montgomery *et al*., (2003) did not find a significant field association between liquefaction occurrence and streamflow changes.

4. *Ruptured subsurface reservoirs*. Following the M 6.5 San Simeon, CA, earthquake on 22 December 2003, a different type of streamflow increase was revealed in the central Coast Ranges, occurring within 15 min of the earthquake and lasting about an hour. At the same time, several new hot springs appeared across a dry riverbed, a short distance upstream of the stream gauge that registered the abrupt increase in streamflow. These observations suggest that the occurrence of streamflow increase and the appearance of hot springs were causally related and may be the result of earthquake-induced rupture of the seal of a hydrothermal reservoir. Recession analysis of the stream gauge data show $D/L^2 \sim 1000\ s^{-1}$, where L is the depth of the reservoir, implying a small L or large D or both. Assuming a one-dimensional model for flow between the ruptured seal and the stream, Wang *et al*., (2004c) estimated that the extra discharge after this earthquake was $\sim 10^3$ m^3. Abrupt changes in hot-spring discharge are known to occur in the Long Valley, California hydrothermal area (Sorey and Clark, 1981), suggesting that this type of hydrologic response may be characteristic of hydrothermal areas.

Rupturing permeability barriers may occur at depths of many kilometers. Husen and Kissling (2001) suggest that postseismic changes in the ratio of P-wave and S-wave velocities above the subducting Nazca Plate reflect fluid migration into the overlying plate following the rupture of permeability barriers.

5. *Fault valves*. In convergent tectonic regions, large volumes of pore water may be locked in subducted sediments (Townend, 1997). Sealing may be enacted partly by the presence of low-permeability mud and partly by the prevailing compressional stresses in such tectonic settings (Sibson and Rowland, 2003). Earthquakes may rupture the seals and allow pressurized pore water to erupt to the surface and recharge streams. This process may explain time variations in submarine fluid discharge at convergent margins (Carson and Screaton, 1998). Episodes of high discharge are correlated with seismic activity having features similar to tremor but are not correlated with large regional earthquakes (Brown *et al*., 2005).

To illustrate some approaches that are used to test these explanations, we again consider the records from the 1999 M 7.5 Chi-Chi earthquake in central Taiwan. In central Taiwan, 17 stream gauges are located on three streams systems (**Figure 13(a)**) that showed different responses to the 1999 Chi-Chi earthquake (Wang *et al*., 2004b). In addition, the local rainy season was over well before the earthquake. These records have made it possible to test the different hypotheses listed above. Among the three streams, two (Choshui Stream and the Wushi Stream) have extensive tributaries in the mountains (**Figure 13(a)**), but the third (Peikang Stream) originates on the sloping side of the Choshui fan, with no tributaries in the mountains. All the streams in the mountains showed large postseismic streamflow increases. On the alluvial fan, both the Choshui Stream and the Wushi Stream, which have tributaries in the mountains, showed large increases in streamflow. The amount of increase in streamflow in the proximal area of the Choshui alluvial fan was about the same as that in the distal area of the fan, indicating that most of the excess water originated in the mountains and that the contribution from the

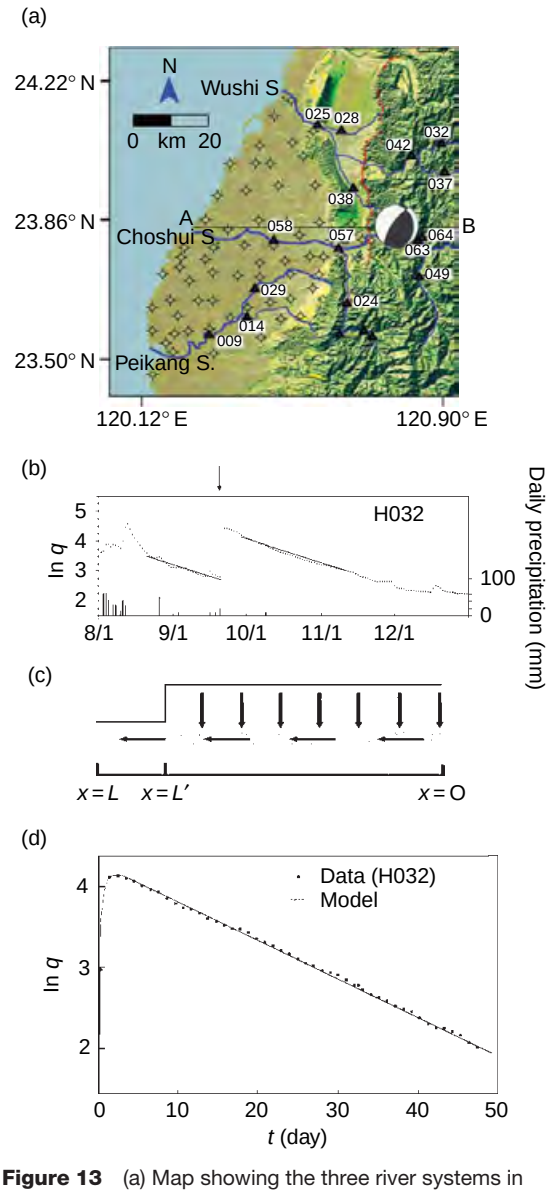


Figure 13 (a) Map showing the three river systems in central Taiwan and the locations of the stream gauges (in triangles). Note that both the Wushi River and the Choshui River have extensive tributaries in the mountains, but the Peikang River does not. Hydrological monitoring wells are marked by bull's-eyes. (b) Logarithm of discharge of a stream as recorded by a gauge in central Taiwan as a function of time before and after the Chi-Chi earthquake, together with precipitation record at a nearby station. Note that the recession curve before and after the earthquake (indicated by arrow) has the same slope. (c) Conceptual model showing increased recharge of aquifer by groundwater from mountains due to enhanced vertical permeability during and after the earthquake. (d) Comparison of excess discharge as a function of time with model result (thin curve). Excess discharge at the time of earthquake ($t = 0$) is taken to be zero and therefore does not appear on the diagram.

sediments in the fan was insignificant. By contrast, the Peikang Stream system, which does not have tributaries in the mountainous area, did not show any noticeable postseismic increase in streamflow. Thus it may be concluded that the excess water in the streams after the Chi-Chi earthquake requires an earthquake-induced discharge from the mountains. Any coseismic consolidation and liquefaction of sediments on the floodplain (alluvial fan) were not substantial enough to cause noticeable postseismic increase in streamflow.

Figure 13(b) shows the discharge of a stream in the mountains before and after the Chi-Chi earthquake, together with the precipitation record from a nearby station. In addition to a coseismic increase in streamflow and a slow postseismic return to normal, the figure shows that the baseflow recession rates both before and after the earthquake are identical. This observation is of interest because it suggests that the Chi-Chi earthquake did not cause significant changes in the permeability of the aquifer that fed the stream, as Manga (2001) first noticed in his study of the coseismic responses of streams in the United States. However, the excess water in streams in central Taiwan after the Chi-Chi earthquake requires extra discharge from the mountains that can appear only through enhanced permeability during the earthquake. This apparent contradiction may be resolved by recognizing that the earthquake-induced change in permeability may be anisotropic. The foothills of the Taiwan mountains consist of alternating beds of sandstone and shale. Thus, before the earthquake, the vertical permeability would be controlled by the impervious shales and groundwater flow would be mostly subhorizontal. During the earthquake, subvertical fractures breached the impervious shales and enhanced vertical permeability, allowing rapid downward draining of water to recharge underlying aquifers. The horizontal conductivity of the aquifers, however, was essentially unaffected. Field surveys after the Chi-Chi earthquake revealed numerous subvertical tensile cracks in the hanging wall of the thrust fault (Angelier *et al.*, 2000; Lee *et al.*, 2000, 2002b). Many wells in the foothills above the thrust fault showed a significant drop in water level (Lin, 2000; Yan, 2001; Chia *et al.*, 2001; Wang *et al.*, 2001), and a sudden downpour occurred in a tunnel beneath the foothills right after the earthquake. These reports are consistent with the model of enhanced vertical permeability during the Chi-Chi earthquake.

A one dimensional leaky aquifer model (**Figure 13(c)**) may be used to quantify this conceptual model. Flow is governed by eqn [7], and the recharge to the aquifer resulting from enhanced vertical permeability is treated as a source. Even though this model is highly simplified, several studies (e.g., Roeloffs, 1998; Manga, 2001; Manga *et al.*, 2003; Brodsky *et al.*, 2003) have demonstrated that it is useful for characterizing the first-order response of hydrological systems to earthquakes.

We treat the recharge of the leaky aquifer as coseismic with constant recharge for $0 < x \leq L'$ (length of aquifer beneath the mountain with enhanced permeability) and no recharge for $L' > x > L$ (length of aquifer beneath the floodplain). If we further assume a boundary condition $h = 0$ at $x = L$ and no-flow condition at $x = 0$, the excess discharge to the stream at time t is given by

$$Q_{ex} = \frac{2DVQ_0}{L^2(L'/L)} \sum_{n=1}^{\infty} (-1)^{n-1} \sin\left[\frac{(2r-1)\pi L'}{2L}\right] \times \exp\left[-\frac{(2r-1)^2\pi^2 D}{4L^2} t\right] \quad [14]$$

where L is the total length of the aquifer and V is the volume of the aquifer between $x = 0$ and $x = L'$ (**Figure 13(a)**); thus, VQ_0 is the total excess water recharging the aquifer. For large times, eqn [14] results in an exponential decrease in discharge, the so-called baseflow recession given by eqn [13]. **Figure 13(d)** compares the data for stream gauge 032 (adjusted to $Q_{ex} = 0$ just before earthquake) with the modeled excess discharge Q_{ex} determined by eqn [14]. An excellent fit is obtained with $D/L^2 = 2.4 \times 10^{-7}$ s^{-1}(from recession analysis) and $L'/L = 0.8$. The latter value is consistent with the fact that gauge 032 is located in the mountains, where the floodplain is small in comparison with the total aquifer length. The amount of excess water VQ_0 may be determined from the value of DVQ_0/L^2 from model fitting. By summing the excess discharges in the Choshui Stream and in the Wushi Stream, we obtain a total amount of 0.7–0.8 km^3 for the excess water released from the mountains after the Chi-Chi earthquake.

For the example shown in **Figure 13**, the peak postseismic discharge occurs within a day of the earthquake. **Figure 14** shows an example in which the peak discharge is reached 9–10 days later, though discharge begins to increase coseismically. This example is for Sespe Creek, CA, responding to the 1952 M 7.5 Kern County earthquake located 63 km away from the center of the drainage basin. Again, the model for excess discharge given by eqn [14], shown in the solid curve in **Figure 14**, fits the observed postseismic discharge very well (baseflow described by eqn [13] has been added back to the calculated excess discharge).

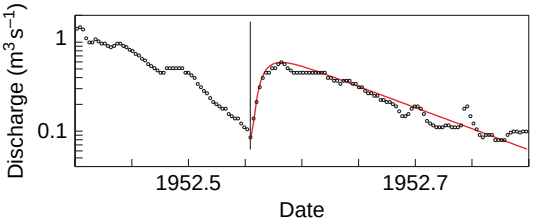


Figure 14 Response of Sespe Creek, CA to the 1952 M 7.5 Kern County earthquake. Daily discharge measurements collected and provided by the US Geological Survey are shown with circles. Curve is solution to eqn [14] for the excess flow with $L'/L = 0.4$ added to the baseflow, eqn [13], to recover the entire hydrograph. Vertical line shows the time of the earthquake. There was no precipitation during the time interval shown in this graph.

The different hypotheses listed in this section imply different crustal processes and different water–rock interactions during an earthquake cycle. In most instances, the hydrological models are underconstrained. A reasonable approach is to test the different hypotheses against cases (such as Chi-Chi) in which abundant and accurate data are available. We note that a single explanation need not apply for all cases of increased streamflow, so that identifying when and where different mechanisms dominate is important.

4.10.3.3 Mud Volcanoes

Mud volcanoes also appear to erupt in response to earthquakes (e.g., Panahi, 2005) and thus provide another probe of earthquake–hydrology interactions. Unfortunately though, the number and quality of reports prevent a rigorous statistical analysis of the correlation (Manga and Brodsky, 2006). Mud volcanoes range from small, centimeter-sized structures to large edifices up to few hundred meters high and several kilometers across. Mud volcanoes erupt predominantly water and fine sediment. They occur in regions where high sedimentation rates and fine-grained materials allow high pore pressures to develop (Kopf, 2002). Large mud volcanoes erupt from depths of more than several hundred meters.

A necessary condition to create mud volcanoes is the liquefaction or fluidization of erupted materials,

so that the erupted materials lose strength and can behave in a liquid-like manner (Pralle *et al.*, 2003). Unconsolidated sediment can be liquefied and fluidized through mineral dehydration, gas expansion, tectonic stresses, inflow of externally derived fluids, and even ocean waves (Maltman and Bolton, 2003).

Figure 15 shows that the occurrence of triggered mud volcanism falls within the liquefaction limit for shallow (upper few to tens of meters) liquefaction. Reports of mud volcanoes erupting within a day of large, distant earthquakes are compiled in Manga and Brodsky (2006) and are listed in the caption of **Figure 15.**

Liquefaction is usually thought to be a shallow phenomenon confined to the upper few to tens of meters. This is because the increase in fluid pressure needed to reach lithostatic pressure usually increases with depth. However, sedimentary basins with high sedimentation rates, fine sediments (with low permeability), and lateral compression often have high fluid pressures. Indeed, it is these settings in which mud volcanoes seem to occur (Milkov, 2000). Seismically triggered liquefaction may not necessarily be only a shallow phenomenon.

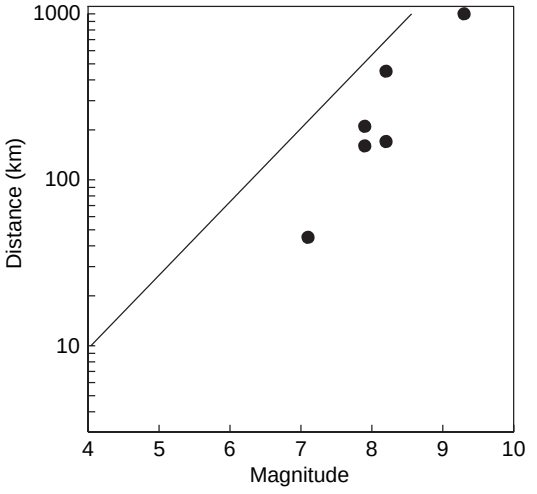


Figure 15 Relationship between earthquake magnitude and distance from the epicenter of large mud volcanoes that originate from depths greater than hundreds of meters and erupt within 2 days of the earthquake; shown are Andaman Islands (responding to M 9.2 Sumatra earthquake) and Niikappu Mud Volcano, Hokkaido, Japan (responding to events with M 7.1, 7.9, 7.9, 8.2, and 8.2, in 1982, 1968, 2003, 1952, and 1994, respectively). The solid line is an empirical upper bound for the maximum distance over which shallow liquefaction has been observed (Wang *et al.*, 2005): $M = -5.0 + 2.26 \log R_{max}$, where R_{max} is in meters Niikappu data provided by Nakamukae Makoto.

4.10.3.4 Geysers

Geysers change eruption frequency following distant earthquakes that generate coseismic static strains smaller than 10^{-7} or dynamic strains less than 10^{-6} (e.g., Hutchinson, 1985; Silver and Vallette-Silver, 1992). Geysers, despite requiring very special thermal conditions and hydrogeological properties to occur (Steinberg *et al.*, 1982a), thus provide another opportunity to understand how earthquakes can influence hydrogeological processes and properties.

Figure 16 shows the response of Daisy geyser in Yellowstone National Park to the 2002 M 7.9 Denali earthquake located more than 3000 km away (Husen *et al.*, 2004). The eruption interval decreases by about a factor of 2 and then slowly increases to the pre-earthquake frequency over a period of a few months. Among the many geysers at Yellowstone, the eruption frequency of some increased, whereas for others it decreased.

Although geysers often respond to earthquakes, their response to barometric pressure changes (e.g., White, 1967), solid Earth tides (e.g., Rinehart, 1972), and hypothetical preseismic strains (Silver and Vallette-Silver, 1992) is weaker. The apparent

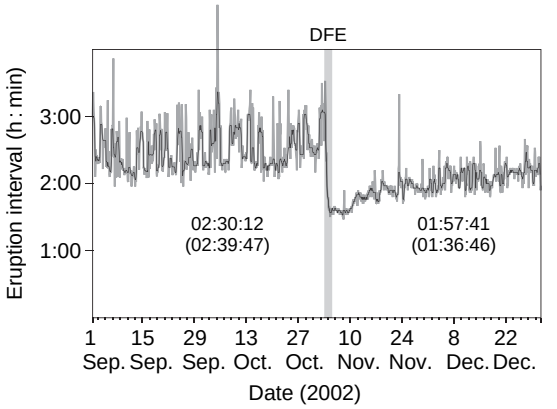


Figure 16 Eruption interval (time between two successive eruptions) in h : min format as a function of times Daisy geyser, Yellowstone National Park. Thin grey line shows raw data, that is, the actual measured eruption intervals. The black line is smoothed data obtained by averaging over several measurements. Median eruption intervals are also shown in h : min : s format. Median eruption intervals are computed over a several-week period; intervals in parentheses are based on a few days. Time of the 2002 M 7.9 Denali earthquake is shown be the vertical line. Figure provided by Stephan Husen and based on data presented in Husen S, Taylor R, Smith RB, and Healser (2004) Changes in geyser eruption behavior and remotely triggered seismicity in Yellowstone. National park, produced by the 2002 M 7.9 Denali fault earthquake, Alaska. *Geology* 32: 537–540.

correlation of eruption interval and nonseismic strains is, in fact, controversial. Rojstaczer *et al.* (2003) reanalyzed the eruption records at Yellowstone and found that geysers were insensitive to earth tides and diurnal barometric pressure changes, and thus concluded that geysers were not sensitive to strains less than about 10^{-8}–10^{-7}. One of these records was from the Daisy geyser (Figure 16). Given that Daisy geyser responded to earthquakes that caused equally small static strain, we can reasonably conclude that it is the dynamic strain, rather than static strain, that causes the observed responses (e.g., Husen *et al.*, 2004).

The mechanism by which the eruption interval changes is not known. Steinberg *et al.* (1982b) note that one mechanism for initiating a geyser eruption is superheating the water. Using a lab model of a geyser, Steinberg *et al.* (1982c) show that vibrations can reduce the degree of superheating and can thus increase the eruption frequency. It is worth noting that while earthquakes can decrease the interval between eruptions, sometimes the frequency of eruptions decreases. Ingebritsen and Rojstaczer (1993) propose instead that much of the observed temporal variability in eruption frequency can be explained by changes in matrix permeability, which in turn governs the recharge of the conduit. Geyser frequency is also sensitive to the (often complex) conduit geometry (Hutchinson *et al.*, 1997) and hence permeability of the geyser conduit. However, because the conduit is already much more permeable, small strains within the conduit itself are unlikely to have a significant effect on geyser eruptions (Ingebritsen *et al.*, 2006). As a consequence, the observed temporal variability in eruption frequency is probably caused by changes in the permeability of the matrix surrounding the main geyser conduit–it is through this matrix that the conduit is recharged between eruptions.

The high sensitivity and long-lasting response of geysers to small seismic strains requires reopening, unblocking, or creating fractures to induce large enough permeability changes to influence eruption frequency. Geysers thus provide evidence that dynamic strains are able to create permanent changes in permeability from small dynamic strains and at great distances from the earthquake.

In summary, the response of geysers to distant earthquakes is most easily explained by changes in permeability, and the sensitivity of geysers to earthquakes compared with earth tides and changes in barometric pressure indicates that dynamic strains cause the response.

4.10.4 Feedback between Earthquakes and Hydrology

The presence of water in the subsurface, and changes in the amount of water on the surface or within the subsurface, can influence the occurrence of earthquakes, as summarized in **Figure 17**. There are three basic ways water can influence seismicity.

First, changes in loading of the surface can increase deviatoric stresses. The most clear examples follow the impoundment of water by reservoirs (e.g., Simpson *et al.*, 1988; Gupta, 1992). Natural examples are more difficult to identify, probably in part because the magnitude of the surface load is much smaller; proposed examples include snow loading (Heki, 2003) and loading from ocean tides (Cochran *et al.*, 2004), or other seasonal processes (Wolf *et al.*, 1997).

Second, changes in fluid pressure p reduce the effective stress

$$\sigma_{ij}^{\text{eff}} = \sigma_{ij} - \alpha p \delta_{ij} \qquad [15]$$

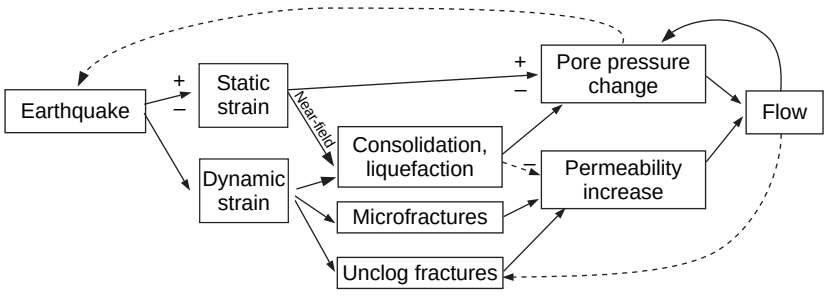


Figure 17 Schematic illustration of the relationship between earthquakes and hydrologic responses; + and–signs indicate the sign of the response. Figure based on an illustration from David Mays.

where α is again the Biot–Willis coefficient. The classic example of seismicity induced by changes in pore pressure is caused by fluid injection (e.g., Raleigh *et al.*, 1976). Some cases of reservoir-induced seismicity, in particular those in which the seismicity lags changes in water level, can be attributed to changes in pore pressure p (e.g., Talwani and Acree, 1985; do Nascimento *et al.*, 2005). Changes in pore pressure also diffuse, governed by eqn [6], so the telltale signature of earthquakes triggered by changes in pore pressure would be a migration of the locus of seismicity at a rate consistent with the (usually unknown) hydraulic diffusivity. Spatial migration of reservoir-induced seismicity, seismicity induced by fluid injection (Tadokoro *et al.*, 2000; Parotidis and Shapiro, 2004), and some aftershock sequences (e.g., Nur and Booker, 1972; Bosl and Nur, 2002; Hainzl, 2004; Miller *et al.*, 2004) have been attributed to pore pressure diffusion. Pore pressure changes caused by natural loading at the surface (e.g., groundwater recharge) have also been invoked to trigger seismicity (e.g., Costain *et al.*, 1987; Saar and Manga, 2003; Christiansen *et al.*, 2005).

Third, fluid extraction, rather than fluid injection, can change stresses through poroelastic deformation (Segall, 1989). Earthquakes appear to have been induced by fluid extraction in gas fields (Segall *et al.*, 1994) and oil fields (Gomberg and Wolf, 1999; Zoback and Zinke, 2002). A decrease in pore pressure might be expected to stabilize faults by increasing the effective stress. Poroelastic deformation, however, caused changes in fluid pressure, increases the magnitude of deviatoric stresses away from the region of fluid extraction.

One general conclusion from all these studies of hydrologically triggered seismicity, regardless of the mechanism, is that very small stress changes, typically 0.01–1 Mpa, appear to trigger the earthquakes. Such small stress changes, however, are also similar to the stress changes thought to trigger earthquakes through nonhydrological means (e.g., Harris, 1998; Stein, 1999).

If earthquakes cause changes in hydrogeological properties and changes in fluid pressure promote seismicity, then hydrogeological and seismological processes are coupled. Triggered seismicity. (*see* Chapter 4.09) is one possible example of this coupling. Interaction is promoted if the state of stress is close to failure so that the small changes in stress associated with either natural hydrologic processes or hydrological responses to earthquakes can in turn influence seismicity. At least in tectonically active areas, many faults do appear to be critically stressed (Zoback *et al.*, 1987), consistent with earthquakes being triggered by small stress changes.

Rojstaczer and Ingebritsen (2005, personal communication) suggest that one manifestation of this interaction is the value of the mean large-scale permeability of the crust – it should be of a size to accommodate internal (fluid generation by metamorphic and magmatic processes) and external (groundwater recharge and discharge) forcing. Diagenesis, in general, tends to seal fractures and fill pores, thus decreasing permeability. If groundwater recharge does not change with time, water levels and pore pressures will increase, promoting seismicity, which in turn generates new fractures and increases permeability, A mean crustal permeability of $10^{-14}\,m^2$ (Manning and Ingebritsen, 1999) is consistent with the present mean rate of groundwater recharge and hydraulic head gradients driving basin-wide groundwater flow. Rojstaczer and Ingebritsen propose that this balance is reached by a feedback between hydrological processes and seismicity. A similar balance must occur in the lower crust, though the source of water is from metamorphic reactions rather than of meterological orgin. The mean permeability obtained from such a balance, however, is a temporal and spatial average. Instead, evidence of short-lived and locally high permeabilities are preserved in the spatial distribution of mineral deposits in ancient fault systems (e.g., Micklethwaite and Cox, 2004) and transient, localized high temperatures in the lower crust (Camacho *et al.*, 2005).

4.10.5 Hydrologic Precursors

As noted in Section 4.10.2.1, failure of brittle rocks under deviatoric stress is usually preceded by pervasive microcrack formation. In the subsurface, this process (i.e., microfracturing) would greatly increase the surface area of the affected rock in contact with groundwater and thus allow the release of gases and dissolved ions from the rock into the groundwater, changing its chemical composition. Coalescing of microcracks into larger fractures may connect hydraulically isolated aquifers, causing both changes in the hydraulic heads and mixing of groundwaters with initially distinct hydrogeochemistry. These processes may further cause changes in the electrical conductivity of the rocks, and thus of the crust. Scenarios of this kind have led to the not

unreasonable expectation that hydrological, hydrogeochemical, and related geophysical precursors may appear before the occurrence of large earthquakes.

During the 1960s and 1970s, several groups reported precursory changes in the crustal seismic velocity ratio (V_p/V_s) before some earthquakes in the then Soviet Union (Savarensky, 1968; Semenov, 1969), in New York (Aggarwal *et al.*, 1973), and in California (Whitcomb *et al.*, 1973). Coupling the failure processes of brittle rocks, as noted previously, to a groundwater flow model, Nur (1974) proposed a 'dilatancy model' to explain the sequence of precursory changes in (V_p/V_s). Scholz *et al.*, (1973) further develop the dilatancy model as a common physical basis for precursory phenomena. Later seismic experiments designed to detect changes in both V_p and V_s, however, failed to reveal any precursory changes before many major and moderate earthquakes (e.g., McEvilly and Johnson, 1974). In 1988, in response to Bakun and Lindh's (1985) prediction that a large earthquake would occur near Parkfield before 1993, Park installed a network of electrodes across the fault zone near Parkfield to detect any precursory changes in the electrical resistivity of the crust (Park, 1997) – one of the major predictions of the dilatancy model (Scholz *et al.*, 1973). The predicted Parkfield earthquake did not occur; but an M 6.0 earthquake occurred on 28, September 2004 near Parkfield, providing an excellent opportunity to test the telluric method for detecting any precursory changes in crustal resistivity before a large earthquake. Careful data processing was applied to the time series of the dipole fields before the earthquake, but no precursory changes were found (Park, 2005). These failures, among others, have cast questions about the general validity of the dilatancy model, and about earthquake precursors in general.

An intense search for hydrological and hydrochemical precursors before earthquakes has also been pursued. For example, a few days prior to the 1946 M 8.3 Nankaido earthquake in Japan, water levels in some wells reportedly fell by more than 1 m and some wells went dry (Sato, 1978; Linde and Sacks, 2002). Three days before the M 6.1 Kettleman Hill, CA, earthquake, Roeloffs and Quilty (1997) found a gradual, anomalous rise in water level of 3 cm. This observation was included in the IASPEI Preliminary List of Significant Precursors (Wyss and Booth, 1997). Following the 1995 M 7.2 Kobe earthquake, several papers reported precursory changes in the concentrations of radon, chlorine, and sulfate ions in groundwater (e.g., Tsunogai and Wakita, 1995; Igarashi *et al.*, 1995) and in groundwater level (King *et al.*, 1995). Changes in radon concentration are the most commonly reported and discussed hydrogeochemical precursor (e.g., Wakita *et al.*, 1988; Trique *et al.*, 1999) because the release of radon is especially sensitive to crustal strains.

Definitive and consistent evidence for hydrological and hydrochemical precursors, however, has remained elusive (Bakun *et al.*, 2005). Difficulties include that most reported changes were not corrected for the fluctuations in temperature, barometric pressure, earth tides, and other environmental factors, so that some changes taken to be earthquake related may in fact be 'noise' (e.g., Hartman and Levy, 2005), that changes were recorded at some sites but often were not recorded at other nearby sites (e.g., Biagi *et al.*, 2001), and that instrument failures and personnel/program changes often do not allow persistent and consistent monitoring over long periods of time (King *et al.*, 2000) – a necessary condition for obtaining reliable precursory data. Distinguishing a precursor from a response to a previous earthquake, for example, the 1946 Nankaido earthquake was preceded by the 1944 M 8.2 Tonankai event, creates additional ambiguity. Notwithstanding these difficulties, progress has been made in the past decade. For example, intensive and continued observations of various kinds of precursory hydrological and hydrochemical changes have been made in Japan during the past half century (Wakita, 1996), and records are corrected to remove the noise introduced by fluctuations in temperature, barometric pressure, earth tides, and other factors (Igarashi and Wakita, 1995); both high- and low-pass filtering has been applied to the time series of raw hydrochemical data in Kamchatka, Russia, to remove long- and short-period changes unrelated to earthquake processes (Kingsley *et al.*, 2001); statistical, rather than deterministic, procedures have been introduced (Maeda and Yoshida, 1990) to assess the conditional probability of future seismic events; multicomponent, hydrochemistry analysis was applied to groundwater samples in Iceland before and after a major earthquake to enhance the possibility of detecting possible precursors (Claesson *et al.*, 2004); and, finally, relationships among various types of hydrological and hydrochemical precursory signals were sought to improve future earthquake prediction strategies (Hartman and Levy, 2005). Thus, although we may still be far from achieving a genuine understanding of the underlying mechanisms of the various earthquake-related anomalies, significant efforts are underway.

Static strains of 10^{-8} can reasonably be expected to produce changes in water level of about 1 cm for optimal hydrogeologic properties. Even greater sensitivity to strain (though perhaps only dynamic strains) is implied by some of the very distant hydrologic responses seen at geysers and some very distant changes in water in wells (e.g., Brodsky *et al.*, 2003). The strong sensitivity of hydrologic processes and properties to small strains is the primary basis for hope that hydrological and hydrogeochemical monitoring may detect any hypothetical pre-earthquake strains. Recognizing precursors, and distinguishing between responses from previous earthquakes and precursors to new earthquakes, is a different matter and may continue to be problematic (Hartman and Levy, 2005). Although it is unclear whether any future documented hydrological precursors could actually be used for earthquake prediction, they may at least provide new insight into the physics of earthquakes.

4.10.6 Concluding Remarks

Hydrologic responses to earthquakes provide constraints on hydrogeologic processes in regions that might otherwise be inaccessible, for example, fault zones or the deep subsurface. Measured responses may also provide information at spatial and temporal scales that are difficult to study with more conventional hydrogeological measurements such as well tests. However, these novel features of hydrologic responses also mean that it can be difficult, and perhaps even impossible in some cases, to obtain the information needed to distinguish between competing models for hydrologic responses.

Explaining the hydrologic responses to earthquakes should in principle be simple because they reflect the strain caused by earthquakes. The great variety of hydrologic responses, however, highlights the complexity of deformation and structure of geological materials and the interaction between processes. Over the last decade there has thus been a trend toward developing quantitative physically based models to explain hydrogeological responses to earthquakes with an emphasis on explaining phenomena that cannot be explained by linear poroelastic models alone.

We conclude by noting that advances in physically based models may not be sufficient for understanding the interactions between earthquakes and water. Although expensive and time consuming, continued monitoring of wells, springs, and streams, ideally at high sampling rates and with complementary data sets (e.g., chemistry, temperature, pressure), provides the data needed to test models and may also lead to the discovery of new hydrological phenomena.

Acknowledgments

We thank Alex Wong for sharing the results of his study of the Chi-Chi earthquake, Fu-Qiong Huang for the information in **Figure 3** and three reviewers for comments. This study is supported by the Sloan Foundation.

References

Aggarwal YP, Sykes LR, Armbruster J, and Sbar ML (1973) Premonitory changes in seismic velocities and prediction of earthquakes. *Nature* 241: 101–104.

Akita F and Matsumoto N (2004) Hydrological responses induced by the Tokachi-oki earthquake in 2003 at hot spring well in Hokkaido, Japan. *Geophysical Research Letters* 31, doi:10.1029/2004GL020433.

Ambraseys NN (1988) Engineering seismology. *Earthquake Engineering and Structural Dynamics* 17: 1–105.

Andrew DJ (2003) A fault constitutive relation accounting for thermal pressurization of pore fluid. *Journal of Geophysical Research* 107, doi:10.1029/2002JB001942.

Angelier J, Chu HT, Lee JC, *et al.* (2000) Geologic knowledge and seismic risk mitigation: Insight from the Chi–Chi earthquake: Taiwan. In: Lo C-H and Liao W-I (eds.) *Proceeding of International Workshop Annual Commemoration of Chi-Chi Earthquake, Science Aspect,*, pp. 13–24. Taipei, Taiwan: National Center for Research on Earthquake Engineering.

Arias A (1970) A measure of earthquake intensity. In: Hansen RJ (ed.) *Seismic Design for Nuclear Power Plant*, pp. 438–483. Cambridge, MA: Massachusetts Institute of Technology.

Bakun W and Lindh A (1985) The Parkfield earthquake prediction experiment. *Science* 229: 619–624.

Bakun WH, Aagaard B, Dost B, *et al.* (2005) Implications for prediction and hazard assessment from the 2004 Parkfield earthquake. *Nature* 437: 969–974.

Bardet JP (2003) Advances in analysis of soil liquefaction during earthquakes. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology, Part B*, pp. 1175–1201. San Diego: Academic press.

Berryman JG and Wang HF (2000) Elastic wave propagation and attenuation in a double-porosity dual-permeability medium. *International Journal of Rock Mechanics and Mining Sciences* 37: 67–78.

Biagi PF, Ermini A, Kingsley SP, Khatkevich YM, and Gordeev EI (2001) Difficulties with interpreting changes in groundwater gas content as earthquake precursors in Kamchatka, Russia. *Journal of Seismology* 5: 487–497.

Biot MA (1941) General theory of three-dimensional consolidation. *Journal of Applied Physics* 12: 155–164.

Biot MA (1956a) Theory of propagation of elastic waves in fluid-saturated porous solid. I: Low-frequency range. *Journal of the Acoustical society of America* 28: 168–178.

Biot MA (1956b) Theory of propagation of elastic waves in fluid-saturated porous solid. I: Higher frequency range. *Journal of the Acoustical Society of America* 28: 179–191.

Biot MA (1962) Mechanics of deformation and acoustic propagation in porous media. *Journal of Applied Physics* 33: 1482–1498.

Blanchard FB and Byerly P (1935) A study of a well gauge as a seismograph. *Bulletin of the Seismological Society of America* 25: 313–321.

Bosl WJ and Nur A (2002) Aftershocks and pore fluid diffusion following the 1992 Landers earthquake. *Journal of Geophysical Research* 107: 2366.

Bower DR and Heaton KC (1978) Response of an aquifer near Ottawa to tidal forcing and the Alaskan earthquake of 1964. *Canadian Journal of Earth Sciences* 15: 331–340.

Brace WF, Paulding J, and Scholz C (1966) Dilatancy in the fracture of crystalline rocks. *Journal of the Geophysical Research* 71: 3939–3953.

Briggs RO (1991) Effects of Loma Prieta earthquake on surface water in Waddell Valley. *Water Resources Bulletin* 27: 991–999.

Brodsky EE and Kanamori H (2001) Elastohydrodynamic lubrication of faults. *Journal of Geophysical Research* 106: 16357–16374.

Brodsky EE and Prejean SG (2005) New constraints on mechanisms of remotely triggered seismicity at Long Valley Caldera. *Journal of Geophysical Research* 110, dio:10.1029/2004JB003211.

Brodsky EE, Roeloffs E, Woodcock D, Gall I, and Manga M (2003) A mechanism for sustained ground water pressure changes induced by distant earthquakes. *Journal of Geophysical Research* 108, doi:10.1029/2002JB002321.

Brown KM, Tryon MD, DeShon HR, Schwartz S, and Dorman LM (2005) Correlated transient fluid pulsing and seismic tremor in the Costa Rica subduction zone. *Earth and Planetary Science Letters* 238: 189–203.

Camacho A, Lee JKW, Hensen BJ, and Braun J (2005) Short-lived orogenic cycles and the eclogitization of cold crust by spasmodic hot fluids. *Nature* 435: 1191–1196.

Carson B and Screaton EJ (1998) Fluid flow in accretionary prisms: Evidence for focused, time-variable discharge. *Reviews of Geophysics* 36: 329–352.

Charmoille A, Fabbri O, Mudry J, Guglielmi Y, and Bertrand C (2005) Post-seismic permeability change in a shallow fractured aquifer following a M-L 5.1 earthquake (Fourbanne karst aquifer, Jura outermost thrust unit, eastern France). *Geophysical Research Letters* 32: L18406 (doi:10.1029/2005GL023859).

Chester FM, Chester JS, Kirschner DL, Schulz SE, and Evans JP (2004) Structure of large-displacement strike-slip fault zones in the brittle continental crust. In: Karner GD, Taylor B, Driscoll NW, and Kohlstedt DL (eds.) *Rheology and Deformation in the Lithosphere at Continental Margins*, pp. 223–260. New York: Columbia University Press.

Chester FM, Evens JP, and Biegel RL (1993) Internal structure and weakening mechanisms of the San Andreas fault. *Journal of Geophysicall Research* 98: 771–786.

Chia YP, Wang YS, Wu HP, Chiu JJ, and Liu CW (2001) Changes of groundwater level due to the 1999 Chi-Chi earthquake in the Choshui River fan in Taiwan. *Bulletin of the Seismological Society of America* 91: 1062–1068.

Christiansen LB, Hurwitz S, Saar MO, Ingebritsen SE, and Hsieh PA (2005) Seasonal seismicity at western United States volcanic centers. *Earth and Planetary Science Letters* 240: 307–321.

Claesson L, Skelton A, Graham C, *et al.* (2004) Hydrogeochemical changes before and after a major earthquake. *Geology* 32: 641–644.

Cochran ES, Vidale JE, and Tanaka S (2004) Earth tides can trigger shallow thrust fault earthquakes. *Science* 306: 1164–1166.

Cooper HH, Bredehoeft JD, Papadopulos IS, and Bennett RR (1965) The response of well-aquifer systems to seismic waves. *Journal of Geophysical Research* 70: 3915–3926.

Costain JK, Bollinger GA, and Speer JA (1987) Hydroseismicity: A hypothesis for the role of water in the generation of intro-plate seismicity. *Seismological Research Letters* 58: 41–64.

do Nascimento AF, Lunn RJ, and Cowie PA (2005) Numerical modelling of pore-pressure diffusion in a reservoir-induced seismicity site in northeast Brazil. *Geophysical Journal International* 160: 249–262.

Dobry R (1989) Some basic aspects of soil liquefaction during earthquakes. *Annals of the New York Academy of Sciences* 558: 172–182.

Dvorkin J, Nolen-Hoeksema R, and Nur A (1994) The squirt-flow mechanism: Macroscopic description. *Geophysics* 59: 428–438.

Elkhoury JE, Brodsky EE, and Agnew DC (2006) Seismic waves increase permeability. *Nature* 441: 1135–1138.

Forster CB, Evens JP, Tanaka H, Jeffreys R, and Noraha T (2003) Hydrologic properties and structure of the Mozumi Fault, central Japan. *Journal of Geophysical Research* 30, doi:10.1029/2002GL014904.

Galli P (2000) New empirical relationships between magnitude and distance for liquefaction. *Tectonophysics* 324: 169–187.

Gavrilenko P, MeliKadze G, Chelidze T, Gibert D, and Kumsiashvili (2000) Permanent water level drop associated with the Spitak earthquake: Observations at Lisi borehole (Republic of Georgia) and modeling. *Geophysical Journal International* 143: 83–98.

Gomberg J and Wolf L (1999) A possible cause for an improbable earthquake: The 1997 Mw 4.9 southern Alabama earthquake and hydrocarbon recovery. *Geology* 27: 367–370.

Gorokhovich Y (2005) Abandonment of Minoan palaces on Crete in relation to the earthquake induced changes in groundwater supply. *Journal of Archeological Science* 32: 217–222.

Gudmundsson A (2000) Active fault zones and groundwater flow. *Geophysical Research Letters* 27: 2993–2996.

Gupta HK (1992) *Reservoir-Induced Earthquakes*, 264 pp. New York: Elsevier.

Hainzl S (2004) Seismicity patterns of earthquake swarms due to fluid intrusion and stress triggering. *Geophysical Journal International* 159, doi: 10.1111/j.1365-246X.2004.02463.x.

Hamiel Y, Lyakhovsky V, and Agnon A (2005) Rock dilation, nonlinear deformation, and pore pressure change under shear. *Earth and Planetary Science Letters* 237: 577–589.

Harris RA (1998) Introduction to the special section: Stress triggers, stress shadows, and the implication for seismic hazard. *Journal of Geophysical Research* 103: 24247–24258.

Hartman J and Levy JK (2005) Hydrogeological and gasgeochemical earthquake precursors–a review for application. *Natural Hazards* 34: 279–304.

Heki K (2003) Snow load and seasonal variation of earthquake occurrence in Japan. *Earth and Planetary Science Letters* 207: 159–164.

Hickman SH, Zoback MD, and Ellsworth WL (2005) Structure and composition of the San Andreas fault zone at Parkfield: Intial results from SAFOD, phase 1 and 2. *EOS Transaction, American Geophysical Union* 83: 237.

Hsu SK, Chin CV, Cheng LH, and Lin WS 1999 The change of groundwater level of earthquake 921 in central Taiwan. In: *Proceedings of Conference on Sustainable Development and Geosciences Prospect*, *National Taiwan University*, *Department of Geololgy*, pp. 7–35 (in Chinese). Taipei, Taiwan: Department of Geology, National Taiwan University.

Husen S and Kissling E (2001) Postseismic fluid flow after the large subduction earthquake of Antofagasta, Chile. *Geology* 29: 847–850.

Husen S, Taylor R, Smith RB, and Healser H (2004) Changes in geyser eruption behavior and remotely triggered seismicity in Yellowstone. National park produced by the 2002 M 7.9 Denali fault earthquake, Alaska. *Geology* 32: 537–540.

Hutchinson RA (1985) Hydrothermal changes in the upper geyser basin, Yellowstone National Park, after the 1983 Borah Peak, Idaho, earthquake. *US Geological Survey Open File Report 85-290,* 612–624.

Hutchinson RA, Westphal JA, and Kieffer SW (1997) *In situ* observations of Old Faithful Geyser. *Geology* 25: 875–878.

Igarashi G, Saeki S, Takahata N, *et al.* (1995) Ground-water radon anomaly before the Kobe earthquake in Japan. *Science* 269: 60–61.

Igarashi G and Wakita H (1991) Tidal responses and earthquake-related changes in the water level of deep wells. *Journal of Geophysical Research* 96: 4269–4278.

Igarashi G and Wakita H (1995) Geochemical and hydrological observations for earthquake prediction in Japan. *Journal of Physics of the Earth* 43: 585–598.

Ingebritsen SE and Rojstaczer SA (1993) Controls on geyser periodicity. *Science* 262: 889–892.

Ingebritsen SE, Sanford WE, and Neuzil CE (2006) *Groundwater in Geologic Processes,* 2nd edn. New York: Cambridge University Press.

Ishihara K (1996) *Soil Behavior in Earthquake Geotechnics*, 350p. Oxford: Clarendon Press.

Jonsson S, Segall P, Pedersen R, and Bjornsson G (2003) Post-earthquake ground movements correlated to pore-pressure transients. *Nature* 424: 179–183.

King C-Y, Azuma S, Igarashi G, Ohno M, Saito H, and Wakita H (1999) Earthquake-related water-level changes at 16 closely clustered wells in Tono, central Japan. *Journal of Geophysical Research* 104: 13073–13082.

King C-Y, Azuma S, Ohno M, *et al.* (2000) In search of earthquake precursors in the water-level data of 16 closely clustered wells at Tono, Japan. *Geophysical Journal International* 143: 469–477.

King C-Y, Koizumi N, and Kitagawa Y (1995) Hydrogeochemical anomalies and the 1995 Kobe earthquake. *Science* 269: 38–39.

King CY, Zhang W, and Zhang ZC (2006) Earthquake-induced groundwater and gas changes. *Pure and Applied Geophysics* 163: 633–645.

Kingsley SP, Biagi PF, Piccolo R, *et al.* (2001) Hydrogeochemical precursors of strong earthquakes: A realistic possibility in Kamachatka. *Physics and Chemistry of the Earth (C)* 26: 769–774.

Kitagawa Y, Koizumi N, Takahashi M, Matsumoto N, and Sato T (2006) Changes in groundwater levels or pressures associated with the 2004 earthquake off the west coast of northern Sumatra (M9.0). *Earth, Planets and Space* 58: 173–179.

Koizumi N, Lai W-C, Kitagawa Y, and Matsumoto N (2004) Comment on ''Coseismic hydrological changes associated with dislocation of the September 21, 1999 Chichi earthquake, Taiwan'' by Lee M *et al*. *Geophysical Research Letters* 31, doi:10:1029/2004GL019897.

Kono Y and Yanagidani T (2006) Broadband hydroseismograms observed by closed borehole wells in the Makioka mine, central Japan: Response of pore pressure to seismic waves from 0.05 to 2 Hz. *Journal of Geophysical Research* 111: B03410 (doi:10.1029/2005JB003656).

Kopf AJ (2002) Significance of mud volcanism. *Reviews of Geophysics* 40, doi:10.1029/2000RG000093.

Kuribayashi E and Tatsuoka F (1975) Brief review of liquefaction during earthquakes in Japan. *Soil and Foundations* 15: 81–92.

Lachenbruch A (1980) Frictional heating, fluid pressure, and the resistance to fault motion. *Journal of Geophysical Research* 85: 6097–6112.

Lachenbruch AH and Sass JH (1977) Heat flow in the United States and the thermal regime of the crust. In: Heacock JH (ed.) *The Earth's Crust, Geophys. Monogr, Ser., 20*, pp. 625–675. Washington, DC: American Geophysical Union.

Lawson AC (1908) The California Earthquake of April 18, 1906. Report of the state earthquake investigation commission, vol. 1, Carnegie Institution of Washington, Washington, DC.

Lay T and Wallace TC (1995) *Modern Global Seismology*. New York: Academic Press.

Lee CS and Wu CL (1996) Pumping tests of the Choshuichi alluvial fan. In: *Conference on Groundwater and Hydrology of the Choshui River Alluvial Fan*, *Taiwan*, *Water Resources Bureau*, pp.165–179 (in Chinese). Taipei, Taiwan: Water Resources Bureau, MOEA.

Lee CT, Kelson KI, and Kang KH (2000) Hanging wall deformation and its effect on buildings and structures as learned from the Chelungpu faulting in the 1999 Chi-Chi Taiwan earthquake. In: Lo C-H and Liao W-I (eds.) *Proceedings of International Workshop Annual Commemoration of Chi-Chi Earthquake, Science Aspect*, pp. 93–104. Taipei, Taiwan: National Center for Research on Earthquake Engineering.

Lee J-C, Chu H-T, Angelier J *et al.* (2002a) Geometry and structure of northern rupture surface ruptures of the 1999 $M_w = 7.6$ Chi-Chi Taiwan earthquake: Influence from inherited fold belt structures. *Journal of Structural Geology* 24: 173–192.

Lee M, Liu T-K, Ma K-F, and Chang Y-M (2002b) *Coseismic hydrological changes associated with dislocation of the September 21, 1999 Chichi earthquake, Taiwan.* 29, doi:10.1029/2002GL015116.

Li H and Zhu L (2006) High-resolution structures of the Landers fault zone inferred from aftershock waveform data. *Geophysical Journal International* (in press).

Li Y-G, Vidale JE, Aki K, and Xu F (2000) Depth-dependent structure of the Lander Fault zonefrom trapped waves generated by aftershocks. *Journal of Geophysical Research* 105: 6237–6254.

Li YG, Vidale JE, and Malin PE (2006) Parkfield fault-zone guided waves: High-resolution delineation of the low-velocity damage zone on the San Andreas at depth near SAFOD site. *The Proceeding of ICDP-IODP Fault-Zone Drilling*, pp. 1–4. Miyazaki, Japan.

Lin WY (2000) Unstable groundwater supply after the big earthquake. *United Daily Newspaper*, Taiwan (in Chinese).

Linde AT and Sacks IS (2002) Slow earthquakes and great earthquakes along the Nankai trough. *Earth and Planetary Science Letters* 203: 265–75.

Linde AT, Sacks IS, Johnston MJS, Hill DP, and Bilham RG (1994) Increased pressure from rising bubles as a mechanism for remotely triggered seismicity. *Nature* 371: 408–410.

Liu E, Vlatsos S, Li S-Y, Main IG, and Schoenberg M (2004) Modeling seismic wave propagation during fluid injection in a fractured network: Effects of pore fluid pressure on time-lapse seismic signatures. *Leading Edge* 23: 778–783.

Liu LB, Roeloffs E, and Zheng XY (1989) Seismically induced water level oscillations in the Wali well, Beijing, China. *Journal of Geophysical Research* 94: 9453–9462.

Lockner DA and Beeler NM (2002) Rock failure and earthquakes. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake and*

Engineering Seismology, pp. 505–537. San Diego: Academic Press.

Ma KF, Lee CT, and Tsai YB (1999) The Chi-Chi, Taiwan earthquake: Large surface displacement on an inland thrust fault. *EOS Transaction, American Geophysical Union* 80: 605–611.

Maeda K and Yoshida A (1990) A probabilistic estimation of earthquake occurrence on the basis of the appearance times of multiple precursory phenomena. *Journal of Physics of the Earth* 38: 431–444.

Maltman AJ and Bolton A (2003) How sediments become mobilized. In: Van Rensberger P, Hillis RR, Maltman AJ, and Morley CK (eds.) *Subsurface Sediment Mobilization,* Geological Society of London Special Publications 216, pp. 9–20. London: Geological Society of London.

Manga M (2001) Origin of postseismic streamflow changes inferred from baseflow recession and magnitude-distance relations. *Geophysical Research Letters* 28: 2133–2136.

Manga M and Brodsky EE (2006) Seismic triggering of eruptions in the far field: Volcanoes and geysers. *Annual Review of Earth and Planetary Sciences* 34: 263–291.

Manga M, Brodsky EE, and Boone M (2003) Response of streamflow to multiple earthquakes and implications for the origin of postseismic discharge changes. *Geophysical Research Letters* 30, doi:10.1029/2002GL016618.

Manning CE and Ingebritsen SE (1999) Permeability of the continental crust: Constraints from heat flow models and metamorphic systems. *Reviews of Geophysics* 37, 127–150.

Mase CW and Smith L (1987) Effect of frictional heating on the thermal, hydrological, and mechanical response of a fault. *Journal of Geophysical Research* 92: 6249–6272.

Masson Y and Pride SR (2006) Computation of seismic attenuation and dispersion due to microscopic heterogeneity in porous materials. *Journal of Geophysical Research* (in press).

Matsumoto N (1992) Regression analysis for anomalous changes of ground water level due to earthquakes. *Geophysical Research Letters* 19: 1192–1196.

Matsumoto N and Roeloffs EA (2003) Hydrological response to earthquakes in the Haibara well, central Japan. II: Possible mechanism inferred from time-varying hydraulic properties. *Geophysical Journal International* 155: 899–913.

Mavko G, Mukerji T, and Dvorkin J (1998) *The Rock Physics Handbook: Tools for Seismic Data Analysis in Porous Media*. New York: Cambridge University Press.

McEvilly T and Johnson L (1974) Stability of P and S velocities from central California quarry blasts. *Bulletin of the Seismological Society of America* 64: 343–353.

Micklethwaite S and Cox SF (2004) Fault-segment rupture, aftershock-zone fluid flow, and mineralization. *Geology* 32: 813–816.

Milkov AV (2000) Worldwide distribution of submarine mud volcanoes and associated gas hydrates. *Marine Geology* 167: 29–42.

Miller SA, Collettini C, Chiaraluce L, Cocco M, Barchi M, and Kaus BJP (2004) Aftershocks driven by a high-pressure CO_2 source at depth. *Nature* 427: 724–727.

Mizuno T (2003) The subsurface observation of fault-zone trapped waves: Applications to investigations of the deep structure of active faults. *Bulletin of the Earthquake Research Institute* 78: 91–106.

Mogi K, Mochizuki H, and Kurokawa Y (1989) Temperature changes in an artesian spring at Usami in the Izu Peninsula (Jaoan) and their relation to earthquakes. *Tectonophysics* 159: 95–108.

Montgomery DR, Greenberg HM, and Smith DT (2003) Streamflow response to the Nisqually earthquake. *Earth and Planetary Science Letters* 209: 19–28.

Montgomery DR and Manga M (2003) Streamflow and water well responses to earthquakes. *Science* 300: 2047–2049.

Mount VS and Suppe J (1987) State of stress near the San Andreas Fault – Implications for wrench tectonics. *Geology* 15: 1143–1146.

Muir-Wood R and King GCP (1993) Hydrological signatures of earthquake strain. *Journal of Geophysical Research* 98: 22035–22068.

National Research Council (1985) *Liquefaction of Soils during Earthquakes*, 204 pp. Washington, DC: National Academy Press.

Neuzil CE (2003) Hydromechanical coupling in geologic processes. *Hydrogeological Jounal* 11: 41–83.

Nur A (1974) Matsushiro, Japan, earthquake swarm: Confirmation of the dilatancy-fluid diffusion model. *Geology* 2: 217–221.

Nur A and Booker JR (1972) Aftershocks caused by pore fluid flow? *Science* 175: 885–887.

Ohtani T, Tanaka H, Fujimoto K, Higuchi T, Tomida N, and Ito H (2001) Internal structure of the Nojima fault zone from Hirabayshi GSJ drill core. *Island Arc* 10: 392–400.

Panahi BM (2005) Mud volcanism, geodynamics and seismicity of Azerbaijan and the Caspian sea region. In: Martinelli G and Panahi B (eds.) *Mud Volcanoes, Geodynamics and Seismicity*, pp. 89–104. Dordrecht: Springer.

Papadopoulos GA and Lefkopulos G (1993) Magnitude-distance relations for liquefaction in soil from earthquakes. *Bulletin of the Seismological Society of America* 83: 925–938.

Park SK (1997) Monitoring resistivity changes in Parkfield, California: 1988–1995. *Journal of Geophysical Research* 102: 24545–24559.

Park SK (2005) The agony and ecstasy of earthquake prediction (talk presented at US Geological Survey, Menlo Park, CA).

Parotidis M and Shapiro SA (2004) A statistical model for the seismicity rate of fluid-injection-induced earthquakes. *Geophysical Research Letters* 31, doi:10.1029/2004GL020421.

Pralle N, Külzer M, and Gudehus G (2003) Experimental evidence on the role of gas in sediment liquefaction and mud volcanism. In: Van Rensberger P, Hillis RR, Maltman AJ, and Morley CK (eds.) *Subsurface Sediment Mobilization* pp. 159–172. Geological Society of London Special Publications 216.

Pride SR, Berryman JG, and Harris JM (2004) Seismic attenuation due to wave-induced flow. *Journal of Geophysical Research* 109: B01201, (doi:10.1029/2003JB002639).

Quilty EG and Roeloffs EA (1997) Water level changes in response to the December 20, 1994 M4.7 earthquake near Parkfield, California. *Bulletin of the Seismological Society of America* 87: 1018–1040.

Raleigh CB, Healy JH, and Bredehoeft (1976) An experiment in earthquake control at Rangely, Colorado. *Science* 191: 1230–1237.

Rice JR (1992) Fault stress state, pore pressure distributions, and the weakness of the san Andreas fault. In: Evans B and Wong T-F (eds.) *Fault Mechanics and Transport Properties of Rocks*, pp. 475–504. San Diego: Academic Press.

Rice JR and Cleary MP (1976) Some basic stress diffusion solutions for fluid saturated elastic porous media with compressible constituents. *Reviews of Geophysics and Space Physics* 14: 227–241.

Rinehart JS (1972) Fluctuations in geyser activity caused by variations in earth tidal forces, barometric pressure, and tectonic stresses. *Journal of Geophysical Research* 77: 342–350.

Roeloffs EA (1996) Poroelastic techniques in the study of earthquake-related hydrologic phenomena. *Advances in Geophysics* 37: 135–195.

Roeloffs EA (1998) Persistent water level changes in a well near Parkfield, California, due to local and distant earthquakes. *Journal of Geophysical Research* 103: 869–889.

Roeloffs EA (2006) Evidence for aseismic deformation-rate changes prior to earthquakes. *Annual Review of Earth and Planetary Sciences* 34: 591–627.

Roeloffs EA and Quilty E (1997) Water level and strain changes preceding and following the August 4, 1985 Kettleman Hills, California, earthquake. *Pure and Applied Geophysics* 149: 21–60.

Roeloffs EA, Sneed M, Galloway DL, *et al.* (2003) Water-level changes induced by local and distant earthquakes at Long Valley caldera, California. *Journal of Volcanology and Geothermal Research* 127: 269–303.

Rojstaczer S and Ingebritsen SE 2005 Crustal permeability constrained by internal and external forcing (in review).

Rojstaczer S and Wolf S (1992) Permeability changes associated with large earthquakes: An example from Loma Prieta, California. *Geology* 20: 211–214.

Rojstaczer S, Wolf S, and Michel R (1995) Permeability enhancement in the shallow crust as a cause of earthquake-induced hydrological changes. *Nature* 373: 237–239.

Rojstaczer S, Galloway DL, Ingebritsen SE, and Rubin DM (2003) Variability in geyser eruptive timing and its causes: Yellowstone National Park. *Geophysical Research Letters* 30: (art No. 1953), doi: 10.1029/2003GL017853.

Rudnicki JW, Yin J, and Roeloffs EA (1993) Analysis of water level changes induced by fault creep at Parkfield, California. *Journal of Geophysical Research* 98: 8143–8152.

Saar MO and Manga M (2003) Seismicity induced by seasonal groundwater recharge at Mt. Hood, Oregon. *Earth and Planetary Science Letters* 214: 605–618.

Sato H (1978) Precursory land tilt prior to the Tonankai earthquake of 1944. In some precursors prior to recent great earthquakes along the Nankai trough. *Journal of Physics of the Earth* 25: s11521.

Sato T, Sakai R, Furuya K, and Kodama T (2000) Coseismic spring flow changes associated with the 1995 Kobe earthquake. *Geophysical Research Letters* 27: 1219–1222.

Sato T, Matsumoto N, Kitagawa Y, *et al.* (2004) Changes in water level associated with the 2003 Tokachi-oki earthquake. *Earth, Planets and Space* 56: 395–400.

Savarensky EF (1968) On the prediction of earthquakes. *Tectonophysics* 6: 17–27.

Scholz CH, Sykes LR, and Aggarwal YP (1973) Earthquake prediction – Physical basis. *Science* 181: 803–810.

Seed H and Lee KL (1966) Liquefaction of saturated sands during cyclic loading. *Journal of the Soil Mechanics and Foundation Division* 92: 105–134.

Segall P (1989) Earthquakes triggered by fluid extraction. *Geology* 17: 942–946.

Segall P, Grasso JR, and Mossop A (1994) Poroelastic stressing and induced seismicity near the Lacq gas field, southwestern France. *Journal of Geophysical Research* 99: 15423–15438.

Semenov AM (1969) Variations in the travel-time of transverse and longitudinal waves before violent earthquakes. *Physics of the Solid Earth* 3: 245–248.

Sibson RH (1973) Interactions between temperature and pore-fluid pressure during earthquake faulting and a mechanism for partial or total stress relief. *Nature* 243: 66–68.

Sibson RH and Rowland JV (2003) Stress, fluid pressure and structural permeability in seismogenic crust, North Island, New Zealand. *Geophysical Journal International* 154: 584–594.

Sil S and Freymueller JT (2006) Well water level changes in Fairbanks, Alaska, due to the great Sumatra-Andaman earthquake. *Earth Planets Space* 58: 181–184.

Silver PG and Valette-Silver NJ (1992) Detection of hydrothermal precursors to large northern California earthquakes. *Science* 257: 1363–1368.

Simpson DW, Leith WS, and Scholz CH (1988) Two types of reservoir-induced seismicity. *Bulletin of the seismological Society of America* 78: 2025–2050.

Sleep NH and Blanpied ML (1992) Creep, compaction and the weak rheology of major faults. *Nature* 359: 687–692.

Sorey ML and Clark MD (1981) Changes in the discharge characteristics of thermal springs and fumaroles in the Long Valley caldera, California, resulting from earthquakes on May 25–27, 1980. *US Geological Survey Open-File Report* 81–203.

Stein RS (1999) The role of stress transfer in earthquake occurence. *Nature* 402: 605–609.

Steinberg GS, Merzhanov AG, and Steinberg AS (1982a) Geyser process: Its theory, modeling and field experiment. Part 1: Theory of the geyser process. *Modern Geology* 8: 67–70.

Steinberg GS, Merzhanov AG, and Steinberg AS (1982b) Geyser process: Its theory, modeling and field experiment. Part 3: On metastability of water in geysers. *Modem Geology* 8: 75–78.

Steinberg GS, Merzhanov AG, and Steinberg AS (1982c) Geyser process: Its theory, modeling and field experiment. Part 4: On seismic influence on geyser regime. *Modern Geology* 8: 79–86.

Stidham C, Antolik M, Dreger D, Larsen S, and Romanowicz B (1999) Three-dimensional structure influences on the strong-motion wavefield of the 1989 Loma Prieta earthquake. *Bulletin of the Seismological Society of America* 89: 1184–1202.

Su T-C, Chiang K-W, Lin S-J, Wang F-G, and Duann S-W (2000) Field reconnaissance and preliminary assessment of liquefaction in Yuan-Lin area. *Sino-Geotechnics* 77: 29–38.

Tadokoro K, Ando M, and Nishigami K (2000) Induced earthquakes accompanying the water injection experiment at the Nojima fault zone, Japan: Seismicity and its migration. *Journal of Geophynical Research* 105: 6089–6104.

Talwani P and Acree S (1985) Pore pressure diffusion and the mechanism of reservoir-induced seismicity. *Pure and Applied Geophysics* 122: 947–965.

Terzaghi K, Peck RB, and Mesri G (1996) *Soil Mechanics in Engineering Practice,* 3rd edn. New York: Wiley.

Thurber C, Roecker S, Ellsworth W, Chen Y, Lutter W, and Sessions R (1997) Two-dimensional seismic image of the San Andreas fault in the Northern Gablian range, Central California: Evidence for fluids in the fault zone. *Geophysical Research Letters* 24: 1591–1594.

Tokunaga T (1999) Modeling of earthquake-induced hydrological changes and possible permeability enhancement due to 17 January 1995 Kobe earthquake, Japan. *Journal of Hydrology* 223: 221–229.

Townend J (1997) Subducting a sponge; minimum estimates of the fluid budget of the Hikurangi Margin accretionary prism. *Geological Society of New Zealand Newsletter* 112: 14–16.

Trique M, Richon P, Perrier F, Avouac JP, and Sabroux JC (1999) Radon emanation and electric potential variations associated with transient deformation near reservoir lakes. *Nature* 399: 137–141.

Tsunogai U and Wakita H (1995) Precursory chemical changes in ground water: Kobe earthquake, Japan. *Science* 269: 61–62.

Unsworth M, Bedrosian P, Eisel M, Egbert G, and Siripunvaraporn W (2000) Along strike variations in the electrical structure of the San Andreas Fault at Parkfield, California. *Geophysical Research Letters* 27: 3021–3024.

Wakita H (1975) Water wells as possible indicators of tectonic strain. *Science* 189: 553–555.

Wakita H (1996) Geochemical challenge to earthquake prediction. *Proceeding of the National Academic Science* 93: 3781–3786.

Wakita H, Nakamura Y, and Sano Y (1988) Short-term and intermediate-term geochemical precursors. *Pure and Applied Geophysics* 126: 78–89.

Waller RM (1966) Effects of the March 1964 Alaska earthquake on the hydrology of south-central Alaska. *USGS Professional Paper 544A,* pp 28.

Wang C-H, Wang CY, Kuo C-H, and Chen W-F (2005c) Some isotopic and hydrological changes associated with the 1999 Chi-Chi earthquake, Taiwan. *Island Arc* 14: 37–54.

Wank C-Y (1984) On the constitution of the San Andreas fault. *Geophysical Research Letters* 89: 5858–5866.

Wang C-Y, Manga M, and Wong A (2005a) Floods on Mars released from groundwater by impact. *Icarus* 175: 551–555.

Wang C-Y, Cheng L-H, Chin C-V, and Yu S-B (2001) Coseismic hydrologic response of an alluvial fan to the 1999 Chi-Chi earthquake, Taiwan. *Geology* 29: 831–834.

Wang C-Y, Dreger D, Manga D, and Wong A (2004c) Streamflow increase due to rupturing of hydrothermal reservoirs: Evidence from the 2003 San Simeon, California, earthquake. *Geophysical Research Letters* 31: L10502 (doi: 10.1029/2004GL020124).

Wang C-Y, Dreger DS, Wang C-H, Mayeri D, and Berryman JG (2003) Field relations among coseismic ground motion, water level change and liquefaction for the 1999 Chi-Chi (M_w=7.5) earthquake, Taiwan. *Geophysical Research Letters* 17, doi:10.1029/2003GL017601.

Wang C-Y, Rui F, Yao Z, and Shi X (1986) Gravity anomaly and density structure of the San Andreas fault zone. *Pure and Applied Geophysics* 124: 127–140.

Wang C-Y, Wong A, Dreger D, and Manga M (2006) Liquefaction limit during earthquakes and underground explosions – implications on ground-motion attenuation. *Bulletin of the seismological Society of America* 96: 355–363.

Wang C-Y, Wang C-H, and Kuo C-Y (2004a) Temporal change in groundwater level following the 1999 (M_w = 7.5) Chi-Chi earthquake, Taiwan. *Geofluids* 4: 210–220.

Wang C-Y, Wang C-H, and Manga M (2004b) Coseismic release of water from mountains: Evidence from the 1999 (M_w = 7.5) Chi-Chi, Taiwan, earthquake. *Geology* 32: 769–772.

Wang R, Woith H, Milkereit C, and Zschau J (2004d) Modelling of hydrogeochemical anomalies induced by distant earthquakes. *Geophysical Journal International* 157: 717–726.

Wang HF (2000) *Theory of Linear Poroelasticity*, 287 pp. Princeton, NJ: Princeton University Press.

Whitcomb JH, Garmany JD, and Anderson DL (1973) Earthquake prediction: Variation of seismic velocities before the San Fernando earthquake. *Science* 180: 632–635.

White DE (1967) Some principles of geyser activity, mainly from Steamboat Springs, Nevada. *American Journal of Science* 265: 644–684.

Wibberley C and Shimamoto T (2003) Internal structure and permeability of major strike-slip fault zones: The Median Tectonic Line in W. Mie Prefecture, S.W. Japan. *Journal of Structural Geology* 25: 59–78.

Wibberley CAJ and Shimamoto T (2005) Earthquake slip weakening and asperities explained by thermal pressurization. *Nature* 436: 689–692.

Wolf LW, Rowe CA, and Horner RB (1997) Periodic seismicity near Mt. Ogden on the Alaska–British Columbia border: A case for hydrologically-triggered earthquakes? *Bulletin of the seismological Society of America* 87: 1473–1483.

Wong A (2005) *Relations between the Spectral Composition of Ground Motion and Hydrological Effects*, *and a Robustness Test of a Finite-Fault Model in the 1999 Chi-Chi (Taiwan) Earthquake*. Masters Thesis, University of California, Berkeley.

Wyss M and Booth DC (1997) The IASPEI procedure for the evaluation of earthquake precursors. *Geophysical Journal International* 131: 423–424.

Yan HR (2001) Water problems in the Yuenlin Mountains: Changes in groundwater after the Chi-Chi earthquake, Taiwan. *Chinese Daily* (in Chinese).

Youd TL (2003) Liquefaction mechanisms and induced ground failure. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology, Part B*, pp. 1159–1173. San Diego: Academic Press.

Zoback MD and Zinke J (2002) Production-induced normal faulting in the Valhall and Ekofisk oil fields. *Pure and Applied Geophysics* 159: 403–420.

Zoback MD, Zoback ML, Mount VS, *et al.* (1987) New evidence on the state of stress of the San Andreas fault system. *Science* 238: 1105–1111.

4.11 Deep Earthquakes

H. Houston, University of Washington, Seattle, WA, USA

4.11.1 Introduction

Earthquakes occur below Earth's crust to nearly 700 km depth in planar regions of seismicity termed Wadati–Benioff zones, which delineate the cold core of subducting slabs. The rates of occurrence of earthquakes deeper than 70 km, however, are several times lower than those of crustal quakes. This chapter reviews the properties and possible physical mechanisms of deep earthquakes, generally defined here as earthquakes that occur below about 70 km depth. At times the discussion distinguishes between deep and intermediate earthquakes (setting the boundary between them at about 350 km), while other times the specific depth ranges are indicated.

The physical processes that permit the occurrence of deep earthquakes are not well understood. The brittle–frictional processes that operate to produce shallow seismic rupture, fracture, and frictional sliding appear to be suppressed by conditions at depth. In particular, increasing pressure with depth tends to inhibit fracture and sliding, while increasing

temperature promotes ductile flow (e.g., Griggs and Handin, 1960; Scholz, 2002). Classic experiments on olivine show that the shear stresses necessary to overcome the high normal stresses imposed by the overburden are far greater than can be sustained at the temperatures at depth (e.g., Brace and Kohlstedt, 1980). Thus, failure must occur at lower shear stresses by one or more other mechanisms.

In addition to posing a long-standing puzzle to earthquake seismologists, deep earthquakes pose significant seismic hazards in several regions around the globe. The hazard stems primarily from relatively shallow intermediate-depth earthquakes beneath populated continental regions, because event frequency falls off exponentially with depth below about 30 km. Regions that have suffered damage from such subcrustal earthquakes include Romania (4 March 1977 M 7.5, depth = 90 km), Chile (25 January 1939 M 7.8, depth = 80 km), Washington State in the USA (28 February 2001 M 6.8, depth = 54 km), and Japan (26 May 2003 M 7.0, depth = 61 km). See Frohlich (2006a) for a discussion of the seismology and severe destruction associated with the 1977 Romania and 1939 Chile intermediate-depth earthquakes, including death tolls of about 1500 and 28 000, respectively.

Finally, deep earthquakes have played a key role in illuminating the structure of Earth's mantle and core. This is largely because their waveforms are less complicated than those of shallow earthquakes due to the separation in time of the first-arriving body waves from later-arriving body waves reflected off Earth's surface. In addition, direct body waves from deep earthquakes need pass through the highly layered and laterally heterogeneous crust of the Earth beneath only the receiver, rather than beneath both the receiver and source.

Deep earthquakes remain mysterious. This overview does not attempt an exhaustive survey of the topic, which impinges on many aspects of geophysics including plate tectonics, mineral physics, mantle convection, slab and mantle structure, composition, and rheology, as well as seismology. Rather, the aim is to highlight some of the more recent work on the subject and raise several issues, some perhaps overlooked. Previous reviews that the reader may find helpful include Frohlich (1989), Green and Houston (1995), Kirby *et al.* (1996), Green and Marone (2002), and Frohlich (2006a), an entire book devoted to deep earthquakes.

This chapter discusses seismological properties of deep earthquakes, followed by the leading candidates for physical mechanisms proposed to generate deep earthquakes, the geophysical setting (temperatures and stresses) of deep earthquakes, seismic structures in slabs, and finally focuses on implications for seismogenesis and key outstanding issues.

4.11.2 Deep Earthquake Source Properties

Section 4.11.1 presents salient properties of deep earthquakes. Some of these have been known for decades, while others have come into focus only recently. Particular attention is drawn to changes in source properties with depth, magnitude, or subduction zone, with an eye to whether these may provide constraints on possible deep earthquake generation mechanisms.

The bimodal nature of the distribution of earthquakes with depth has been pondered for decades. The rate of occurrence of earthquakes with depth falls roughly exponentially from near 50 to ~300 km, remains at a low level until ~500 km depth, then shows a marked resurgence before abruptly terminating near 680 km (**Figure 1**). A similar bimodal distribution is observed individually, with some variation, in every subduction zone with deep seismicity (Helffrich and Brodholt, 1991), and provides the basis of the longstanding distinction between intermediate and deep earthquakes.

To explore the three-dimensional (3-D) seismicity distributions in various subduction zones

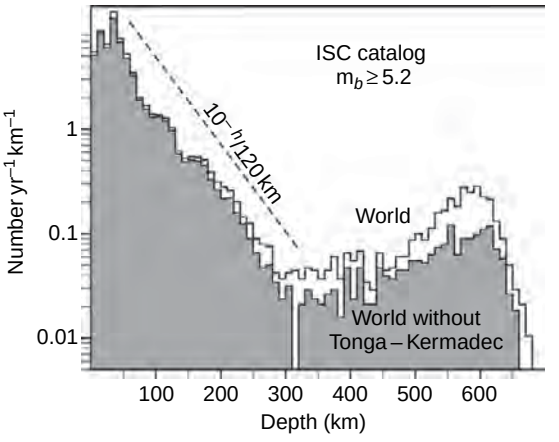


Figure 1 Earthquake occurrence rates vs depth for ISC catalog events from 1964 to 2001. The upper line corresponds to rates for the entire world, the lower line to the entire world excluding Tonga-Kermadec. The rate decreases roughly exponentially between 50 and 300 km. Adapted from Frohlich C (2006a) *Deep Earthquakes*. Cambridge, UK: Cambridge University Press.

interactively, reader may visit the website listed at the end of the chapter.

4.11.2.1 Nature of the Seismic Source

The seismic signals from deep earthquakes are described by a combination of double-couple and compensated linear vector dipole (CLVD) seismic sources. Although an implosive isotropic component of deep earthquakes has long been sought, in connection with the abrupt increases in density associated with upper-mantle phase changes, a nonzero isotropic component has not been reliably detected for a deep earthquake (Kawakatsu, 1996; Hara *et al.*, 1996).

Careful reanalysis of low-frequency spectra of the great 1970 Columbia earthquake indicates no significant isotropic component (Dziewonski and Gilbert, 1974; Russakoff *et al.*, 1997). Similarly, a very broadband analysis of the 1994 Bolivian earthquake limited an isotropic component to less than 1–2% (Hara *et al.*, 1995). The CLVD component, however, does appear more prevalent for deep than for shallow earthquakes, in that a higher proportion of deep earthquakes has a large CLVD component (Frohlich, 1995, table 1). However, the favored interpretation of this observation is that, except for some shallow earthquakes in volcanic regions, large CLVD components of both deep and shallow earthquakes usually result from a change in focal mechanism during rupture (e.g., McGuire *et al.*, 1997; Tibi *et al.*, 1999), as double-couple motion on two fault planes can sum to an apparent CLVD component (Frohlich, 1995). Thus, because an isotropic component does not seem to be present and a CLVD component can be explained as double-couple motions on more than one fault plane, deep earthquakes are believed to result from shear deformation, similar to the vast majority of shallow earthquakes. However, resolution of the fault-plane geometry is insufficient to constrain the deformations to lie on a single planar surface. Thus, the possibility that a deep earthquake could consist of deformation over a wide shear zone or on multiple en echelon planes (e.g., Chen, 1995) has not been definitively ruled out.

4.11.2.2 Numbers and Magnitudes of Deep Earthquakes

Less than a fifth of the Earth's seismicity occurs below 100 km (**Figure 1**). Eighteen percent of the ~6000 earthquakes with moment magnitude greater than M_w 5.8 in the Harvard CMT catalog from 1977 to 2006 have CMT depths greater than 100 km, and only 6% occur at depths between 400 and 700 km, below which no earthquakes have been reliably located (see Frohlich (2006a) for a discussion). Only about 2.2% occur within the seismically quiet interval from 300 to 500 km depth.

On the other hand, the vast majority of Earth's intraplate (within a tectonic plate) seismicity occurs below 50 km as deep earthquakes within subducted slabs (particularly so if continental seismicity near diffuse plate boundaries is excluded). Clearly, oceanic lithosphere becomes more seismogenic after it has subducted beneath Earth's surface. Thus, it appears that one or more of the following processes associated with subduction has seismically activated the oceanic lithosphere – the damage and/or hydration from slab bending at the trench and outer rise, the high temperatures and pressures, or the high internal stresses imposed on subducting slabs by negative buoyancy, thermal expansion, and volumetric phase changes. These phenomena are discussed in Sections 4.11.4 and 4.11.5.1.

4.11.2.2.1 Gutenberg–Richter statistics

Like shallow earthquakes, deep earthquakes generally follow Gutenberg–Richter statistics; that is, the number of earthquakes with magnitude greater than M is given by $n = a\,M^b$ where a and b are constants. For global sets of both deep and shallow earthquakes, the value of b is near 1.0 (Frohlich and Davis, 1993; Kagan, 1999). However, although regional variations are, in general, weak for shallow earthquakes, a remarkable systematic regional variation in the values of b is found for the deepest events (e.g., Giardini, 1988; Frohlich, 1989, table 1; Frohlich and Davis, 1993; Frohlich, 2006a). For events below 400 km Giardini and Lundgren (1995) obtained b-values that varied from ~0.4 for South America to ~1.2 for Tonga (**Figure 2**). Although the Tonga subduction zone lays claim to the largest number of deep earthquakes, South America hosts the largest-magnitude deep ruptures. While Tonga has about 10 times more M-5.5 events, the two regions have similar numbers of M-6.5 events, and the biggest deep earthquakes occur in South America (**Figure 2**). With respect to inferred slab temperatures (discussed in Section 4.11.4.1), Tonga and South America are cold and warm end members, respectively, of subducted slabs with deep seismicity. Strong thermal control of the deep magnitude–frequency relation is, therefore, implicated (e.g., Wiens and Gilbert, 1996; Wiens, 2001). Interestingly, the b-values of

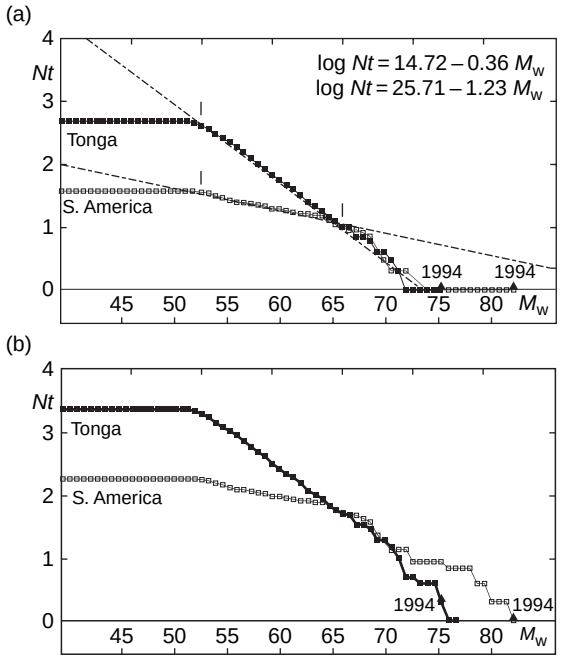


Figure 2 Magnitude–frequency relations for earthquakes below 400 km depth in Tonga (solid squares) and South America (open squares). The logarithm of the number of events with magnitude greater than a given value is plotted. (a) From Harvard CMT catalog from 1977 to 1994. Dashed lines, for which equations are given in the upper right, show fits to the data for magnitudes between M_W 5.3 and 6.6. Very different *b*-values (slopes) for the two regions were determined. The 1994 Bolivia and Tonga earthquakes are roughly consistent with the magnitude–frequency behavior at lower magnitudes. (b) From hybrid catalog extended back to 1900 by Giardini and Lundgren (1995) by including magnitudes from the literature for larger events and assuming that the magnitude–frequency behavior (of events up to M_W 7.2) during 1977–1994 is typical of that from 1900 to 1994. Adapted from Giardini D and Lundgren P (1995) The June 9 Bolivia and March 9 Fiji deep earthquakes of 1994: II. Geodynamic implications. *Geophysical Research Letters* 22: 2281–2284.

seismicity between 100 and 350 km deep are not significantly different for Tonga and South America. Evidently, thermal influence on earthquake generation does not operate in the same way above 350 km.

4.11.2.2.2 Isolated giant earthquakes

At depths greater than 400 km the very largest earthquakes tend to occur in South America, in isolated settings, and produce very few aftershocks for their size. A remarkable disparity, one related to the regional contrast in *b*-value noted above, is that the three largest-known earthquakes deeper than 300 km occurred in South America, whereas only 10% of M_w 6.0 earthquakes occur there compared to 60% in Tonga (from the Harvard CMT catalog between 525 and 650 km depth). Furthermore, 7 out of 19 of the largest earthquakes below 300 km occurred in South America compared to only 5 in Tonga-Fiji (see **Table 1**). The fifth largest deep earthquake, the 1954 *M*7.9 Spanish earthquake, also occurred in an isolated setting, in lithosphere associated with a previous episode of subduction (Blanco and Spakman, 1993).

There are several peculiarities of the 'South American giants' (e.g., Estabrook, 1999). The 1994 *M*8.3 Bolivia (Section 4.11.2.4.5), 1970 M_w 8.1 Colombia, and 1922 M_w 7.9 North Peru earthquakes are the three largest-known deep earthquakes below 300 km (**Table 1**). They occurred in seismically very quiet to relatively quiet regions that lie one to several hundred kilometers from well-defined Wadati–Benioff zones. Body-wave studies of the rupture directivity of the 1994 and 1970 events, as well as the 1963 M_w 7.7 Peru-Bolivia earthquake which occurred about 200 km west of the 1994 event, indicate that their rupture velocities are slower than typical, about 0.2–0.4 of the shear-wave velocity (Furumoto, 1977; Estabrook, 1999). Their seismic efficiencies (Sections 4.11.2.4.4 and 4.11.2.4.5) are correspondingly low as well (Estabrook, 1999). Source time functions of these three South American earthquakes (1994 Bolivia, 1970 Colombia, 1963 Peru-Bolivia) begin with a small-to-moderate subevent followed by the main moment release 10–20 s later (Furumoto, 1977; Brustle and Muller, 1987; Fukao and Kikuchi, 1987; Estabrook, 1999). The 1963 M_w 7.7 Western Brazil earthquake (**Table 1**) did not have such a precursor in its time function, although it did have a low rupture velocity (Fukao, 1972; Fukao and Kikuchi, 1987).

The contrasts between seismicity in the Tonga and South American slabs may indicate different physical mechanisms for Tonga and South America deep earthquakes (e.g., Wiens and McGuire, 1995), perhaps implicating different mechanisms for nucleation and propagation of deep rupture. It appears that earthquakes can nucleate more easily in Tonga, but propagate into larger ruptures more easily in South America. Thus, the properties that allow rupture to initiate appear to differ from those that promote its propagation. But further, the disparity between the contrast in *b* values below ~350 km and the similarity of *b* values above ~350 km suggests a difference between rupture mechanisms above and below that

Table 1 Largest deep earthquakes below 300 km from 1906 to 1 Sept. 2006, ordered by seismic moment[a]

Date	Area	Depth (km)	Moment (Nm)	M_w
09 June 1994	Bolivia	647	2.6×10^{21}	8.3
31 July 1970	Colombia	623	1.4×10^{21}	8.1
17 January 1922	North Peru	664	9.4×10^{20}	7.9
17 June 1996	Flores Sea	589	7.9×10^{20}	7.9
29 March 1954	Spain	630	7.0×10^{20}	7.9
29 September 1973	North Korea	593	5.0×10^{20}	7.8
11 June 1972	Celebes Sea	332	4.7×10^{20}	7.7
19 August 2002	Fiji	699	4.3×10^{20}	7.7
26 May 1932	Fiji	560	4.0×10^{20}	7.7
15 August 1963	Peru-Bolivia	573	3.9×10^{20}	7.7
28 February 1950	Sea of Okhotsk	339	3.9×10^{20}	7.7
25 May 1907	Sea of Okhotsk	548	3.7×10^{20}	7.7
27 January 2006	Banda	397	3.5×10^{20}	7.7
09 November 1963	Western Brazil	573	3.5×10^{20}	7.7
19 August 2002	Fiji	631	3.5×10^{20}	7.7
19 August 1961	Peru-Bolivia	620	3.4×10^{20}	7.7
09 March 1994	Fiji	563	2.7×10^{20}	7.6
23 May 1956	Fiji	436	2.7×10^{20}	7.6
26 July 1958	Peru-Bolivia	592	2.6×10^{20}	7.6

[a]Updated from table 1.2 of Frohlich (2006a), which was based on Huang and Okal (1998) augmented by the CMT catalog for events occurring since 1996.

depth. Possible mechanisms are discussed in Sections 4.11.3 and 4.11.6.

4.11.2.3 Aftershock Occurrence

Deep earthquakes are notable for producing very few aftershocks (Frohlich, 1987), although there are a few cases (e.g., 1994 Tonga) in which the aftershock sequences of deep earthquakes are as numerous as those for shallow earthquakes (Wiens and McGuire, 2000). In general, the incidence of aftershocks for mainshocks with depths greater than ~50 km is an order of magnitude lower than for shallow mainshocks (Frohlich, 1987). Nevertheless, the aftershocks produced appear to follow, on average, Omori's Law of aftershock decay (Nyffenegger and Frohlich, 2000).

Perhaps due to the paucity of deep aftershocks, Frohlich (1987) found little difference in deep aftershock production by geographic region. Subsequently, Wiens and Gilbert (1996) investigated regional variations in aftershock production for large mainshocks below 400 km and found that the most productive aftershock sequences are limited to cooler slabs, as defined by the thermal parameter (Section 4.11.4.1.1). Mainshocks in slabs with smaller thermal parameters, such as South America, Japan, and Izu Bonin, exhibited lower aftershock productivity than mainshocks in Tonga, Indonesia, and Marianas, which have larger thermal parameters. A more recent study of 290 somewhat smaller mainshocks failed to detect a clear relation of thermal parameter to aftershock productivity for either intermediate or deep earthquakes (Persh and Houston, 2004b). Nevertheless, for aftershocks deeper than 400 km, the situation probably parallels that of b values of seismicity in general as discussed above, that is, b values and aftershock productivity are high (i.e., more small events relative to large ones) for cooler slabs and low for warmer slabs.

Strong depth-dependence of aftershock productivity is seen in **Figure 3** from Persh and Houston (2004b); their study analyzed 290 mainshocks with $M_w \geq 6.3$ below 100 km from 1977 to 2004. Between 300 and 550 km depth, aftershock productivity falls to one-third of the relatively low level that prevails between 100 and 300 km compared to shallow crustal events. But below 550 km, productivity rises abruptly to twice the level between 100 and 300 km. **Figure 4(a)** shows the number of aftershocks per mainshock normalized to an M_w 8.2 mainshock (the largest in the data set), assuming the number of aftershocks scales linearly with $M_0^{2/3}$. Bootstrap analyses of the data indicate that the differences in aftershock productivity are statistically significant. The strong dependence of aftershock productivity on depth suggests that the reason few earthquakes occur in the depth range 300–550 km may be due to a difficulty in

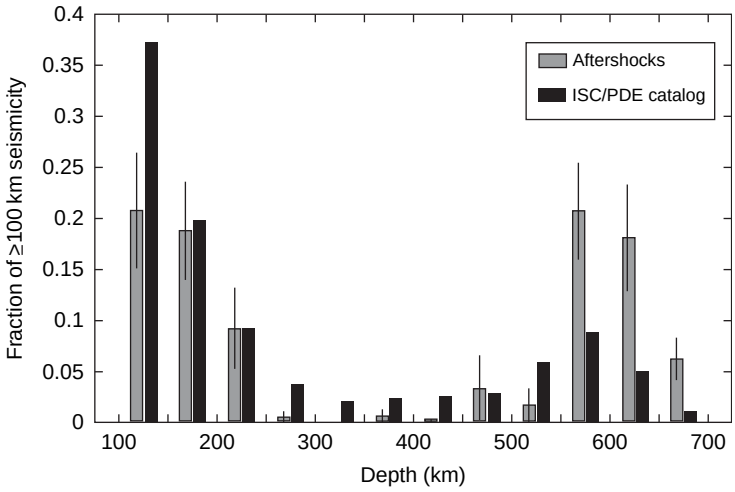


Figure 3 Depth distribution of aftershocks (light shading) compared to all deep seismicity (dark shading) from the ISC/PDE catalog. Aftershock numbers were normalized to account for mainshock moment (Persh and Houston, 2004b). The heights of the bars of each type sum to 1.0. The error bars on the aftershocks are one standard deviation and were determined by bootstrap resampling of the mainshocks in each depth bin. The distribution of aftershocks with depth differs from that of the background seismicity; at 300–550 km aftershock productivity relative to all seismicity is much weaker than at intermediate depths, whereas deeper than 550 km relative productivity is much stronger than at intermediate depths. Adapted from Persh SE and Houston H (2004b) Strongly depth-dependent aftershock production in deep earthquakes. *Bulletin of the Seismological Society of America* 94: 1808–1816.

nucleating ruptures rather than due to a lack of ambient deviatoric stress in that depth range, which is the conventional explanation and implied by the simple thermomechanical stress model discussed in Section 4.11.4.2. A mainshock of a given size between 300 and 550 km presumably produces a stress change similar in magnitude to that produced by a shallower or deeper event. The abrupt increase in aftershock productivity below 550 km is striking and perhaps indicative of a change in physical mechanism near that depth.

4.11.2.3.1 Long-range triggering

Large deep earthquakes can trigger other large deep quakes at great distances. Such long-range triggering is known to occur for shallow crustal earthquakes (e.g., Hill *et al.*, 1993). In one notable case in 2002, a M_w-7.6 event deep in the Tonga subduction zone (598 km deep) likely triggered a M_w-7.7 event (664 km deep) 313 km away 7 min later (Tibi *et al.*, 2003b). The latter event occurred well below the bottom of the previously seismically active zone in the region. Tibi *et al.* (2003b) give several other examples of triggered deep seismicity below 450 km, for which the time delays range from 2 to 143 min and the spatial separations range from 70 to 320 km. Such long-range triggering occurs mainly in Tonga, but is also seen in Bolivia and Japan. For example, just 10 min after the 1994 M_w 8.2 Bolivian earthquake, an M_w 6 event was triggered about 330 km east of the mainshock at 671 km depth (Myers *et al.*, 1995).

The triggered events were deeper than the mainshocks except in one case, and often occurred in nearly aseismic regions next to or below the seismically active zone (Tibi *et al.*, 2003b). The time delay of the triggered events does not correspond to the arrival of large seismic phases from the triggering events, suggesting that the triggered events take a few minutes to several tens of minutes to nucleate following the stress disturbance associated with the passage of the dynamic waves. The time delay could be related to the development of plastic instabilities (Tibi *et al.*, 2003b), or to the growth of anticracks in a metastable phase, to single out two possible mechanisms (Sections 4.11.3.2 and 4.11.3.3). The triggered seismicity is located, presumably, in regions on the verge of failure. The location of the triggered events, near the edges of the seismically active slab and/or deeper than the triggering events, seems more suggestive of transformational faulting because compared to conditions at the triggering event, the edge of the slab is a region of less kinetic hindrance, and the overstep of a metastable phase transition (i.e., heat

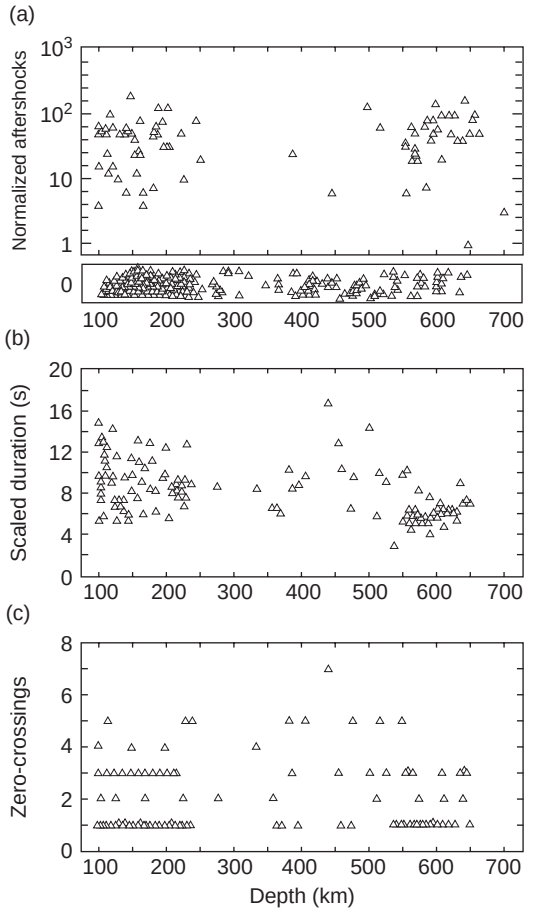


Figure 4 Depth-dependence of three source properties. Note differences in properties for three depth ranges (100 to ~350 km, ~350 to 550 km, 550 to 700 km), as well as abrupt change of properties near 550 km. (a) Normalized aftershocks. Number of aftershocks was scaled to remove effect of mainshock size. There are almost no aftershocks in the middle depth range, and a larger fraction of mainshocks below 550 km have aftershocks than mainshocks in the first depth range (Persh and Houston, 2004b). Events with no aftershocks are plotted below with random jitter added to reduce overlap. (b) Scaled duration. Durations determined from source time functions were scaled to remove the effect of earthquake size (Persh and Houston, 2004a). (c) Source complexity measured by zero-crossings of the time derivative of low-passed, scaled source time functions (Persh and Houston, 2004a). Complexity appears greater in the middle depth range. Adapted from Persh SE and Houston H (2004b) Strongly depth-dependent aftershock production in deep earthquakes. *Bulletin of the Seismological Society of America* 94: 1808–1816.

release and Gibbs free energy) increases with depth, both of which would favor transformation. In contrast, at the relevant depths (~600 km), the instability parameter for thermal shear instability typically decreases with depth, and thus is smaller near the triggered event than near the triggering event, so that shear instability would be less favored at the site of the triggered event than at the site of the triggering event (Section 4.11.3.3 and **Figure 13**).

4.11.2.4 Rupture Processes of Deep Earthquakes

As with shallow earthquakes, our knowledge of the rupture process of deep earthquakes (i.e., the progression of rupture through space and time) comes from analysis of the shape and duration of seismic waveforms, particularly the direct P and S body waves. Unlike for shallow earthquakes, for deep earthquakes these waves are not complicated by the depth-reflected phases pP, sP, or sS, because those phases arrive long after the direct wave. In addition, the paths of body waves from deep earthquakes traverse a smaller length of the highly heterogeneous material near Earth's surface and, therefore, suffer less attenuation, both intrinsic and scattering. Thus, deep earthquake body waves are relatively clean, simple, and sharp (**Figure 5**) compared to the body waves from shallow earthquakes. In the analysis of deep earthquake source properties, it is advisable to use depth-dependent values for such parameters as seismic velocity, rigidity, attenuation (e.g., t^*), and geometrical spreading (which is also distance dependent), as these vary significantly from the top to the base of the upper mantle.

4.11.2.4.1 Duration of rupture

It has long been clear that deep earthquakes (especially very deep ones) tend to be briefer than shallow ones of similar magnitude. It is relatively direct to obtain the duration of rupture from broadband seismograms, although care must be taken to determine the time of termination (**Figure 5**). Although there is significant variability, earthquake durations generally scale as seismic moment to the one-third (e.g., Kanamori and Anderson, 1975; Vidale and Houston, 1993; Houston *et al.*, 1998; Houston, 2001). Thus, to determine how duration may depend on various parameters, it is helpful to consider scaled duration, which is duration divided by moment (relative to a reference moment) to the one-third. Scaled durations from a global set of time functions of shallow and deep earthquakes generally decrease gradually with increasing depth, but abrupt decreases are seen near 40 and 550 km (Houston, 2001). The abrupt decrease near 40 km is interpreted as due to the contrast in stress drop and/or shear velocity between ruptures

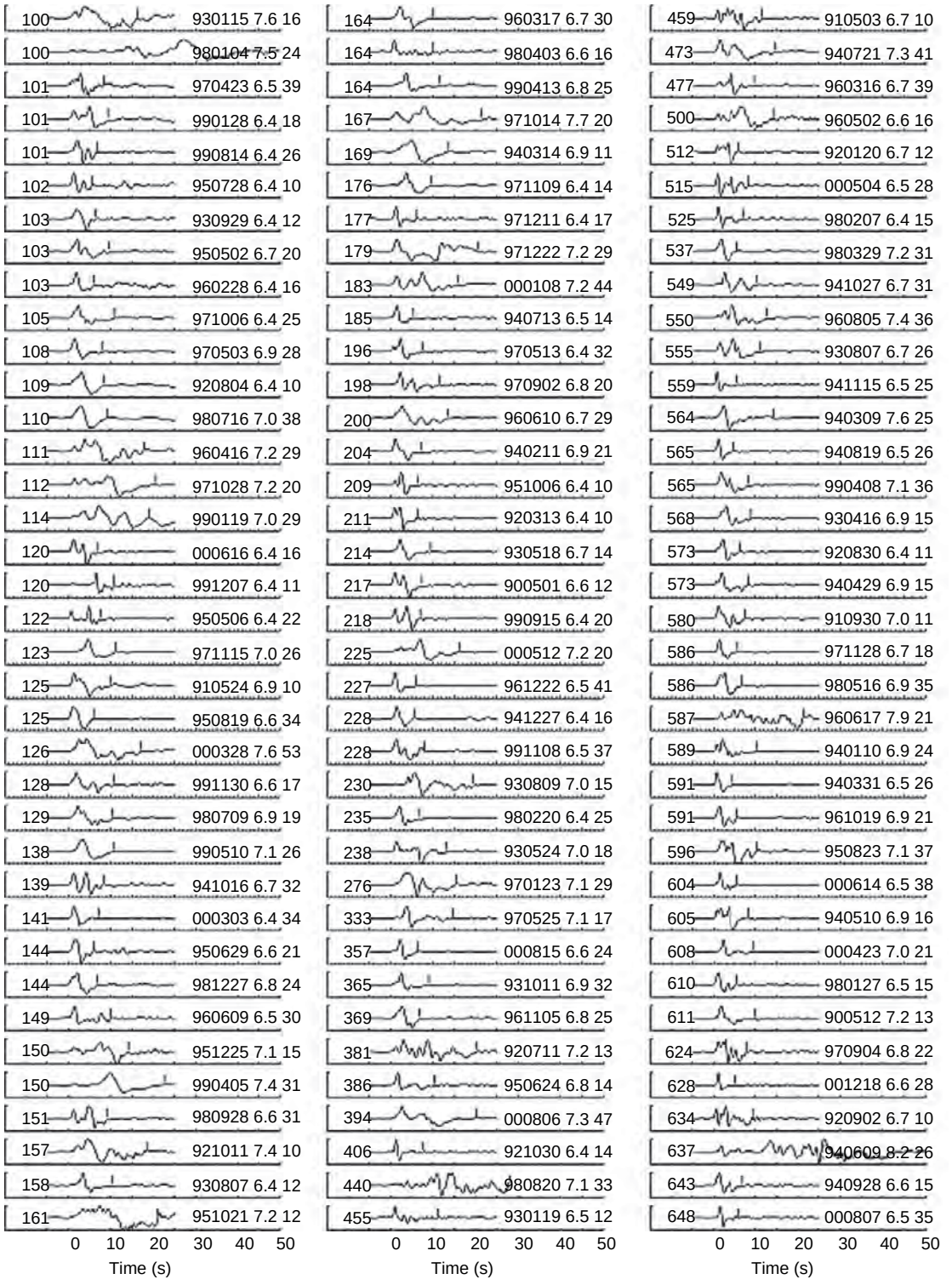


Figure 5 Stacks of P-wave velocity seismograms for 111 large earthquakes deeper than 100 km. Stacks were constructed by aligning P waves on a small feature early in the rupture, then summing seismograms. Depth, date, M_W, and number of stations stacked are given for each event. Stacks are ordered by depth from the top left down each column to the bottom right. Adapted from Persh SE and Houston H (2004a) Deep earthquake rupture histories determined by global stacking of broadband P waveforms. *Journal of Geophysical Research* 109 (doi:10.1029/2003JB002762).

occurring between plates (the global data set used consisted mostly of thrust earthquakes in subduction zones) and those within plates (below 40 km). The abrupt decrease near 550 km was corroborated by Persh and Houston (2004a) using stacks of carefully aligned broadband P waves of 111 large events below 100 km (**Figures 4(b)**, **5**, and **6**). Tocheport *et al.* (2006) obtain similar results using a similar data set and a more automated alignment and stacking method. The durations deviate somewhat from a simple model in which rupture velocity is a constant fraction of shear velocity and all else (e.g., stress drop, shape of rupture area) remains constant (see Persh and Houston (2004a), figure 6). Such a deviation near 550 km, within considerable variability, can also be detected in Frohlich (2006a)'s figure 6.4, although that author prefers an interpretation involving no deviation from the simple model. As with the aftershock productivity discussed above (Section 4.11.2.3) and the rupture complexity discussed next, the abrupt change in scaled duration near 550 km suggests the possibility of a change in the physical mechanism of rupture.

4.11.2.4.2 Complexity of rupture

As in shallow earthquakes, deep earthquake ruptures are often complex, with multiple rupture subevents (e.g., **Figure 6**). A significant fraction, roughly half, of events below 100 km depth exhibit multiple subevents in their waveforms and time functions (e.g., Fukao and Kikuchi, 1987; Houston and Vidale, 1994; Houston *et al.*, 1998; Antolik *et al.* 1999; Tibi *et al.*, 2002; Tibi *et al.*, 2003a; Persh and Houston, 2004a). Houston *et al.* (1998) proposed that complexity be measured quantitatively by the number of zero-crossings of the time derivative of the scaled source time function, after low-pass filtering to equalize the effect of event magnitude. Using this approach, Houston *et al.* (1998) and Persh and Houston (2004a) found that events between 350 and 550 km depth have greater complexity (**Figures 4(c)** and **6**) compared to events between 100 and 350 km, in addition to being longer than expected, while events below 550 km are simpler than either group above them, in addition to being briefer as discussed in Section 4.11.2.4.1. Thus, in the 350–550 km depth range, events are more complex, longer in scaled duration relative to shear velocity, and have fewer aftershocks, compared to events above and below that depth range (**Figure 4**). Perhaps the aftershocks of events in the 350–550 km depth range occur immediately as part of the main rupture, leading to fewer measured aftershocks but greater complexity of the mainshock. The abruptness of these changes in several attributes (i.e., complexity, aftershock productivity, duration) at about 550 km depth may suggest a change in earthquake generation mechanism (Persh and Houston, 2004a, 2004b).

4.11.2.4.3 Spatial pattern of rupture, rupture velocity, and stress drop

The progression of rupture in space and time is a key characteristic that bears on the mechanics and physics of fault failure. The variation with azimuth of body waveforms provides the main constraint on the progression of rupture (**Figure 7**). Under most circumstances, what can be resolved is the velocity of propagation of a significant amount of slip, rather than the propagation of a crack tip. Many studies have forward-modeled or inverted P and SH waveforms based on parametrizations of deep ruptures as line sources or multiple point sources (e.g., Glennon and Chen, 1995; Tibi *et al.*, 2002; Tibi *et al.*, 2003a). Such parametrizations yield some aspects of the spatiotemporal distribution of rupture, including fault length and rupture velocity. For example, Glennon and Chen (1995) and Tibi *et al.* (2002) modeled eight and six events, respectively, with a line source, obtaining a length of rupture, source time functions, and average rupture velocities. Although the two-dimensional (2-D) pattern of slip over a fault plane is more difficult to constrain, a number of studies have determined 2-D slip in large deep earthquakes (e.g., Fukao and Kikuchi, 1987; McGuire *et al.*, 1997; Antolik *et al.*, 1999, as well as numerous studies of the 1994 Bolivia earthquake summarized in Section 4.11.2.4.5).

Determination of the spatial distribution of slip in deep earthquakes is complicated by several issues. Because few or no aftershocks occur to delineate the rupture surface, it is usually difficult to identify which of the nodal planes is the fault plane, although Warren *et al.* (2007) developed an approach to detect which of the nodal planes is the fault plane based on directivity. Also, deep earthquakes are more compact than most shallow ones. And there is rarely an abundance of well-distributed low-noise seismometers.

Taken together, the results of existing studies indicate that there is much variability in fault geometry and aspect ratios, rupture velocities, and stress drops (e.g., see Frohlich, 2006a). There appears to be a tendency for ruptures to propagate along strike and near horizontally (Antolik *et al.*, 1999; Tibi *et al.*, 2002; Warren *et al.*, 2007).

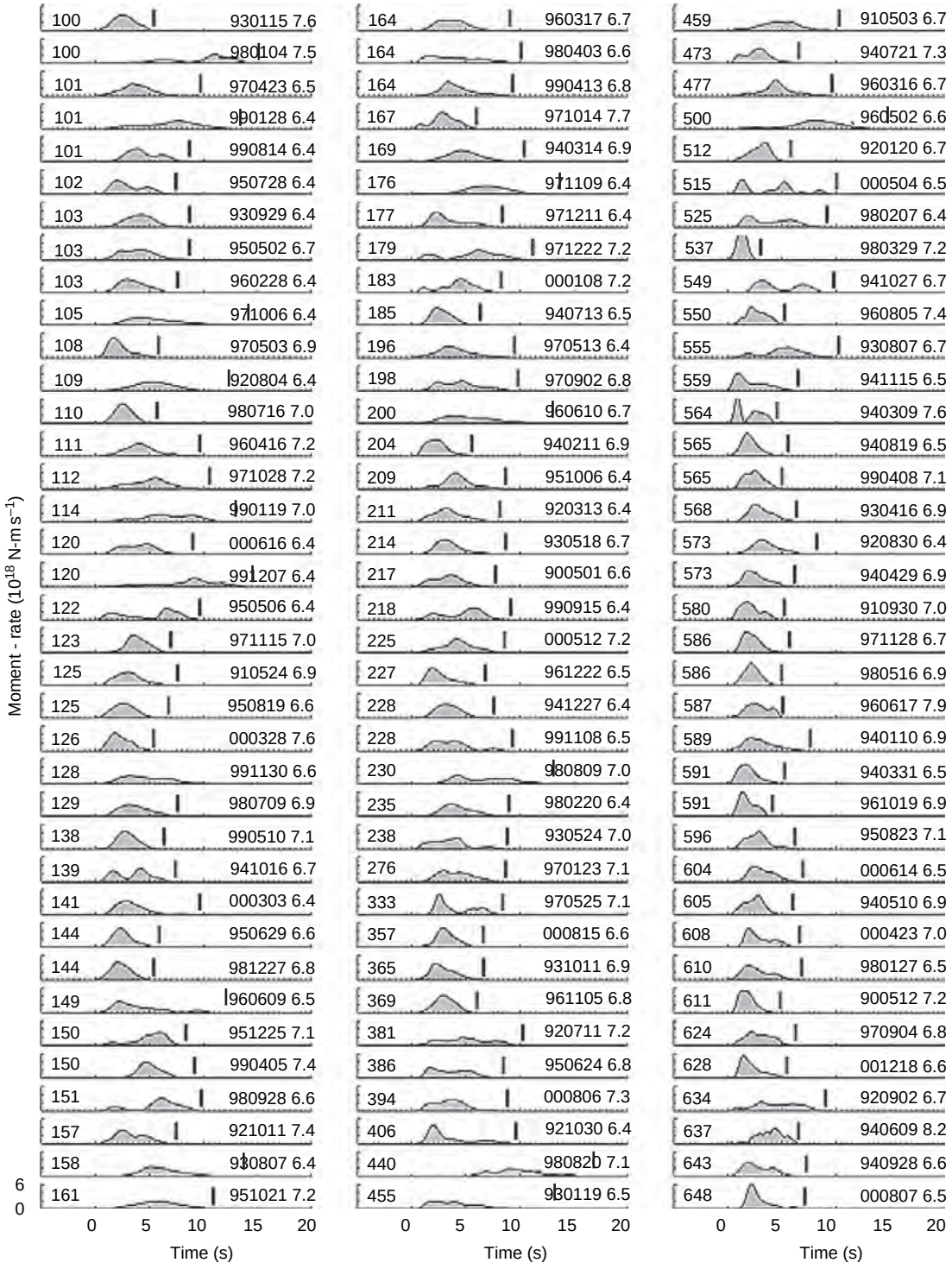


Figure 6 Scaled source time functions in displacement from P-wave stacks in **Figure 5**. Time functions were scaled to removed effect of earthquake size as follows: time is scaled by the cube-root of seismic moment, while amplitude is scaled by the square of the cube root of seismic moment so that the area under the time function equals the seismic moment. Stacks are ordered by depth from the top left down each column to the bottom right. Note simpler shapes for time functions deeper than ~550 km. Adapted from Persh SE and Houston H (2004a) Deep earthquake rupture histories determined by global stacking of broadband P waveforms. *Journal of Geophysical Research* 109 (doi:10.1029/2003JB002762).

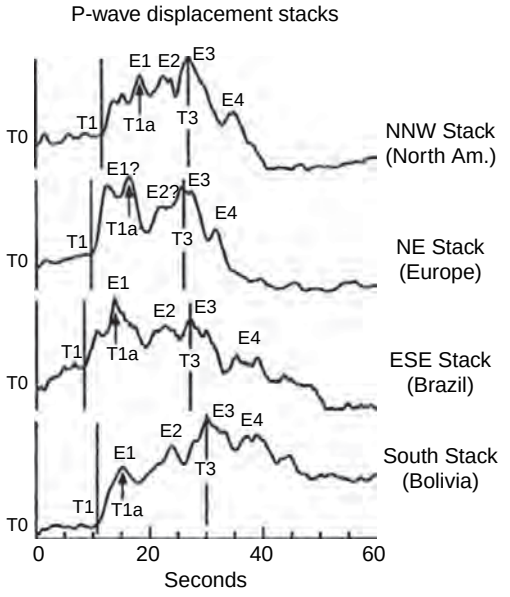


Figure 7 Stacked P waveforms of the 1994 Bolivian earthquake in different azimuthal directions. The variations with azimuth constrain the location and timing of high-slip regions. Labels indicate timing of subevents within the rupture. The earthquake initiated at time *T*0, but moment-release rate increased sharply about 11 s later. Adapted from Beck SL, Silver P, Wallace TC, and James D (1995) Directivity analysis of the deep Bolivia earthquake of June 9, 1994. *Geophysical Research Letters* 22: 2257–2260.

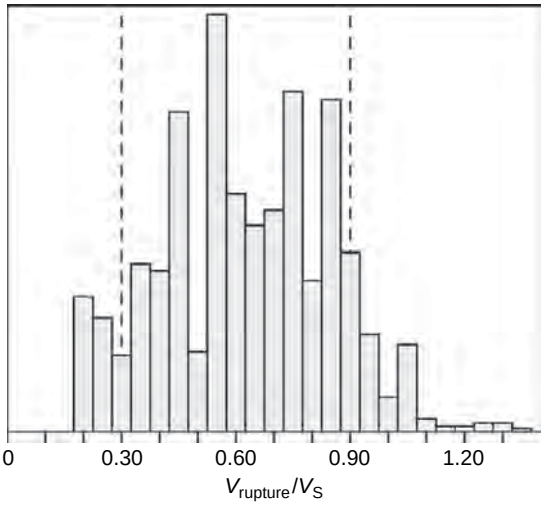


Figure 8 Histogram of reported rupture velocities from a compilation of source studies of events with $M_W \geq 6.3$ and depth between 60 and 700 km. Velocities are plotted as a fraction of shear velocity at the source depth. About 80% of the estimates fall between the dashed lines. Adapted from Frohlich C (2006a) *Deep Earthquakes*. Cambridge, UK: Cambridge University Press.

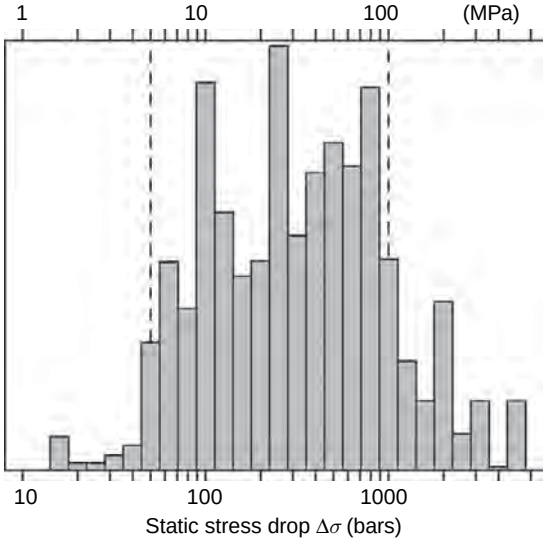


Figure 9 Histogram of reported static stress drops from a compilation of source studies of events with $M_W \geq 6.3$ and depth between 60 and 700 km. About 80% of the estimates fall between the dashed lines. Adapted from Frohlich C (2006a) *Deep Earthquakes*. Cambridge, UK: Cambridge University Press.

Rupture velocities are typically given as a fraction of shear-wave velocity β in the source region. The distribution of ~100 estimates of rupture velocities from various studies is shown in **Figure 8** after Frohlich (2006a). Notably, rupture velocities appear to vary more for deep than for shallow earthquakes–from 0.2β to 0.9β (Frohlich, 2006a) or, arguably, even faster (e.g., Kuge, 1994). For most intermediate and deep earthquakes, rupture velocities range from 0.4β to 0.9β (e.g., Tibi *et al.*, 2002, 2003a; Warren *et al.*, 2007), but they are notably slower (0.2–0.4β) for very large deep South American events–1963 Peru-Bolivia, 1970 Colombia, and 1994 Bolivia (Furumoto, 1977; Beck *et al.*, 1995; Estabrook, 1999). A general dependence of rupture velocity on slab thermal state has been proposed, in which earthquakes in warmer slabs rupture more slowly (Wiens, 2001; Tibi *et al.*, 2003a). The variability and difficulty of resolution of rupture velocity currently precludes a definitive evaluation of that proposal.

Static stress drops are difficult to determine accurately, but appear to range more than an order of magnitude from ~1 to ~50 MPa (e.g., Antolik *et al.*, 1999; Tibi *et al.*, 2002, 2003a), and even higher for the Bolivian event, discussed below. **Figure 9** taken from Frohlich (2006a) shows ~100 estimates of static stress drops of deep earthquakes compiled from numerous

studies. As Antolik *et al.* (1999) pointed out, difficulties in resolving details of slip due to attenuation at teleseismic distances and limited azimuthal station coverage, result in an inferred spatial slip distribution that is smoothed and broadened compared to the actual slip, so that actual static stress drops of deep earthquakes are likely to be more heterogeneous and, in particular, larger than most studies infer.

4.11.2.4.4 Radiated seismic energy

Seismically radiated energy is a fundamental measure of an earthquake source, and can be determined from P and S body waves. Whereas longer-period waves are suitable for measuring seismic moment, seismic energy is best measured by broadband waves that contain the corner frequency, which is the inverse of the rupture duration. Therefore, radiated energy is typically determined from body waveforms. It is proportional to the integral over time (or frequency) of squared far-field velocity. Houston and Williams (1991) calculated radiated energies from broadband records and used them to obtain Orowan stresses, which are twice the apparent stresses (the rigidity times the energy/moment ratio), for 68 globally distributed earthquakes deeper than 100 km. Values of Orowan stresses and apparent stresses are three to five times larger than those for most shallow earthquakes (Choy *et al.*, 2006), which occur in plate boundaries (interplate), but are comparable to those of outer-rise earthquakes, which occur within the oceanic lithosphere (Green and Houston, 1995). No robust dependence on depth was detected (Houston and Williams, 1991).

Radiation efficiency is the ratio of radiated energy to radiated plus fracture energy (see e.g., Venkataraman and Kanamori, 2004), and is equal to twice the ratio of apparent stress to static stress drop. High values (i.e., near 1) imply that much of the available nonfrictional energy change in the earthquake went into the seismic radiation; low values imply that a high proportion of the energy was dissipated in a 'fracture' process. While the energy/moment ratios of deep earthquakes are, perhaps surprisingly, not dissimilar to those of shallow quakes, some large deep events have very low radiation efficiency due mainly to inferred high static stress drops. A few large events below 400 km have very low values of radiation efficiency (<0.1) (Winslow and Ruff, 1999; Tibi *et al.*, 2003a; Venkataraman and Kanamori, 2004); these events occurred in warmer-to-moderate slabs (of slabs possessing deep events) with somewhat lower thermal parameters (Estabrook, 1999; Tibi *et al.*, 2003a). The events include the 1994 Bolivia, 1970 Colombia, 1963 Peru-Bolivia, 1994 Japan Sea, 1999 Eastern China, and 2002 Eastern China earthquakes. This has been interpreted as evidence for energy-dissipation on the fault zone through a thermal process such as melting during rupture (e.g., Kanamori *et al.*, 1998; Tibi *et al.*, 2003a; Venkataraman and Kanamori, 2004).

4.11.2.4.5 The M 8.2 1994 Bolivian earthquake

Many studies have examined the spatiotemporal distribution of slip in the largest-known deep earthquake, the 1994 M_w 8.2 Bolivia earthquake (e.g., Kikuchi and Kanamori, 1994; Chen, 1995; Estabrook and Bock, 1995; Lundgren and Giardini, 1995; Silver *et al.*, 1995; Ihmle, 1998). The data sets used for these studies were largely similar suites of body waveforms, although somewhat different methods of inversion were applied. The results consistently locate the highest slip ~20 km northeast of the hypocenter occurring ~20 s after seismic rupture initiation. The spatial distribution implies a high static stress drop of ~100 MPa (e.g., Kikuchi and Kanamori, 1994). Because the high slip is located northeast of the hypocenter, but at about the same depth, the near-horizontal nodal plane of the focal mechanism is generally thought to be the rupture plane, and a slow rupture velocity of 1–2 km s^{-1} is inferred. This is roughly consistent with the aftershock distribution (Myers *et al.*, 1995). However, Chen (1995) fit the waveforms with relatively fast-growing circular ruptures on three closely spaced near-vertical planes in an en echelon arrangement. In that study, the centroids of the large vertical planes are located near the high-slip region found in the other studies, so that the apparent rupture speed is again 1–2 km s^{-1}. Chen's model has not been ruled out, which illustrates the difficulty of resolving detailed patterns of slip even in large deep earthquakes (this difficulty still persists for shallow earthquakes in which the fault plane is well known). Thus, the argument that the rupture plane of the Bolivian earthquake is too wide to fit within a metastable olivine wedge can perhaps be circumvented (Section 4.11.6.1). In this regard, thickening of the slab above 670 km due to resistance to penetration into the lower mantle has also been suggested (Kirby *et al.*, 1996). However, most thermokinetic modeling to date argues against the persistence of metastable olivine to the base of the upper mantle in any but the coolest slabs, and particularly not in

warmer ones such as South America (Devaux *et al.*, 1997; Mosenfelder *et al.*, 2001; see Section 4.11.4.1).

The shape of the 1994 Bolivian earthquake source time function is quite distinctive (**Figures 6** and **7**). After an initial rapid rise of ~2 s, a relatively low level of moment release (corresponding to an M_W 7.3 event) is sustained for ~10 s, followed by a strong breakout phase of ~30 s duration consisting of multiple strong pulses (e.g., Kikuchi and Kanamori, 1994; Beck *et al.*, 1995; Persh and Houston, 2004a). A similar delay for the major episode of moment release is seen in the 1970 Columbia and the 1963 Peru-Bolivia time functions (Furumoto, 1977; Brustle and Muller, 1987; Fukao and Kikuchi, 1987; Estabrook, 1999). Such a time function shape suggests that one physical mechanism of failure may occur in the initial stages of the deep South American giants followed by a different mechanism which generates most of the slip. Consideration of the space-time slip distribution, as well as the shape of the time function, suggests a difference between the locations within the slab where a rupture can initiate easily and where it can grow largest (Section 4.11.2.2.2).

Kanamori *et al.* (1998) consider the energy balance of the Bolivian earthquake. The radiation efficiency, which is the maximum seismic efficiency, is estimated from the radiated energy-to-moment ratio and the static stress drop. A small value for radiation efficiency, 0.036, is obtained which leads to a minimum frictional stress of 55 MPa. From the minimum frictional stress, Kanamori *et al.* (1998) calculated that melting would occur if the thickness of the fault zone over which the displacement takes place was less than 0.3 m. The very low value of radiation efficiency indicates that most of the energy released was dissipated rather than radiated as seismic waves, which is consistent with the low rupture velocities determined by most studies (e.g., Kikuchi and Kanamori, 1994; Beck *et al.*, 1995). Kanamori *et al.* (1998) inferred that once slip was initiated and sufficient displacement took place, melting probably prolonged the rupture (Section 4.11.3.3).

4.11.3 Possible Mechanisms of Deep Earthquake Generation

Here the author outlines the three primary mechanisms proposed for the generation of deep earthquakes. These mechanisms have some distinct implications for deep earthquake locations and source processes, but data are as yet insufficient to distinguish definitively between them. The mineralogy of subducting slabs is sufficiently complex that a number of somewhat more specific mechanisms have been proposed over the years, but limitations on space preclude their discussion.

4.11.3.1 Dehydration Embrittlement

Dehydration embrittlement refers to brittle failure assisted by high fluid pore pressures that counteract the high normal stresses due to large overburden pressures. In the brittle failure of an intact material in a deviatoric stress field, failure occurs by growth and linkage of tensile (Mode I) microcracks (Lockner *et al.*, 1991; Scholz, 2002) oriented perpendicularly to the maximum compressive stress (**Figure 10**). The tensile cracks interact via the stress concentrations around their tips, self-organize, and coalesce into a through-going shear crack, which is oriented at an oblique angle to the microcracks and on which macroscopic failure and sliding takes place. A shear crack cannot grow in its own plane, but needs such Mode I microcracks to create a process zone through which the shear crack can extend (Petit and Barquins, 1988). Formation of Mode I microcracks, and thus the resulting shear failure, is inhibited by the application of pressure. High fluid pressures (i.e., negative pressures) can hold open Mode I microcracks that would otherwise be closed by overburden pressures. Deformation experiments in sepentinite showed the viability of this mechanism (Rayleigh and Paterson, 1965). In subducting slabs, the fluid may come from pore fluids squeezed from downgoing sediments or various metamorphic phase transitions in the subducting oceanic plate (Davies, 1999). Oceanic crust is pervasively hydrated at mid-ocean ridges, while oceanic lithosphere may be hydrated by fluid transport along outer-rise normal faults (e.g., Peacock, 2001). Bending-related faulting penetrates deep (at least 20 km) into the mantle at the Middle America trench and probably most subduction zones, and likely promotes hydration of the deep crust and mantle surrounding the faults (Ranero *et al.*, 2003).

The viability of dehydration embrittlement as a mechanism to generate deep earthquakes depends on the availability of fluids at relevant depths. The isotope Be^{10} is formed only at Earth's surface, so its presence in arc magmas is strong evidence that water in sediments is carried down by subducting slabs (Morris and Tera, 1989). There is abundant evidence that slabs may dewater via various phase transitions down to about 200–250 km depth (e.g., see

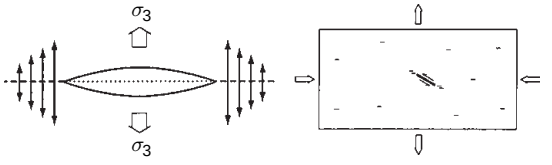


Figure 10 Relation between Mode I microcracks and brittle shear failure. On left, a tensile stress σ_3 opens a Mode I microcrack perpendicular to σ_3. The microcrack concentrates tensile stress at its tips, promoting further growth. On right, brittle shear failure occurs when such microcracks interact and link up. Macroscopic shear slip occurs at an oblique angle to the direction of greatest principal stress. Adapted from Frohlich C (2006a) *Deep Earthquakes*. Cambridge, UK: Cambridge University Press and Green HW and Houston H (1995) The mechanics of deep earthquakes. *Annual Review of Earth and Planetary Sciences* 23: 169–213.

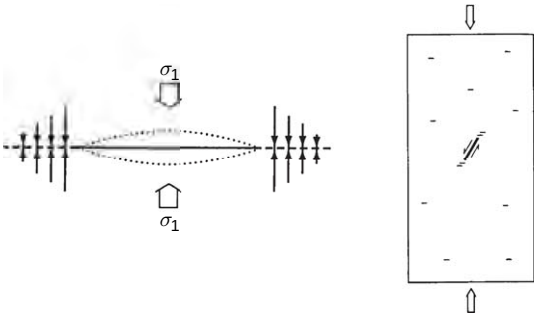


Figure 11 Relation between Mode I anticracks and transformational faulting. On left, a compressional stress σ_1 promotes transformation to a denser phase in a lense-shaped anticrack oriented perpendicular to σ_1. The anticrack concentrates compressional stress at its tips, promoting further growth. On right, anticracks can interact, link up, and coalesce to form a plane on which macroscopic shear slip occurs at an oblique angle to the direction of greatest principal stress. Adapted from Frohlich C (2006a) *Deep Earthquakes*. Cambridge, UK: Cambridge University Press and Green HW and Houston H (1995) The mechanics of deep earthquakes. *Annual Review of Earth and Planetary Sciences* 23: 169–213.

Thompson, 1992). Laboratory experiments show that there are numerous possible mineralogies and transitions that could occur, many of them involving release of water (e.g., table 1 of Omori *et al.*, 2001). A popular candidate transition, due to the abundance of antigorite serpentinite, is the breakdown of serpentinite to olivine plus enstatite plus water. A special issue of the *Bulletin of the Earthquake Research Institute* (2001, vol. 76, no. 3 and 4) focuses on 'The role of water in earthquake generation'.

More recently, attention has turned to dense hydrous magnesium silicate phases (so-called alphabet phases A through E), which could carry water deeper into the mantle as their stability fields range from ~200 km depth to within the lower mantle. However, the ability of such hydrous phases to produce deep earthquakes is not clear, because as slab material subducts following a low-temperature path to the higher pressures and temperatures at depth, the hydrous phases are capable of storing increasing amounts of water (e.g., Kohlstedt *et al.*, 1996; Angel *et al.*, 2001). Thus, net water would not be released during the phase transitions (although it might be available transiently), and the availability of free fluid to promote Mode I cracks is questionable (Green and Marone, 2002).

4.11.3.2 Anticrack Faulting in a Metastable Phase

The possible relationship between deep earthquakes and the well-known phase transitions of olivine silicate to progressively denser forms through the upper mantle has long been pondered (e.g., see reviews by Green and Houston, 1995; Kirby *et al.*, 1996; Frohlich, 2006a). However, seismic signals clearly show the deformation is shear, rather than implosive, in character (Section 4.11.2.1). Anticrack faulting provides a self-organizing mechanism analogous to fluid-assisted brittle faulting (i.e., dehydration embrittlement), by which the volume change and heat release of a phase transformation can trigger shear instability in a deviatoric stress field.

Anticracks are mechanically analogous to Mode I cracks except that the signs of the displacement and stresses are reversed (Fletcher and Pollard, 1981). In the context of olivine phase transitions, anticracks comprise lenticular 'pods' of the transformed and denser phase (i.e., β- or γ-spinel). In contrast to Mode I cracks, anticracks are oriented perpendicular rather than parallel to the maximum compressive stress (**Figure 11**). Similarly to mode I cracks, the stresses induced around anticrack tips promote the lengthening or extension of the anticrack, that is, high normal stresses near their tips promote further phase transformation, which leads to their crack-like lenticular shape (**Figure 11**). Laboratory experiments on analog materials indicate that these structures can link up to form a through-going shear zone oriented not parallel, but obliquely, to the anticracks (**Figure 11**), on which macroscopic shear faulting takes place (e.g., Green and Burnley, 1989; Kirby *et al.*, 1991; and see Green and Houston 1995 for a

review). The shear instability requires that the starting phase exist metastably in the stability field of the final phase, so that anticracks filled with the final phase can grow slowly (e.g., aseismically). The final (equilibrium) phase was found in an extremely fine-grained form in the shear zone, as well as in the anticracks, suggesting that it lubricated the shear zone permitting shear slip to occur. A key factor required for the instability or runaway is that latent heat be released by the transformation (Green and Zhou, 1996). As more anticracks form, heat is released, pushing the local region further away from the stability field of the metastable phase, and resulting in a greater overstep of equilibrium conditions. Experiments on analog materials which have a positive volume change on transformation show that transformational faulting can proceed with either positive or negative volume change, but the reaction must be exothermic (Green and Zhou, 1996; Green and Marone, 2002, table 1).

The viability of transformational faulting as a mechanism for deep earthquakes hinges on the presence of a sufficient amount of metastable phase (**Figure 12**). As the most abundant mineral in oceanic slabs is olivine, the most relevant phases for slabs in the upper mantle are α-olivine $\rightarrow$ β-spinel and α-olivine $\rightarrow$ γ-spinel. The equilibrium boundary of the olivine–spinel phase transition, normally at 410 km depth, is thermodynamically elevated in the vicinity of cold subducting slabs by 50–100 km as evidenced by an elevated 410 km seismic discontinuity near cold slabs (Vidale and Benz, 1992). Transformational faulting in olivine provides, therefore, a plausible mechanism only for earthquakes deeper than ~350 km. Thus, a key requirement for transformational faulting is the kinetic inhibition of transformation of a cold 'wedge' of the low-pressure phase, bounded approximately by a critical isotherm as the slab gradually subducts to deeper higher pressure–temperature environs. This requires sufficiently low temperatures inside the slab, discussed in Sections 4.11.4.1 and 4.11.6.1. Further, the extent to which transformation will continue to be inhibited on the long timescales relevant to subduction (e.g., several million years) is unclear. Seismic evidence for the low velocities in the form of distorted waveforms associated with a metastable olivine wedge has not been found (Vidale *et al.*, 1991; Koper and Wiens, 2000).

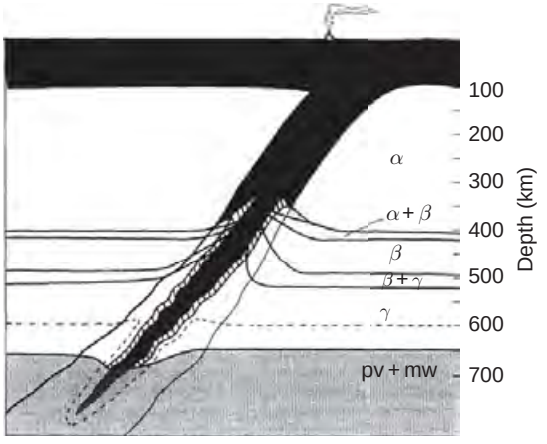


Figure 12 Schematic of metastable olivine wedge (dark region in core of slab below 350 km) in a cold subducting slab. Solid lines show the phase boundaries for the α, β, and γ polymorphs of mantle olivine. The hatchured region represents a partially reacted region, proposed to be available for the generation of aftershocks under sufficiently high deviatoric stress (Green and Zhou, 1996). The dashed line shows the conditions for the metastable enstatite-to-ilmenite transformation. Adapted from Green HW and Zhou Y (1996) Transformation-induced faulting requires an exothermic reaction and explains the cessation of earthquakes at the base of the mantle transition zone. *Tectonophysics* 256: 39–56.

In addition to the high-pressure polymorphs of olivine, phase transitions of other slab minerals could produce transformational faulting, although they are significantly less abundant than olivine. Hogrefe *et al.* (1994) reported that enstatite should persist metastably to greater depths than olivine, and proposed that enstatite $\rightarrow$ ilmenite, an exothermic reaction with a large volume decrease (12%), could generate deep earthquakes via transformational faulting of metastable enstatite. Although enstatite comprises only ~25% of the slab by volume, that may result in a sufficient amount of metastable phase to permit transformational faulting.

Thus, this mechanism provides a natural explanation of the abrupt cessation of seismicity at the base of the upper mantle. Transformational faulting would not be expected to occur in the lower mantle because the possible reactions are either endothermic (γ-olivine $\rightarrow$ perovskite + magnesiowustite and ilmenite $\rightarrow$ perovskite) or disproportionation reactions (α-olivine $\rightarrow$ perovskite + magnesiowustite and garnet $\rightarrow$ pervoskite phases) that require diffusion on the scale of the grain size and thus likely run too slowly to propagate a fault (Green and Zhou, 1996).

It is instructive to consider the similarities between dehydration embrittlement and transformational faulting (cf. **Figures 10** and **11**). In both cases, mode I tensile cracks or compressional anticracks filled with fluid or transformed material, respectively, self-organize in a

deviatoric stress field via the stress concentrations at their tips to form a through-going shear zone (Green and Zhou, 1996). In both cases shear zones are oriented, not in the plane of the mode I features, but about 30° to the direction of maximum compressive stress (**Figures 10** and **11**). Given this fundamental mechanical similarity despite large differences in the microphysics, together with the way that dynamic stresses drive rupture, the lack of dramatic changes in earthquake source properties with depth (e.g., stress drop, duration, double-couple nature) is not necessarily indicative of the operation of only a single physical mechanism of failure.

4.11.3.3 Thermal Shear Instability

Although transformational faulting is, in a sense, a type of thermal shear instability (in that heat is released), in the deep earthquake literature, the term 'thermal shear instability' refers to shear localization produced by a positive feedback between temperature-dependent rheology and shear deformation that generates viscous heating. The nature of the feedback is such that under certain conditions the localization of shear strain increases and accelerates exponentially, leading to apparently abrupt failure on a shear zone.

Variants of this mechanism have been proposed as a mechanism for deep earthquakes for many years (e.g., Griggs and Handin, 1960; Ogawa, 1987; Hobbs and Ord, 1988). Recently, Karato *et al.* (2001) observed that the most deformed slabs tend to be cold slabs. In their study of the dependence of slab rheology on temperature, Karato *et al.* (2001) incorporated the evolution in grain size associated with the transformation of olivine to spinel, plastic flow laws of the two phases, and a slab thermal model with latent heat release; three deformation mechanisms for olivine and spinel were included. The rheological structure of the slab is complex with weak, fine-grained spinel in the cold core surrounded by stronger, coarse-grained spinel regions on both sides. The cooler slabs are essentially weaker, and rapid deformation of a weak, cool slab focuses stress onto the strong regions. In this model, shear instability can occur in the strong regions if strain rate is large; thus earthquakes could occur in the regions surrounding the weak fine-grained core. An instability parameter defines the regions subject to thermal shear instability (**Figure 13**). A 40–60-km-thick zone is available to host deep earthquakes below ~450 km depth.

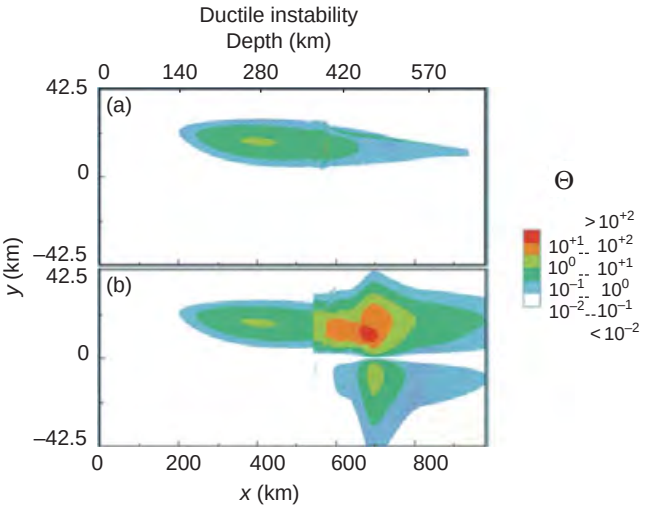


Figure 13 Portions of a slab subject to thermal shear instability. The assumed subduction rate is 10 cm yr^{-1}. Distribution of the instability parameter Θ (a) when latent heat release and grain-size reduction are ignored, and (b) when these effects are included (Karato *et al.*, 2001). Thermal shear instability (and earthquake nucleation) is predicted to occur where the instability parameter exceeds 1. Note that the upper horizontal scale shows depth, while the lower gives distance along the dipping slab. The vertical scale shows distance perpendicular to slab dip. The thickness of the seismogenic regions can be up to 40–60 km. A cold slab is predicted to have a weak fine-grained spinel core; Θ exceeds the critical value for instability in the high-strength regions surrounding the core. Thus, a deep double seismic zone of limited extent is possible. Adapted from Karato S, Riedel MR, and Yuen DA (2001) Rheological structure and deformation of subducted slabs in the mantle transition zone: Implications for mantle circulation and deep earthquakes. *Physics of the Earth and Planetary Interiors* 127: 83–108.

If displacements are large and rapid enough, the fault surface or shear zone may melt. Fault slip on seismic timescales will typically generate heat much faster than it can be diffused away. If enough heat is generated by sufficiently large displacements across a sufficiently thin shear zone, temperatures may rise to the melting point. The temperature rise is proportional to the frictional stress and inversely proportional to the thickness of the shear zone (Kanamori *et al.*, 1998). Estimates of temperature rise versus fault zone thickness for the M_w 8.2 1994 Bolivia earthquake suggest that melting would occur if the shear zone or slip zone was thinner than 0.3 m (Kanamori *et al.*, 1998); it is difficult to assess whether such a thin shear zone was attained, but the shear instability process is inherently self-localizing. The relationship between rupture velocity, radiated seismic energy, moment, and static stress drop for the Bolivian event (in particular, the very low radiation efficiency, discussed in Section 4.11.2.4.5) imply that a large amount of energy was dissipated in the shear zone, perhaps by melting (Kanamori *et al.*, 1998).

Shear localization in the shear instability mechanism is due to a progressive reduction in temperature-dependent viscosity due to heat generated by viscous shearing. If boundary conditions are suitable, for example, under conditions of constant stress, the viscosity reduction can lead to an increased rate of deformation that leads to greater reduction, and an instability can result. But, under typical conditions, strain may reduce the driving stress. Thus, it is not clear whether the actual boundary conditions in slabs (e.g., constant strain rate) can support a runaway positive feedback (Green and Marone, 2002). Thermal shear instability is known to occur in metals, but has not yet been definitively associated with earthquakes at any depth.

In the model of Karato *et al.* (2001), the abrupt cessation of seismicity near 670 km in slabs subducting below that depth may be explained by superplasticity of perovskite in the lower mantle (Ito and Sato, 1991). But, the low level of seismicity between ~250 and ~550 km is not well-explained by the shear instability mechanism alone. In fact, the instability parameter of the mechanism predicts that seismicity should be strongest between 350 and 550 km depth (**Figure 13**), whereas observed seismicity increases below ~550 km (**Figure 1**). The instability parameter and the stress field must be considered together, but it is not clear that shear instability can easily explain the low levels of seismicity between 250 and 550 km, nor the resurgence of seismicity between 550 and 650 km.

4.11.4 Geophysical Setting of Subducting Slabs

4.11.4.1 Thermal State of Subducting Slabs

The thermal structure of slabs has been studied extensively, but only a few results are cited here. Thermal structure is central to the deep earthquake problem because all three potential mechanisms discussed in Section 4.11.3 are controlled by slab temperature. Thermal control of the microphysics is exerted via dehydration rate, phase transformation rate, temperature-dependent rheology, or some combination of these. Kirby *et al.* (1996) related slab thermal structure to the depth of seismicity in the context of transformational faulting in metastable olivine. A more complete analysis of slab temperature by Devaux *et al.* (1997) included olivine–spinel reaction kinetics, latent heat release, and 2-D heat conduction (up slab dip). Including these factors tends to increase slab temperatures and to reduce the length of a possible wedge of metastable olivine. The presence of metastable olivine at a given depth depends on the reaction kinetics and the minimum temperature in the slab at that depth. The critical temperature below which metastable olivine can be present may be weakly pressure-dependent (Kirby *et al.*, 1991) and is usually estimated in the range 550–800°C. Devaux *et al.* (1997) find that a metastable olivine wedge can exist down to ~660 km in only the coldest slabs (see their table 3), but a few deep earthquakes occur as deep as ~660 km in apparently warmer slabs, such as South America. Mosenfelder *et al.* (2001) estimate somewhat higher temperatures and, thus, a shorter metastable wedge. However, the reaction kinetics are not well known. In sum, the viability of a metastable wedge through the entire upper mantle is uncertain.

4.11.4.1.1 Thermal parameter

Since the minimum temperature at a given depth in a slab is of interest for most plausible physical mechanisms of deep earthquakes, the thermal parameter was developed as a crude way to compare this aspect of thermal structure for different subduction zones (Kirby *et al.*, 1991). It is crude because it is based on a heat conduction model and does not include the effects of adiabatic heating or latent heats of transformation. Furthermore, the thermal parameter is less useful for studying the relation between intermediate-depth seismicity and dehydration. Much of the dehydration takes place at the top of the slab, and

temperatures there are not described well by the thermal parameter.

The thermal parameter ϕ is defined as the product of a subducting slab's age at entering the trench, the convergence rate perpendicular to the trench, and the sin of slab dip; globally, it ranges from ~50 to ~12 000 in units of km. ϕ is proportional to the inverse of minimum temperature in slabs at a given depth; thus, large ϕ signifies a relatively cool slab, and small ϕ, a warm one. Frohlich (2006b) recently developed an explicit relation of thermal parameter to minimum temperature, which does not include, however, the effects of adiabatic heating, latent heat, or 2-D heat conduction.

Subduction zones with $\phi > \sim 3000$ are cool enough to host deep earthquakes (Kirby *et al.*, 1991). This result was corroborated by Gorbatov and Kostglodov (1997), who studied the relation of maximum depth of seismicity to ϕ for 110 locations along 10 subduction zones. They find complex behavior (see their figure 13), with a near-linear dependence for $\phi < 2500$ and depths of ~50 to ~300 km, then the dependence of depth on ϕ weakens, then strengthens as maximum depth of seismicity increases rapidly from 300 to 600 km for just a small increase in ϕ. This sharp transition was also seen by Kirby *et al.* (1991), who argue that the reaction kinetics are such that a small decrease in slab temperature near $\phi \sim 3500$ results in a large change in the region of potential metastabilty. Finally, for ϕ higher than 6000 (i.e., the coolest slabs), the maximum depth of seismicity remains near 670 km for many subduction zones, raising the question why earthquakes do not continue into the lower mantle, which is addressed in Section 4.11.3. Gorbatov and Kostoglodov (1997) interpret the change in the ϕ-to-maximum-depth relation near 300 km as due to the olivine–spinel transformation, which has an equilibrium boundary near 350 km in subduction zones, in contrast to 410 km in ambient mantle. As latent heat of transformation is not included in ϕ, it may not represent temperatures well in that region, although if olivine remained metastable, ϕ would not change its relation to the minimum temperature.

In addition to the depth of the deepest seismicity, a number of earthquake properties seem related to ϕ. Wiens (2001) compared thermal parameter to magnitude–frequency relations (b-values) and aftershock occurrence rate (productivity) for events below 400 km. In cool slabs the ratio of large to small events is smaller, and mainshocks produce more aftershocks, in contrast to the situation in warm slabs; the magnitude–frequency result is consistent with the results of Giardini (1988). Wiens (2001) also found a systematic relation of ϕ with the rupture properties of seven very large events ($M_w \geq 7.6$) deeper than 400 km, including source duration, static stress drop, rupture velocity, and maximum seismic efficiency. This approach was extended by Tibi *et al.* (2003a) who report source properties for 14 large earthquakes ($M_w \geq 7.0$) below 400 km. All of the above-mentioned properties, except stress drop, show a systematic relationship with ϕ. Large earthquakes in cooler slabs have shorter scaled durations, higher rupture velocities, and greater radiation efficiencies. The dependence indicates the sensitivity of the deep earthquake rupture process to slab temperature, and was interpreted by Wiens (2001) and Tibi *et al.* (2003a) as support for the thermal shear instability mechanism. However, Persh and Houston (2004a, 2004b) compared ϕ to scaled duration and aftershock productivity for data sets of 111 and 290 earthquakes deeper than 100 km. The comparisons were made in three depth intervals (100–350 km, 350–550 km, and 550–700 km), but no clear relationship with ϕ was seen within the scatter. Perhaps the largest events (those on which Wiens (2001) and Tibi *et al.* (2003a) are based) have a more systematic relationship to temperature. Also, for a given slab, values of ϕ obtained in different studies can differ by up to several thousand, mainly due to different estimates of slab age, rate, and dip (Frohlich (2006a), figure 4.15). It would be useful to have modern, consistent, and comprehensive estimates of thermal parameter and, more importantly, slab thermal structure for all subduction zones, although uncertainty as to whether a metastable phase is present leads to significant uncertainty in temperature.

4.11.4.2 Stresses and Fault Plane Solutions

The stress within a vertical column of material under its own weight provides an analogy to stresses in a subducting slab (Isacks and Molnar, 1971; figure 5.4–11d of Stein and Wysession, 2003). In this simple thermomechanical model, tensional and/or compressional stresses are aligned along the dip of the slab. Down-dip compressional stresses dominate above the base of the upper mantle, if the slab is impeded by a higher-viscosity lower mantle; down-dip tensional stresses dominate at 100–200 km depth where the negatively buoyant slab essentially hangs from the surface.

Focal mechanisms from some subducting slabs are consistent with the thermomechanical model. Globally, mechanisms below 400 km are consistently down-dip compressional, while those above 300 km are often down-dip tensional, particularly in warmer slabs with no deep seismicity. Thus, this simple stress model has been invoked to explain the ubiquitous seismicity minimum around 250–450 km: as stress changes from down-dip tension just below the surface tectonic plates to down-dip compression above the base of the upper mantle, stress in the intervening region is expected to be low. However, a number of slabs in the Western Pacific deviate from the predictions of the thermomechanical model, and exhibit down-dip compression above the seismicity minimum (Green and Houston, 1995; Chen *et al.*, 2004). Furthermore, in detail, focal mechanisms reveal locally heterogeneous strain, with different types of mechanisms coexisting in the same region (Giardini, 1992). Thermal expansion and equilibrium phase transitions in a compositionally layered slab (a 'petrologic' model) can lead to a heterogeneous stress state (Bina, 1996; Guest *et al.*, 2003). Internal stresses due to phase-transition-related volumetric changes are typically an order of magnitude greater than those due to buoyancy or thermal expansion (Devaux *et al.*, 2000; Guest *et al.*, 2003). The presence of a metastable wedge leads to additional spatial complexity (Goto *et al.*, 1987; Yoshioka *et al.*, 1997; Devaux *et al.*, 2000). In their modeling of buoyancy-induced stress related to the olivine phase transitions, Yoshioka *et al.* (1997) find down-dip compression above 670 km, but below the metastable wedge, and maximum shear stresses up to 230 MPa along the wedge. Devaux *et al.* (2000) obtain differential stresses of ~1 GPa associated with a metastable wedge. Guest *et al.* (2003) obtain maximum shear stresses over 1 GPa due to equilibrium transitions with no metastability. Such high stresses are common in slab stress models; they are much higher than typical deep earthquake stress drops (1994 Bolivia excepted), which indicates that, in general, deep earthquake stress drops are not complete.

At intermediate depths, for the regions in which down-dip stresses dominate, Chen *et al.* (2004, see their figure 5) find that down-dip compressional mechanisms are prevalent in old and fast slabs, whereas down-dip tension is prevalent in young and slow slabs; that is, down-dip compression at intermediate depths tends to occur in the coolest and thus densest slabs, while the down-dip tension predicted by the above-mentioned thermomechanical model occurs in the less-dense slabs. This is inconsistent with the thermomechanical model, but may be consistent with a petrologic model. Chen *et al.* (2004) suggest that compressional stresses at intermediate depths may be generated in cold slabs by greater downwarping at the 670 km boundary, the potential presence of buoyant metastable olivine, or the inhibition of slab penetration at 670 km. However, it is not clear what is producing down-dip compression at intermediate depths in some subduction zones, as the results of Yoshioka *et al.* (1997) and Devaux *et al.* (2000), though complex, do not show this feature.

4.11.4.2.1 Focal mechanisms and reactivated faults

Focal mechanisms have been used to evaluate a specific mechanism by which dehydration embrittlement could generate large intermediate-to-deep earthquakes. A key issue involves how and to what depths slab mantle can become hydrated. One possibility involves fluid flow down the fault zones of large outer-rise normal faulting earthquakes prior to subduction. In that case the fluids in the pre-existing fault zones would be carried down with the subducting slab and released at depth to generate large intermediate or deep earthquakes. Jiao *et al.* (2000) analyze patterns of focal mechanism orientations in Tonga and Kuriles. Above 450 km they find (after correcting for the dip of the slab at depth) an asymmetric pattern of nodal planes with the asymmetry similar to that at the outer rise. The asymmetry occurs because, at the outer rise, the more steeply dipping nodal plane usually dips trenchward. At the Middle America trench, reflection data shows many normal faults, most of which dip trenchward (Ranero *et al.* 2003). Below 450 km, the distribution of nodal plane orientations becomes more complex. Thus, the results of Jiao *et al.* (2000) are consistent with earthquake generation above 450 km occurring on pre-existing faults formed at the outer rise.

Ranero *et al.* (2005) compare the strikes of bending faults at the outer rise to the strikes of intermediate-depth earthquakes in the Middle America and Chile subduction zones. They note that the strikes of the bending faults can vary significantly along the trenches, and the variation of fault strikes is very similar to that of the strikes of nodal planes of intermediate-depth earthquakes in the subduction slab below. The systematic similarity strongly suggests that intermediate seismicity in these zones occurs on pre-existing faults formed by plate bending at the trench.

In the Vanuatu subduction zone, however, stress inversions of intermediate-depth focal mechanisms indicate that new faults form if pre-existing ones are not oriented favorably for failure (Christova and Scholz, 2003). The authors find significant differences between the distribution of nodal planes for intermediate-depth earthquakes and that for outer-rise events. The results imply that most intermediate-depth earthquakes in the Vanuatu slab lie on faults newly created at depth; the pre-existing faults are only reactivated if they are oriented appropriately for the stress field. In the Tonga subduction zone, recent work by Warren *et al.* (2007) appears to argue against the occurrence of most intermediate-depth seismicity on pre-existing faults formed at the outer rise. Warren *et al.* (2007) measured differential rupture durations from teleseismic P waves to obtain constraints on source directivity, that allowed the fault plane to be identified in 30% of the 82 cases studied. Fault planes were identified for 9 of 25 intermediate-depth events. All nine were subhorizontal and inconsistent with the orientation of outer-rise normal faults. However, a systematic bias could be present, if subhorizontal fault planes are easier than near-vertical planes to resolve because of the geometry of teleseismic body waves.

4.11.5 Seismic Structures in Slabs

4.11.5.1 Double Seismic Zones

In several subduction zones, intermediate-depth seismicity occurs in two layers up to 40 km apart, vertically separated by an aseismic region (**Figure 14**). Such double seismic zones have been detected at intermediate depths thus far at local sites in Japan, Tonga, Kamchatka, Alaska, Chile, New Britain, New Zealand, Mexico, and Cascadia (e.g., Kawakatsu, 1986; Robinson, 1986; Abers, 1992; Hasegawa *et al.*, 1994; Gorbatov *et al.*, 1994; Comte and Suarez, 1994; Pardo and Suarez, 1995; Cassidy and Waldhauser, 2003; Rietbrock and Waldhauser, 2004). Often, uncertainty in earthquake depths complicates the detection of double seismic zones, so they may be much more common than currently realized. In the best-resolved double zones in northeastern Japan (Hasegawa *et al.*, 1994) and Kamchatka (Gorbatov *et al.*, 1994; Kao and Chen, 1994), the lower zone is seismically active between 30 and 180 km depth, below which it merges with the upper zone.

Focal mechanisms indicate that the majority of seismic double zones are consistent with down-dip compression in the upper zone and down-dip tension in the lower zone. This is consistent with a model of stresses produced by unbending of the slab as it straightens out below the overriding lithosphere, and this model has provided the conventional explanation of the stresses causing double zones (e.g., Isacks and Barazangi, 1977; Kawakatsu, 1986). But, double zone seismicity continues well after the slab has straightened out. Perhaps minor sagging could be involved. Furthermore, the double zones beneath Alaska (Abers, 1992) and New Zealand (Robinson, 1986) have both layers in down-dip tension, while for those beneath Mexico (Pardo and Suarez, 1995) and northern Chile (Comte and Suarez, 1994) the upper zone is in down-dip tension and the lower zone appears in down-dip compression. In northeastern Japan, Igarashi *et al.* (2001) detected a third seismic plane only 5–10 km above the upper down-dip compressional plane; this uppermost plane appears to lie at and just beneath the top surface of the descending slab. It consists of interplate low-angle thrust earthquakes near the top surface of the descending slab from 30 to 60 km depth, transitioning to intraplate normal faulting (down-dip tensional) mechanisms within the upper crust from 60 to 100 km depth. A similar three-plane zone may be present in Kamchatka, where Gorbatov *et al.* (1994) found a handful of events analogous to the down-dip tensional events in the uppermost (third) plane. Thus, there must be extensional stresses around the slab surface, which cannot be explained by large-scale plate unbending alone. Stress heterogeneities due to petrological layering of the slab may provide an explanation. Wang (2002) suggested that the extensional stress state of the uppermost (third) plane could be associated with the formation of eclogites, while the stress state of the other planes may be caused by unbending. Kao and Chen (1994) propose that the compression responsible for the lower plane of the Kamchatka double zone is transmitted to intermediate depths from the base of the upper mantle.

For many double zones, it is clear that the lower seismic zone must lie deep within the subducting mantle lithosphere, raising the question of how originally dry oceanic mantle lithosphere can become hydrated. Peacock (2001) suggested that normal faulting associated with bending at the outer rise and trench allows infiltration of water tens of kilometers into the oceanic lithosphere. Evidence for pervasive bending-related normal faulting that penetrates at least 20 km deep into the plate is seen at sites along the Middle America trench (Ranero

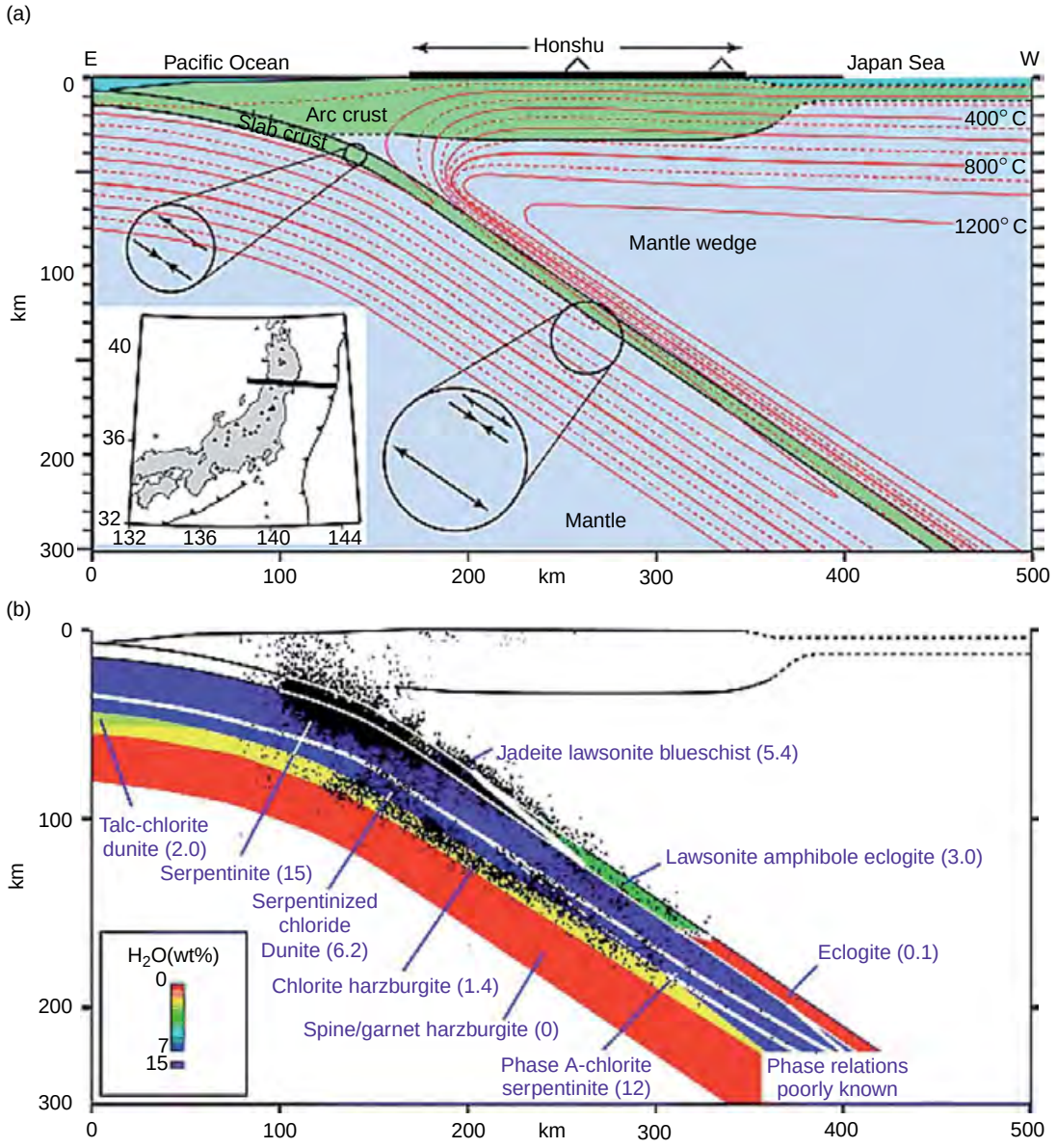


Figure 14 Cross-sections depicting relations between thermal structure, predicted petrology, and earthquake locations in the Tohoku, Japan subduction zone. (a) Thermal structure shown with solid isotherms every 200°. Offset circles show stress states inferred from Igarashi *et al.* (2001). (b) Comparison of petrology predicted by thermal structure with hypocentral locations. The lower zone of a well-developed double seismic zone follows the predicted locations of regions of serpentinite. Adapted from Hacker BR, Peacock SM, Abers GA, and Holloway SD (2003) Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *Journal of Geophysical Research* 108: 24627–24637 (doi:10.1029/2001JB001129).

et al., 2003). Near the trench, reflection data reveal about 1.5 faults per km with offsets of 100–1000 m. An estimate of the amount of water in the slab subducting at the Middle America trench suggests that the serpentinized mantle portion of the slab may hold as much chemically bound water as the slab's crustal portion (Ranero *et al.*, 2003). Furthermore, the patterns of orientation of such faulting are similar to those of intermediate-depth seismicity in the Middle America and Chile subduction zones (Ranero *et al.*, 2005), suggesting that intermediate seismicity occurs on pre-existing faults.

Hacker *et al.* (2003) compared the morphology of intermediate-depth seismicity for four subduction zones (Cascadia, Nankai, Costa Rica, and Tohoku) with the locations and water content of hydrous minerals, predicted based on slab age and thermal structure. The depth and geometry of the seismicity is consistent with the predicted distribution of several hydrous phases. In the coldest zone, Tohoku, double planes of seismicity surround the cold core of the slab: the zones approach one another, nearly following isotherms, but actually cutting across them at a shallow angle. Antigorite serpentinite, one of the most stable hydrous phases, is the main component in cooler slabs at intermediate depths, and is believed to dehydrate progressively down to about 200 km. Yamasaki and Seno (2003) obtained a similar result analyzing double seismic zones in six subduction zones. These studies provide strong evidence for dehydration embrittlement as the mechanism for intermediate earthquakes down to perhaps 250 km (Hacker *et al.*, 2003; Yamasaki and Seno, 2003). Double seismic zones appear easier to detect in cooler slabs, probably because the two layers are farther apart.

Double seismic zones have also been observed below 300 km depth. Wiens *et al.* (1993) relocated hypocenters with P, pP, and PKP arrivals and resolved a deep double seismic zone in Tonga from 350 to 460 km, with about 30 km separating the two layers. The stress state for the Tonga double zone is roughly opposite to the state typical at intermediate depths. Iidaka and Furukawa (1994) detected two planes about 25 km apart in the Izu Bonin subduction zone using S–P converted phases at the upper boundary of the slab. The double zone extends from about 300 to 450 km depth, that is, below the seismicity minimum, which in the Izu Bonin slab occurs in the anomalously shallow range of 200–300 km. The upper layer lies 15–20 km below the top surface of the slab. The double zone occurs in the same region that Iidaka and Suetsugu (1992) inferred a seismically slow core of the slab from traveltimes (see discussion in Green and Houston, 1995). Both these cases were interpreted as evidence for transformational faulting on the upper and lower edges of a metastable olivine wedge (Wiens *et al.*, 1993; Iidaka and Furukawa, 1994). A comparison of the modeled stresses due to a metastable wedge to the characteristics of the deep double zone in Tonga suggested that the lower seismic zone may reside well below a hypothesized metastable wedge, which would indicate the operation of two mechanisms of rupture (Guest *et al.*, 2004).

4.11.5.2 Planar Shear Zones

Planar seismicity structures have been detected below 400 km in several slabs. In local sites in the Tonga, Japan, Bonin, Marianas, Java, Banda, and South America slabs, seismicity is concentrated along large-scale planes, which align with a focal mechanism nodal plane, suggesting that large through-going shear zones host the seismicity (Giardini and Woodhouse, 1984; Lundgren and Giardini, 1992). These features are best developed in Tonga where they can extend over 200 km and cut through the seismically active portion of the slab (Giardini and Woodhouse, 1984). In some cases active conjugate shear zones intersect at ~90° (Lundgren and Giardini, 1992). The sense of deformation is related to slab contortion, buckling, and imbrication in Tonga (Giardini and Woodhouse, 1984), and in general reflects resistance to slab penetration at 670 km and flattening of the slab to a more horizontal position (Lundgren and Giardini, 1992). Tibi *et al.*'s (2001) relocations of 17 aftershocks of the 1995 *M*7 Mariana earthquake (586 km depth) span a plane about 20 km, across the entire thickness of the seismically active slab.

4.11.5.2.1 Repeated earthquakes

Recent more precise earthquake locations reveal structures similar to planar shear zones on a smaller scale. Wiens and Snider (2001) detected clusters of *M*4–5 earthquakes with identical waveforms, 560–600 km deep in the Tonga slab. Precise relocations using waveform cross-correlation yield location accuracies of 1–3 km, and reveal that the events locate on shear zones 10–30 km in length, and up to only 2–3 km thick. Three events are collocated within 1 km and their rupture zones appear to overlap, so they could represent repeated failure of same fault surface. The authors interpret the results in terms of runaway thermal shear instabilities, and suggest that earthquake doublets within a few hours represent occurrence of the second event before heat generated by first event can diffuse away (Wiens and Snider, 2001). However, location accuracies of 1–3 km, while a large improvement, are still insufficient to rule out transformational faulting, as the apparently collocated events could occur on parallel en echelon faults that cannot be spatially resolved.

4.11.5.3 Outboard Earthquakes and Recumbent Slabs

In several locations a few deep earthquakes have been observed well outside the main Wadati–Benioff zone

seismicity; these events, which occur farther from the trench than the dipping Wadati–Benioff zone, are termed 'outboard earthquakes'. They appear to indicate flattening or horizontal deflection of some deep slabs near the base of the lower mantle.

Lundgren and Giardini (1994) document several outboard earthquakes at about 600 km depth in the Izu Bonin, Kurile, and Chile subduction zones, which lie 150–300 km away from the main Wadati–Benioff zone. In the first two zones, Okal (2001) established the countinuity of the subducting lithosphere between the main seismic zone and the outboard event, consistent with horizontal deflection of the actively subducting slab. Lundgren and Giardini (1994) also describe more extensive subhorizontal distributions of seismicity in Banda and Tonga-Fiji, which appear to delineate nearly flat-lying slab remnants recumbent near the base of the transition zone in these regions. Seismicity in the latter region was scrutinized by Okal and Kirby (1998), who reached a similar conclusion. Deep seismicity near Tonga-Fiji has been recognized as highly spatially complex for decades, and has been interpreted as indicating a very contorted slab. However, Chen and Brudzinski (2001) thoroughly analyzed the geometry and focal mechanisms of seismicity in the Tonga-Fiji region, and resolved that two remnants of previously subducted lithosphere are detached from the currently subducting slab; portions of the remnants are nearly horizontal and up to 1000 km away from the actively subducting slab, while a portion lies above the actively subducting slab (**Figure 15**). The detached remnants may be associated with an episode of subduction along the fossil Vitiaz trench 5–8 Ma. Stress orientations from focal mechanisms of outboard earthquakes and the remnant slab seismicity are distinct from orientations in the main Wadati–Benioff zone, and are not those expected if the detached remnants were a stress guide. Their P-axes lie closer to the slab normal, whereas deep in a Wadati–Benioff zone seismicity is typically down-dip compressional.

Thus, rather than sinking undeformed into the lower mantle, in some subduction zones lithosphere is deflected, retained in the transition zone, and, most saliently, still capable of earthquakes. There must, therefore, exist a localized source of stress in the subhorizontal portions, rather than slab pull or resistance to penetration (Brudzinski and Chen, 2005). Based on their global analysis of patterns of strain in subhorizontal slabs at both intermediate depths and deeper, Brudzinski and Chen (2005) emphasize the need for locally variable sources of stress to generate seismicity in subhorizontal portions of slabs.

Interestingly, the subhorizontal swaths of seismicity that delineate slab remnants in the Tonga-Fiji region are not associated with a fast velocity anomaly such as is typical for rapidly subducting zones, but adjacent regions without subhorizontal seismicity

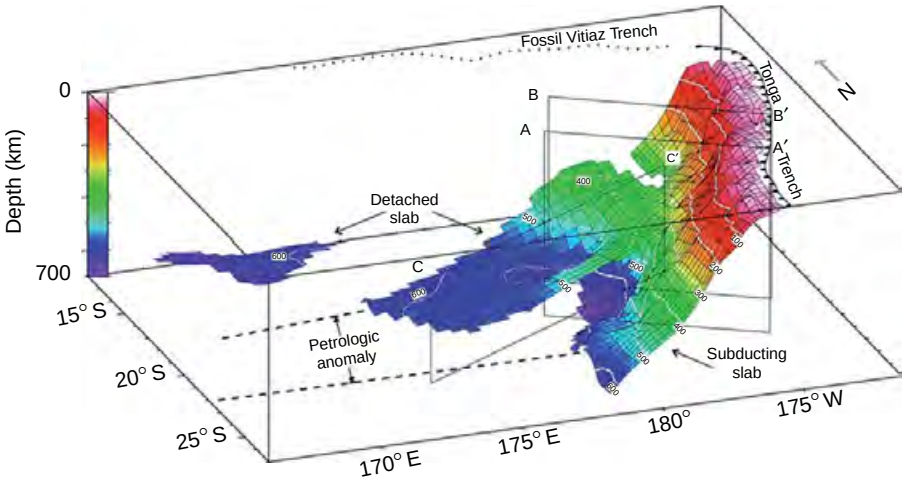


Figure 15 A three-dimensional view of the Tonga slab for the region between 15° S and 25° S. Slab contortions and detached remnants are apparent. Three surfaces were constructed from the hypocentral locations, representing the actively subducting lithosphere, a detached remnant of slab, and an isolated cluster of earthquakes northeast of the Vanuatu trench. Dashed lines indicate the region of an inferred petrologic anomaly, in which anomalously slow wavespeeds are detected. White contour lines (100 km in depth) are shown. Adapted from Chen W-P and Brudzinski MR (2001) Evidence for a large-scale remnant of subducted lithosphere beneath Fiji. *Science* 292: 2475–2479.

exhibit 2–3% faster P and S velocities (Chen and Brudzinski, 2001; Brudzinski and Chen, 2003). The line of demarcation is sharp (Brudzinski and Chen, 2003). The interpretation is that the relatively cold temperature of the slab remnants must have been counteracted by a petrologic anomaly, a compositional or mineralogical difference, such as the presence of metastable olivine. Indeed, the preferred explanation of Brudzinski and Chen (2003) is that metastable olivine is retained in the remnant slabs and generates their seismicity. This explanation is consistent with a number of observations (e.g., Brudzinski and Chen, 2003; Brudzinski and Chen, 2005). However, it contradicts available thermal constraints on olivine metastability (Devaux *et al.*, 1997; Mosenfelder *et al.*, 2001), because after 5–8 My in the transition zone, the remnant slabs must be relatively warm compared to the adjacent currently subducting lithosphere. Whether remnant slabs could contain metastable olivine remains an open question.

4.11.6 Implications for Seismogenesis

4.11.6.1 Viability of the Three Main Physical Mechanisms

Here the author outlines key observations and inferences consistent and inconsistent with the proposed mechanisms of deep earthquake rupture discussed in Section 4.11.3.

Dehydration embrittlement provides a good explanation of the frequency and location of intermediate-depth seismicity from ~50 km down to ~250 km. The exponential decrease with depth in earthquake frequency from 50 to 350 km (**Figure 1**) may be indicative of the gradual exhaustion of water as metamorphic reactions proceed. The morphology of intermediate-depth double seismic zones is consistent with the dewatering of hydrous phases predicted from subduction zone thermal structures (Hacker *et al.*, 2003; Yamasaki and Seno, 2003). Antigorite serpentinite is the most common constituent of subducting lithosphere at intermediate depth in cold slabs and its behavior is well known (Rayleigh and Paterson, 1965). Dehydration embrittlement appears less likely below 250 km, because hydrous phases, if present, probably retain rather than release water as they transform (Green and Marone, 2002).

A plausible mechanism has been put forward for pervasive hydration along bend-related faults at the outer rise (Ranero *et al.*, 2003, 2005); perhaps the hydration is sufficiently pervasive in some cases to allow formation of new faults with different orientations (Christova and Scholz, 2003). Slab unbending does not provide an entirely satisfactory explanation for the stresses driving intermediate-depth seismicity, due to the variety and complexity of stress states observed in double zones. Indeed, Brudzinski and Chen (2005) suggest that patterns of strain in subhorizontal slabs at intermediate depth and deeper indicate the need for locally variable sources of stress.

The micromechanics of transformational faulting is analogous to that in dehydration embrittlement. Transformational faulting has been well-demonstrated in the laboratory in analog materials (e.g., Green and Burnley, 1989; Kirby *et al.*, 1991), though less well in natural silicates. In concert with dehydration embrittlement operating above ~350 km, transformational faulting in metastable olivine below 350 km would explain seismicity's bimodal depth distribution and cessation at 670 km (Green and Zhou, 1996), and has been a conventional and popular explanation (Green and Houston, 1995). However, recent thermal modeling suggests that only the coldest slabs can host metastable olivine down to the base of the upper mantle (Devaux *et al.*, 1997; Mosenfelder *et al.*, 2001), although warmer slabs such as South America exhibit seismicity nearly as deep as that in Tonga. Unfortunately, the relevant reaction kinetics are not well known. The possibility of transformational faulting in other metastable phases, such as enstatite, should not be overlooked (Hogrefe *et al.*, 1994). Metastable olivine provides an explanation for the seismic properties of an apparent remnant slab in the upper mantle beneath Tonga (Chen and Brudzinski, 2001).

One of the main arguments advanced against transformational faulting in metastable olivine is that several recent very large and deep earthquakes exhibit fault planes that appear too large to fit into the expected narrow wedge of cold metastable material (e.g., Silver *et al.*, 1995; McGuire *et al.*, 1997). Furthermore, aftershocks nucleate in a wide zone outside the region of background seismicity (McGuire *et al.*, 1997). This argument, although strong, may not be completely definitive in view of the difficulty resolving the extent of rupture (e.g., compare geometry of rupture planes determined by Kikuchi and Kanamori (1994), Chen (1995), and Ihmle (1998)) as well as the possibility that slabs may thicken near 670 km due to resistance at the base of the mantle (Kirby *et al.*, 1996).

Models of runaway thermal shear instability may be consistent with large, very deep fault planes, although for warmer slabs such as South America a large bending moment must be applied (Karato *et al.*, 2001). However, the shear instability mechanism by itself does not explain the depth distribution of seismicity well, particularly for intermediate-depth earthquakes. Modeling (Karato *et al.*, 2001) indicates that the shear instability mechanism may begin to operate at about 200 km depth. The instability parameter grows with depth, and appears strongest between 350 and 550 km depth (**Figure 13**). In contrast, actual seismicity decreases exponentially from 50 to 350 km, is minimal from 350 to 500 km, then resurges between 500 and 650 km (**Figure 1**). Deviatoric stresses in the slab must be considered in conjunction with the instability parameter; they are sufficiently complex even for the equilibrium case (e.g., Yoshioka *et al.*, 1997; Guest *et al.*, 2003) that it is possible that they could produce the resurgence of seismicity below 500 km.

It is not clear what would stop an earthquake generated by shear instability. Faulting may be expected to run until driving stress falls below the viscous resistance, which must have plummeted to allow the instability. Static stress drops of most deep earthquakes are well below modeled deviatoric stresses in most cases, with 1994 Bolivia providing the main exception. Finally, the boundary conditions must be appropriate to accelerate shear and drive the instability.

Any of the above mechanisms might ultimately result in melting on the fault plane if displacement is large enough, but a progression to melting is most natural via thermal shear instability. Kanamori *et al.* (1998) argue that melting occurred in the 1994 Bolivian earthquake based on the large amount of energy apparently dissipated in the fault zone, as evidenced by the energy budget and the low rupture velocity. On the other hand, it seems odd that the Bolivian earthquake's radiated energy-to-moment ratio and scaled duration are similar to those of other deep and shallow earthquakes. The event's energy budget differs from that of other earthquakes primarily due to its high static stress drop, which is related to its low rupture velocity.

4.11.6.2 Notable Points and Speculations

It is clear that deep earthquakes are primarily double-couple (shearing) sources and trigger fewer aftershocks than shallow earthquakes. Variations in some rupture processes with depth are detectable in the scattered data. Aftershock productivity, scaled duration, and source complexity vary with depth. Although many rupture properties of deep earthquakes are not distinctly different from shallow quakes, it is likely that rupture velocity varies more for deep than for shallow earthquakes. For shallow faulting, rupture velocity usually falls between 0.6 and 0.9 of the shear-wave velocity, whereas for deep earthquakes it varies more widely from 0.2 to near 1.0 of the shear-wave velocity.

What stops earthquake rupture is a longstanding question for deep and shallow events alike. The end of rupture is hard to resolve in seismic waves due to noise, but stacking has increased resolution. In some cases deep earthquake rupture seems to stop abruptly (see figure 1b of Vidale and Houston, 1993). It is not obvious how observations of abrupt termination can be reconciled with mechanisms like shear instability and melting, which appear to call for a gradual decay of moment-release rate and a complete stress drop. Stress drops, while intrinsically variable and difficult to determine, appear to be much lower than many estimates of deviatoric stresses at depth in subducted lithosphere.

A comparison of very deep South American seismicity with that in Tonga suggests that (below 400 km) conditions promoting the nucleation of deep rupture differ from those promoting its propagation. The subduction zone with the most earthquakes, Tonga, is not the site of the largest ruptures, which are hosted in the relatively warm South American subduction zone. This contrast is reflected in the different relation between frequency and magnitude (i.e., *b*-value) for these two zones, with other subduction zones that possess deep seismicity falling intermediate between the two end members (Giardini, 1988; Wiens, 2001). Notably, the contrast is not observed for intermediate-depth earthquakes. Thus, in a way not seen for intermediate-depth events, the propensity to nucleate or propagate rupture of events below 400 km appears controlled by the thermal setting, as are some rupture properties (Wiens, 2001; Tibi *et al.*, 2003a).

The largest deep events occur in isolated seismically quiet regions with anomalously few aftershocks. They have low rupture velocities, a relatively weak initial phase, and a tendency to rupture sideways (Estabrook, 1999). Their isolation leads to a prediction that over the decades, more very deep giant earthquakes are likely to strike in nearby currently quiet zones, including regions north of the 1970

Colombia earthquake. These observations raise the possibility that different mechanisms may operate to generate deep earthquakes in different zones (Wiens and McGuire, 1995), or different mechanisms, with different dependencies on pressure–temperature conditions, may be important during different phases of rupture (e.g., nucleation vs propagation).

More generally, the possibility that multiple mechanisms generate deep earthquakes must be considered. Which of the mechanisms is operative could depend on depth, subduction zone (e.g., slab age, convergence rate, dip), or earthquake size. The author has already mentioned the conventional model that dehydration embrittlement is responsible for intermediate-depth events, and transformational faulting in metastable olivine for events below 350 km (Green and Houston, 1995; Kirby *et al.*, 1996). Persh and Houston (2004b) interpret the dependency on depth seen in some source properties as suggestive of a change in mechanism near 550 km, in addition to the change around 350 km. In the region between ~350 and 550 km, scaled duration, complexity, and aftershock productivity are anomalously long, high, and low, respectively, and the changes at 550 km appear abrupt (Persh and Houston, 2004a, 2004b). Possible mechanisms below 550 km include shear instability, melting, and transformational faulting in metastable enstatite. This scenario circumvents the apparent difficulty preserving metastable olivine below 550 km for most slabs (e.g., Devaux *et al.*, 1997), as well as the difficulty connected with rupture planes larger than the expected thickness of metastable olivine (e.g., Silver *et al.*, 1995; McGuire *et al.*, 1997). On the other hand, the contrast in behavior of the deepest seismicity between Tonga and South America suggests the possibility that transformational faulting in metastable olivine operates in Tonga below 350 km, while a different mechanism operates deep in the warmer South America slab – for example, shear instability or transformational faulting in metastable enstatite. The complete lack of seismicity between about 325 and 500 km depth along the length of the South America slab is striking compared to other zones. Finally, it is also plausible that in very large, deep earthquakes, rupture could initiate by one mechanism but progress to another (e.g., shear instability or melting) after the initial displacement generated sufficient heat. This would result in a variation in the mix of physical mechanisms for different-sized earthquakes. The energy budget and waveform shape suggest that this may have occurred for the 1994 Bolivian earthquake, and, if so, probably for the other South American very large deep earthquakes (Kanamori *et al.*, 1998; Estabrook, 1999).

Other combinations are possible, for example, in Tonga, shear instability, and in South America, transformational faulting in metastable enstatite, followed by shear instability and melting for the largest events. The combinations mentioned here were chosen so that rupture initiation would occur more readily in Tonga, and propagation more readily in South America, due to the effects of different thermal structures or concentrations of a metastable phase; that is, transformational faulting might initiate less readily in metastable enstatite than in metastable olivine, because the enstatite–illmentite component comprises only ~25% of the slab mineralogy, whereas the olivine–spinel component comprises the bulk of the slab mineralogy. Similarly, warmer temperatures favor enhanced propagation via shear instability or melting (once slip is initiated).

At this time no single physical mechanism appears capable of generating all seismicity below 100 km, or even below 200 km, and explaining observations of earthquake depth distribution, stress orientations, *b*-values, aftershock productivity, and rupture processes. In particular, dehydration embrittlement cannot readily explain both the contrast between subduction zones in magnitude–frequency relations for events below 350 km discussed above and the lack of contrast for events above 350 km, nor both the depth distribution and stress state. Given the subduction zone complexity that is emerging from recent studies, perhaps diversity in the mechanisms of rupture generation should be expected; perhaps it is also suggested by the variation in rupture velocities. Subducting slabs are inherently complex in mineralogy, rheology, and stress due to their petrological layering. The contribution of minor mineral components to the complex rheology and stress may be significant, and the mineralogy of subducting lithosphere and crust appears poorly known in detail. Rheology may be controlled by factors beyond thermal structure or the minimum temperature at a given depth; that is, a old, but slow, slab may have a different rheology than a young, but fast, slab, even if they are predicted to have the same minimum temperature at a given depth. Furthermore, the varied deformation history of subducted slabs, for example, the degree of bending at the trench or shear in the mantle, may affect their rheology.

Pre-existing faults likely play a very important role in intermediate-depth seismicity (Ranero *et al.*, 2003, 2005), but if they are not oriented sufficiently

favorably, new faults can form (Christova and Scholz, 2003). Below 400 km seismicity organized in planar structures appear to delineate fault zones or their physical equivalents at 400–700 km depth (e.g., Giardini and Woodhouse, 1984; Lundgren and Giardini, 1992). The occurrence of deep shear zones that persist as stable features of a deep subducted slab suggests that pre-existing faults play a role in deep seismicity, although they need not have formed at Earth's surface. These structures are not predicted by transformational faulting, and depending on their geometry they could be inconsistent with that mechanism. Their existence is evidence of a type of heterogeneity not often considered in modeling the rheology and stress of deep subduction zones.

4.11.7 Concluding Thoughts

Although evidence has accumulated over the past decade that intermediate-depth earthquake is likely generated by dehydration embrittlement, in my view, deeper earthquakes will not give up their secrets easily. Considering that the physical processes that occur during shallow rupture are still not fully understood, it is not entirely surprising that the more remote phenomena remain ambiguous. Progress has been slowed by the variability of earthquake source properties, by the poorly known mineral physics due to the difficulty of achieving in the laboratory conditions at great depth, and perhaps most of all by the lack of robust links between earthquakes source properties and physical mechanisms. Models connecting properties of the source to possible mechanisms lack a firm basis and are not yet well developed. Integration of seismological, laboratory, and modeling efforts is needed to bridge the stubborn gap between source properties, which are extracted under strong assumptions and possess substantial intrinsic variability, and physical mechanisms of rupture generation, which are as yet neither well constrained nor constrained. Heterogeneity in slab structure, stress, and rheology must be considered. Comparative studies of many events in many subduction zones will help overcome scatter and clarify systematic relationships. It would be helpful to have a base of systematic up-to-date fundamental information, such as thermal structures (perhaps versions with and without metastability) for all subduction zones using known tectonics. More precise earthquake locations and focal mechanisms will be valuable. An organized effort to provide these on a global basis would probably drive significant progress.

References

Abers GA (1992) Relationship between shallow- and intermediate-depth seismicity in the eastern Aleutian subduction zone. *Geophysical Research Letters* 19: 2019–2022.

Angel RJ, Frost DJ, Ross NL, and Hemley R (2001) Stabilities and equations of state of dense hydrous magnesium silicates. *Physics of the Earth and Planetary Interiors* 127: 181–196.

Antolik M, Dreger D, and Romanowicz B (1999) Rupture processes of large deep-focus earthquakes from inversion of moment–rate functions. *Journal of Geophysical Research* 104: 863–894.

Beck SL, Silver P, Wallace TC, and James D (1995) Directivity analysis of the deep Bolivia earthquake of June 9, 1994. *Geophysical Research Letters* 22: 2257–2260.

Bina CR (1996) Phase transition buoyancy contributions to stresses in subducting lithosphere. *Geophysical Research Letters* 23: 3563–3566.

Blanco MJ and Spakman W (1993) The P-wave velocity structure of the mantle below the Iberian Peninsula: Evidence for a subducted lithosphere below southern Spain. *Tectonophysics* 221: 13–34.

Brace WF and Kohlstedt DL (1980) Limits on lithospheric stress imposed by laboratory experiments. *Journal of Geophysical Research* 85: 6248–6252.

Brudzinski MR and Chen W-P (2003) A petrologic anomaly accompanying outboard earthquakes beneath Fiji–Tonga: Corresponding evidence from broadband P and S waveforms. *Journal of Geophysical Research* 108 (doi:10.1029/2002JB002012).

Brudzinski MR and Chen W-P (2005) Earthquakes and strain in subhorizontal slabs. *Journal of Geophysical Research* 110 (doi:10.1029/2004JB003470).

Brustle W and Muller G (1987) Stopping phases in seismograms and the spatiotemporal extent of earthquakes. *Bulletin of the Seismological Society of America* 77: 47–68.

Cassidy JF and Waldhauser F (2003) Evidence for both crustal and mantle earthquakes in the subducting Juan de Fuca plate. *Geophysical Research Letters* 30 (doi:10.1029/2002GL015511).

Chen P-F, Bina CR, and Okal EA (2004) A global survey of stress orientations in subducting slabs as revealed by intermediate-depth earthquakes. *Geophysical Journal International* 159: 721–733 (doi:10.1111/j.1365-246x.2004.02450.x).

Chen W-P (1995) En echelon ruptures during the great Bolivian earthquake of 1994. *Geophysical Research Letters* 22: 2261–2264.

Chen W-P and Brudzinski MR (2001) Evidence for a large-scale remnant of subducted lithosphere beneath Fiji. *Science* 292: 2475–2479.

Choy GL, McGarr A, Kirby SH, and Boatwright J (2006) In: Abercrombie RE, McGarr A, Kanamori H, and Di Toro G (eds.) *An Overview of the Global Variability in Radiated Energy and Apparent Stress, Vol. 170, Earthquakes: Radiated Energy and the Physics of Faulting*, pp. 43–58. Washington, DC: American Geophysical Union.

Christova C and Scholz CH (2003) Stresses in the Vanuatu subducting slab: A test of two hypotheses. *Geophysical Research Letters* 30(15): 1790, doi: 10.1029/2003GL017701.

Comte D and Suárez G (1994) An inverted double seismic zone in Chile: Evidence of phase transformation in the subducted slab. *Science* 263: 212–215.

Davies JH (1999) The role of hydraulic fractures and intermediate depth earthquakes in subduction zone magmatism. *Nature* 398: 142–145.

Devaux JP, Fleitout L, Schubert G, and Anderson C (2000) Stresses in a subducting slab in the presence of a metastable olivine wedge. *Journal of Geophysical Research* 105: 13365–13373.

Devaux JP, Schubert G, and Anderson C (1997) Formation of a metastable olivine wedge in a descending slab. *Journal of Geophysical Research* 102: 24627–24637.

Dixon JE, Leist L, Langmuir C, and Schilling JG (2002) Recycled dehydrated lithosphere observed in plume-influenced mid-ocean-ridge basalt. *Nature* 420: 385–389.

Dziewonski AM and Gilbert F (1974) Temporal variation of the seismic moment tensor and the evidence of precursive compression for two deep earthquakes. *Nature* 257: 185–188.

Estabrook CH (1999) Body wave inversion of the 1970 and 1963 South American large deep-focus earthquakes. *Journal of Geophysical Research* 104: 28751–28767.

Estabrook CH and Bock G (1995) Rupture history of the great Bolivian earthquake: Slab interaction with the 660-km discontinuity?. *Geophysical Research Letters* 22: 2277–2280.

Fletcher RC and Pollard DD (1981) Anti-crack model for pressure solution surfaces. *Geology* 9: 419–424.

Frohlich C (1987) Aftershocks and temporal cluster of deep earthquakes. *Journal of Geophysical Research* 92: 13944–13957.

Frohlich C (1989) The nature of deep-focus earthquakes. *Annual Review of Earth and Planetary Sciences* 17: 227–254.

Frohlich C and Davis SD (1993) Teleseismic b-values – Or, much ado about 1.0. *Journal of Geophysical Research* 98: 631–644.

Frohlich C (1994) Earthquakes with non-double-couple focal mechanisms. *Science* 264: 804–809.

Frohlich C (1995) Characteristics of well-determined non-double-couple earthquakes in the Harvard CMT catalog. *Physics of the Earth and Planetary Interiors* 91: 213–228.

Frohlich C (2006a) *Deep Earthquakes*. Cambridge, UK: Cambridge University Press.

Frohlich C (2006b) A simple analytical method to calcuate the thermal parameter and temperature within subducted lithosphere. *Physics of the Earth and Planetary Interiors* 155: 281–285.

Fukao Y (1972) Source process of a large deep earthquake and its tectonic implications: The western Brazil earthquake of 1963. *Physics of the Earth and Planetary Interiors* 5: 61–76.

Fukao Y and Kikuchi M (1987) Source retrieval for mantle earthquakes by iterative deconvolution of long-period P waves. *Tectonophysics* 144: 249–269.

Furumoto M (1977) Spacio-temporal history of the deep Colombian earthquake of 1970. *Physics of the Earth and Planetary Interiors* 15: 1–12.

Giardini D (1988) Frequency distribution and quantification of deep earthquakes. *Journal of Geophysical Research* 93: 2095–2105.

Giardini D (1992) Space–time distribution of deep seismic deformation in Tonga. *Physics of the Earth and Planetary Interiors* 74: 75–88.

Giardini D and Lundgren P (1995) The June 9 Bolivia and March 9 Fiji deep earthquakes of 1994: II. Geodynamic implications. *Geophysical Research Letters* 22: 2281–2284.

Giardini D and Woodhouse JH (1984) Deep seismicity and modes of deformation in Tonga subduction zone. *Nature* 307: 505–509 (doi:10.1038/307505a0).

Glennon MA and Chen W-P (1995) Ruptures of deep-focus earthquakes in the northwestern pacific and their implications on seismogenesis. *Geophysical Journal International* 120: 706–720.

Gorbatov A, Suarez G, Kostoglodov V, and Gordeev E (1994) A double-planed seismic zone in Kamchatka from local and teleseismic data. *Geophysical Research Letters* 21: 1675–1678.

Gorbatov A and Kostoglodov V (1997) Maximum depth of seismicity and thermal parameter of the subducting slab: General empirical relation and its application. *Tectonophysics* 277: 165–187.

Goto K, Suzuki Z, and Hamaguchi H (1987) Stress distribution due to olivine–spinel phase transformation in descending plate and deep focus earthquakes. *Journal of Geophysical Research* 92: 13811–13820.

Green HW and Burnley PC (1989) A new self-organizing mechanism for deep-focus earthquakes. *Nature* 341: 733–737.

Green HW and Houston H (1995) The mechanics of deep earthquakes. *Annual Review of Earth and Planetary Sciences* 23: 169–213.

Green HW and Marone C (2002) Instability of deformation. In: Korato S and Wenk H (eds.) *Reviews in Mineralogy and Geochemistry, Plasticity of minerals and Rocks*, Vol. 51: pp. 181–1999. Oxford, UK: Blackwell.

Green HW and Zhou Y (1996) Transformation-induced faulting requires an exothermic reaction and explains the cessation of earthquakes at the base of the mantle transition zone. *Tectonophysics* 256: 39–56.

Griggs D and Handin J (1960) Observations on fracture and a hypothesis of earthquakes. *Geological Society of America Memoir* 79: 347–373.

Guest A, Schubert G, and Gable CW (2004) Stresses along the metastable wedge of olivine in a subducting slab: Possible explanation for the Tonga double seismic layer. *Physics of the Earth and Planetary Interiors* 141: 253–267.

Guest A, Schubert G, and Gable CW (2003) Stress field in the subducting lithosphere and comparison with deep earthquakes in Tonga. *Journal of Geophysical Research* 108: B6 (doi:10.1029/2002JB002161).

Hacker BR, Peacock SM, Abers GA, and Holloway SD (2003) Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions?. *Journal of Geophysical Research* 108: 24627–24637 (doi:10.1029/2001JB001129).

Hara T, Kuge K, and Kawakatsu H (1995) Determination of the isotropic component of the 1994 Bolivia deep earthquake. *Geophysical Research Letters* 22: 2265–2268.

Hara T, Kuge K, and Kawakatsu H (1996) Determination of the isotropic component of deep focus earthquakes by inversion of normal-mode data. *Geophysical Journal International* 127: 515–528.

Hasegawa A, Horiuchi S, and Umino N (1994) Seismic structure of the northeastern Japan convergent margin: A synthesis. *Physics of the Earth and Planetary Interiors* 99(B11): 22295–22312.

Helffrich G and Brodholt J (1991) Relationship of deep seismicity to the thermal structure of subducted lithosphere. *Nature* 353: 252–255.

Hill DP, Reasenberg PA, Michael A, *et al.* (1993) Seismicity remotely triggered by the magnitude 7.3 Landers California Earthquake. *Science* 260: 1617–1623.

Hobbs BE and Ord A (1988) Plastic instabilities: Implications for the origin of intermediate and deep focus earthquakes. *Journal of Geophysical Research* 93: 10521–10540.

Hogrefe A, Rubie DC, Sharp TG, and Seifert F (1994) Metastability of enstatite in deep subducting lithosphere. *Nature* 372: 351–353.

Houston H and Williams Q (1991) Fast rise times and the physical mechanism of deep earthquakes. *Nature* 352: 520–522.

Houston H and Vidale JE (1994) Temporal distribution of seismic radiation during deep earthquake rupture. *Science* 265: 771–774.

Houston H, Benz HM, and Vidale JE (1998) Time functions of deep earthquakes from broadband and short-

period stacks. *Journal of Geophysical Research* 103: 29895–29913.
Houston H (2001) Influence of depth, focal mechanism, and tectonic setting on the shape and duration of earthquake source time functions. *Journal of Geophysical Research* 106: 11137–11150.
Huang W-C and Okal EA (1998) Centroid moment tensor solutions for deep earthquakes predating the digital era: Discussion and inferences. *Physics of the Earth and Planetary Interiors* 106: 191–218 (doi:10.1016/S0031-9201(97)00111-8).
Igarashi T, Matsuzawa T, Umino N, and Hasegawa A (2001) Spatial distribution of focal mechanisms for interplate and intraplate earthquakes associated with the subducting Pacific plate beneath the northeastern Japan arc: A tripled-planed deep seismic zone. *Journal of Geophysical Research* 106: 2177–2191.
Ihmle PF (1998) On the interpretation of subevents in teleseismic waveforms: The 1994 Bolivia deep earthquake revisited. *Journal of Geophysical Research* 103: 17919–17932.
Iidaka T and Suetsugu D (1992) Seismological evidence for metastable olivine inside a subducting slab. *Nature* 356: 593–595 (doi:10.1038/356593a0).
Iidaka T and Furukawa Y (1994) Double seismic zone for deep earthquakes in the Izu–Bonin subduction zone. *Science* 263(5150): 1116–1118.
Isacks B and Molnar P (1971) Distribution of stresses in the descending lithosphere from a global survey of focal mechanism solutions of mantle earthquakes. *Reviews of Geophysics* 9: 103–174.
Isacks BL and Barazangi M (1977) Geometry of Benioff zones: Lateral segmentation and downwards bending of the subducted lithosphere. In: Talwani M and Pitman WCIII (eds.) *Island Arcs, Deep Sea Trenches and Back-Arc Basins*, pp. 99–114. Washington, DC: AGU.
Ito E and Sato H (1991) Aseismicity in the lower mantle by superplasticity of the descending slab. *Nature* 351: 140–141.
Jiao WJ, Silver PG, Fei YW, and Prewitt CT (2000) Do intermediate- and deep-focus earthquakes occur on preexisting weak zones? An examination of the Tonga subduction zone. *Journal of Geophysical Research* 105(B12): 28125–28138.
Kagan YY (1999) Universality of the seismic moment–frequency relation. *Pure and Applied Geophysics* 155: 537–573.
Kanamori H and Anderson DL (1975) Theoretical basis of some empirical relations in seismology. *Bulletin of the Seismological Society of America* 65: 1073–1095.
Kanamori H, Anderson DL, and Heaton TH (1998) Frictional melting during the rupture of the 1994 Bolivian earthquake. *Science* 279: 839–842.
Kao H and Chen W-P (1994) The double seismic zone in Kurile–Kamchatka – The tale of two overlapping zones. *Journal of Geophysical Research* 99: 6913–6930.
Karato S, Riedel MR, and Yuen DA (2001) Rheological structure and deformation of subducted slabs in the mantle transition zone: Implications for mantle circulation and deep earthquakes. *Physics of the Earth and Planetary Interiors* 127: 83–108.
Kawakatsu H (1986) Downdip tensional earthquakes beneath the Tonga arc: A double seismic zone. *Journal of Geophysical Research* 91: 6432–6440.
Kawakatsu H (1996) Observability of the isotropic component of a moment tensor. *Geophysical Journal International* 126: 525–544.
Kikuchi M and Kanamori H (1994) The mechanism of the deep Bolivia earthquake of June 9, 1994. *Geophysical Research Letters* 21: 2341–2344.
Kirby SH, Durham WB, and Stern L (1991) Mantle phase changes and deep earthquake faulting in subducting lithosphere. *Science* 252: 216–225.
Kirby SH, Stein S, Okal EA, and Rubie DC (1996) Metastable mantle phase transformations and deep earthquakes in subducting oceanic lithosphere. *Reviews of Geophysics and Space Physics* 34: 261–306.
Kohlstedt DL, Keppler H, and Rubie DC (1996) Solubility of water in the α, β and γ-phases of (Mg, Fe)$_2$SiO$_4$. *Contributions to Mineralogy and Petrology* 123: 345–357.
Koper KD and Wiens DA (2000) The waveguide effect of metastable olivine in slabs. *Geophysical Research Letters* 27: 581–584.
Kuge K (1994) Rapid rupture and complex faulting of the May 12, 1990, Sakhalin deep earthquake – Analysis of regional and teleseismic broad-band data. *Journal of Geophysical Research* 99: 2671–2685.
Lockner DA, Byerlee JD, Kuksenko V, Ponomarev A, and Sidorin A (1991) Quasi-static fault growth and shear fracture energy in granite. *Nature* 350: 39–42.
Lundgren PR and Giardini D (1992) Seismicity, shear failure and modes of deformation in deep subduction zones. *Physics of the Earth and Planetary Interiors* 74: 63–74.
Lundgren P and Giardini D (1994) Seismicity, shear failure and modes of deformation in deep subduction zones. *Journal of Geophysical Research* 99: 15833–15842.
Lundgren P and Giardini D (1995) The June 9 Bolivia and March 9 Fiji deep earthquakes of 1994. Part I: Source processes. *Geophysical Research Letters* 22: 2241–2244.
McGuire JJ, Wiens D, Shore P, and Bevis M (1997) The March 9, 1994 (M_w 7.6) deep Tonga earthquake: Rupture outside the seismically active slab. *Geophysical Research Letters* 102: 15163–15182.
Mosenfelder JL, Marton FC, Ross CR, Kerschhofer L, and Rubie DC (2001) Experimental constraints on the depth of olivine metastability in subducting lithosphere. *Physics of the Earth and Planetary Interiors* 127(1–4): 165–180.
Morris J and Tera F (1989) 10 Be and 9 Be in mineral separtates and whole rocks from volcanic arcs: Implications for sediment subduction. *Geochemica et Cosmochimica Acta* 53: 3197–3206.
Myers SC, Wallace TC, Beck SL, *et al.* (1995) Implications of spatial and temporal development of the aftershock sequence for the M_w 8.3 June 9, 1994 deep Bolivian earthquake. *Geophysical Research Letters* 22: 2269–2272.
Nyffenegger P and Frohlich C (2000) Aftershock occurrence rate decay properties for intermediate and deep earthquake sequences. *Geophysical Research Letters* 27: 1215–1218.
Ogawa M (1987) Shear instability in a viscoelastic material as the cause for deep focus earthquakes. *Journal of Geophysical Research* 92: 13801–13810.
Okal EA (2001) 'Detached' deep earthquakes: Are they really? *Physics of the Earth and Planetary Interiors* 127: 109–143.
Okal EA and Kirby SH (1998) Deep earthquakes beneath the Fiji basin, SW Pacific: Earth's most intense deep seismicity in stagnant slabs. *Physics of the Earth and Planetary Interiors* 109: 25–63.
Omori S, Kamiya S, Maruyama S, and Zhao D (2001) Morphology of the intraslab seismic zone and devolatilization phase equilibria of the subducting slab peridotite. *Bulletin of the Earthquake Research Institute* 76: 455–478.
Pardo M and Suarez G (1995) Shape of the subducted Rivera and Cocos plates in southern Mexico: Seismic and tectonic implications. *Journal of Geophysical Research* 100: 12357–12373.
Peacock SM (2001) Are the lower planes of double seismic zones caused by serpentine dehydration in subducting oceanic mantle? *Geology* 29: 299–302.

Persh SE and Houston H (2004a) Deep earthquake rupture histories determined by global stacking of broadband P waveforms. *Journal of Geophysical Research* 109 (doi:10.1029/2003JB002762).
Persh SE and Houston H (2004b) Strongly depth-dependent aftershock production in deep earthquakes. *Bulletin of the Seismological Society of America* 94: 1808–1816.
Petit JP and Barquins M (1988) Can natural faults propagate under Mode II conditions. *Tectonics* 7: 1243–1256.
Preston LA, Creager KC, Crosson RS, Brocher TM, and Trehu AM (2003) Intraslab earthquakes: Dehydration of the Cascadia slab. *Science* 302: 1197–1200.
Ranero CR, Phipps-Morgan J, McIntosh K, and Reichert C (2003) Bending-related faulting and mantle serpentinization at the Middle America trench. *Nature* 425: 367–373.
Ranero CR, Villasenor A, Morgan JP, and Weinrebe W (2005) Relationship between bend–faulting at trenches and intermediate-depth seismicity. *Geochemistry Geophysics Geosystems* 6: Q12002.
Rayleigh CB and Paterson MS (1965) Experimental deformation of serpentinite and its tectonic implications. *Journal of Geophysical Research* 70: 3965–3985.
Rietbrock A and Waldhauser F (2004) A narrowly spaced double-seismic zone in the subducting Nazca plate. *Geophysical Research Letters* 31 (doi:10.1029/2004GL019610).
Robinson R (1986) Seismicity, structure and tectonics of the Wellington region, New Zealand. *Geophysical Journal of the Royal Astronomical Society* 87: 379–409.
Russakoff D, Ekstrom G, and Tromp J (1997) A new analysis of the 1970 Columbia earthquake and its isotropic component. *Journal of Geophysical Research* 102: 20423–20434.
Scholz CH (2002) *The Mechanics of Earthquakes and Faulting*. Cambridge UK: Cambridge University Press.
Silver PG, Beck SL, Wallace TC, *et al.* (1995) Rupture characteristics of the deep Bolivian earthquake of June 9, 1994 and the mechanism of deep-focus earthquakes. *Science* 268: 69–73.
Stein S and Wysession M (2003) *An Introduction to Seismology, Earthquakes, and Earth Structure*. Oxford: Blackwell Publishing Limited.
Thompson AB (1992) Water in the Earth's upper mantle. *Nature* 358: 295–302.
Tibi R, Estabrook CH, and Bock G (1999) The 1996 June 17 Flores Sea and 1994 March 9 Fiji–Tonga earthquakes: Source processes and deep earthquake mechanisms. *Geophysical Journal International* 138(3): 625–642 (doi:10.1046/j.1365-246x.1999.00879.x).
Tibi R, Wiens DA, and Hildebrand JA (2001) Aftershock locations and rupture characteristics of the 1995 Mariana deep earthquake. *Geophysical Research Letters* 28: 4311–4314.
Tibi R, Bock G, and Estabrook CH (2002) Seismic body wave constraint on mechanisms of intermediate-depth earthquakes. *Journal of Geophysical Research* 107 (doi:10.1029/2001JB000361).
Tibi R, Bock B, and Wiens DA (2003a) Source characteristics of large deep earthquakes: Constraint on the faulting mechanism at great depths. *Journal of Geophysical Research* 108 (doi:10.1029/2002JB001948).
Tibi R, Wiens DA, and Inoue H (2003b) Remote triggering of deep earthquakes in the 2002 Tonga sequence. *Nature* 424: 921–925.
Tocheport A, Rivera L, and Chevrot S (2006) A systematic study of source time functions and moment tensors of intermediate and deep earthquakes. *Journal of Geophysical Research* submitted.
Venkataraman A and Kanamori H (2004) Observational constraints on the fracture energy of subduction zone earthquakes. *Journal of Geophysical Research* 109 (doi:10.1029/2003JB002549).
Vidale JE and Benz HM (1992) Upper-mantle seismic discontinuities and the thermal structure of subduction zones. *Nature* 356: 678–683.
Vidale JE and Houston H (1993) The depth dependence of earthquake duration and implications for rupture mechanisms. *Nature* 365: 45–47.
Vidale JE, Williams Q, and Houston H (1991) Waveform effects of a metastable olivine tongue in subducting slabs. *Geophysical Research Letters* 18: 2201–2204.
Wang K (2002) Unbending combined with dehydration embrittlement as a cause for double and triple seismic zones. *Geophysical Research Letters* 29: 1889–1892 (doi:10.1029/2002GL015441).
Warren LM, Hughes AN, and Silver PG (2007) Earthquake mechanics and deformation in the Tonga–Kermadec subduction zone from fault-plane orientations of intermediate- and deep-focus earthquakes. *Journal of Geophysical Research* (in press).
Wiens DA, McGuire JJ, and Shore PJ (1993) Evidence for transformational faulting from a deep double seismic zone in Tonga. *Nature* 364(6440): 790–793.
Wiens DA and Gilbert HJ (1996) Effect of slab temperature on deep-earthquakes aftershock productivity and magnitude–frequency relations. *Nature* 384: 153–156.
Wiens DA and McGuire JJ (1995) The 1994 Bolivia and Tonga events–Fundamentally different types of deep earthquakes. *Geophysical Research Letters* 22: 2245–2248.
Wiens DA and McGuire J (2000) Aftershocks of the March 9, 1994 Tonga earthquake: The strongest known deep aftershock sequence. *Journal of Geophysical Research* 105: 19067–19083.
Wiens DA (2001) Seismological constraints on the mechanisms of deep earthquakes: Temperature dependence of deep earthquake source properties. *Physics of the Earth and Planetary Interiors* 127: 145–163.
Wiens DA and Snider NO (2001) Repeating deep earthquakes: Evidence for fault reactivation at great depth. *Science* 293(5534): 1463–1466.
Winslow NW and Ruff LJ (1999) A hybrid method for calculating the radiated wave energy of deep earthquakes. *Physics of the Earth and Planetary Interiors* 115: 181–190.
Wu L-R and Chen W-P (2001) Rupture of the large (M_w 7.8), deep earthquake of 1973 beneath the Japan Sea with implications for seismogenesis. *Bulletin of the Seismological Society of America* 91: 102–111.
Yamasaki T and Seno T (2003) Double seismic zone and dehydration embrittlement of the subducting slab. *Journal of Geophysical Research* 108 (doi:10.1029/2002JB001918).
Yoshioka S, Daessler R, and Yuen DA (1997) Stress fields associated with metastable phase transitions in descending slabs and deep-focus earthquakes. *Physics of the Earth and Planetary Interiors* 104: 345–361.

Relevant Website

http://www.sionizconter.ucsd.edu – SIO Vizualization Center at Scripps Institute of Oceanography, Institute for Geophysics and Planetary Physics (this website will give you full directions on how to download and use the iView3D software).

4.12 Volcanology 101 for Seismologists

C. G. Newhall, US Geological Survey (emeritus), Sto. Domingo, Albay, Philippines

Glossary

dike A tabular igneous intrusion that cuts across planar structures of surrounding rock, often cutting vertically through horizontal strata.

eruption Extrusion or explosive ejection of volcanic material onto the Earth's surface or into the atmosphere – (1) phreatic eruption: steam blast, without new magma; (2) magmatic eruption: eruption purely of magma; and (3) phreatomagmatic eruption: hybrid of two. Many eruptions progress from phreatic through phreatomagmatic to magmatic, and some revert back to phreatomagmatic in their late stages as groundwater regains access to magma.

felsic Adjective describing minerals and rocks rich in silica and aluminum and poor in iron and magnesium. Felsic rocks include dacite, rhyolite, their intrusive equivalents of granodiorite and granite, and a variety of less common rock types with higher or lower contents of Na and K.

groundmass Fine-grained mixture of glass and microlites, forming the interstitial matrix of volcanic rocks containing phenocrysts.

mafic Adjective describing minerals and rocks rich in iron and magnesium, and relatively poor in silica and aluminum. Mafic rocks include basalt, its intrusive equivalent of gabbro, and a variety of less common rock types with higher or lower contents of Na and K. Andesite and its intrusive equivalent, diorite, with slightly higher silica content, are intermediate between mafic and felsic.

magma Molten rock, consisting of up to three phases – melt, crystals, and exsolved volatiles (bubbles).

magma mixing (and mingling) Mixing is mechanical and chemical homogenization of two magmas. Mingling is incomplete mixing and is seen in 'marble-cake' banded rocks.

microlite Very small crystal in volcanic rock, typically <<1 mm in longest dimension, formed in volcanic conduits upon degassing or cooling during or shortly before eruption.

oxygen fugacity (abbreviated f_{O_2}) A measure of the oxidizing or reducing character of a fluid. More precisely, the chemical activity of oxygen in a chemical solution, that is, the amount of oxygen available for chemical reactions. Typically expressed in log units, absolute or relative to the amount of free oxygen associated with an assemblage of minerals, for example, QFM, quartz (SiO_2) + fayalite (Fe_2SiO_2) + magnetite (Fe_3O_4).

petrologic method Estimation of pre-eruption volatile contents from the highest volatile concentrations trapped in glass inclusions within phenocrysts, and of degassing based on the difference between volatile concentrations in glass inclusions minus concentrations in matrix (= groundmass) glass. A variation of the petrologic method approximates the concentration of volatiles in magma as 100 wt.% minus the analytical total of nonvolatile components.

phenocryst Crystal in igneous rock that is notably larger than the average grain size of that rock, probably because it formed over extended periods within magma reservoirs several kilometers or more beneath the surface. No specific size definition, but almost always >1 mm in length.

pumice Rapidly quenched ('frozen') magma foam. Low-density, full of vesicles (bubble holes). Density often <1 g cm^{-3} so it floats in water, in contrast to nonvesicular rock densities of approx 2.5 g cm^{-2}.

scoria ('cinder') Volcanic rock with high vesicularity but less than that of pumice. Although there is a tendency for scoria to be mafic and pumice felsic, composition is not formally a part of the definition of pumice or scoria.
second boiling Process whereby crystallization concentrates volatiles in the remaining melt phase and eventually causes them to exsolve into bubbles, mimicking a boiling process even though temperature may be constant or decreasing.
sill A tabular igneous intrusion that parallels planar structures of surrounding rock, such as horizontal strata.
ultramafic rocks Rocks containing low silica contents and crystallizing large percentages of mafic minerals, for example, olivine, pyroxenes, and amphiboles. Includes peridotite, eclogite, and a variety of less common combinations of mafic minerals.
unrest An anomalous change from the background level of seismicity, ground deformation, gas emission, thermal output, and other measured parameters of a volcano. Some unrest leads to eruptions but much does not, so we use the term 'unrest' rather than 'precursor' unless the temporal or causal association with an eruption is clear. Eruptions are included within the definition of unrest.
volcano A vent or cluster of vents from which molten or solid debris is erupted. Most volcanoes build edifices from erupted lava and fragmental debris; some that erupt very large volumes of magma create depressions in the ground, as in the case of calderas.

4.12.1 Introduction to Volcanic Systems

Active volcanoes are dynamic systems that can be instructively compared to active fault zones. Both are subject to stress and strain, both produce distinctive patterns of seismicity and, in a sense, both exhibit stick-slip behavior with similar power-law magnitude–frequency relations. The distribution of areas at risk from active volcanoes and fault zones can be mapped and future hazards forecast over decades, centuries, or longer. However, beyond the similarities, volcanic systems involve high-temperature silicate melts, significant volumes of dissolved and undissolved volatiles, and important mass transfer before eruptions. Volcanic eruptions have far more pronounced precursors than do tectonic earthquakes, and, accordingly, in addition to long-term forecasts, both hour-day and week-month scale forecasts are often possible. Some of the precursors to eruptions are similar to those sought for earthquakes (e.g., precursory creep and microseismicity) while others are unique on account of the magma (e.g., leaks of high temperature gas).

Some volcanoes are sensitive to static or dynamic tectonic strain, for example, Long Valley and Yellowstone calderas (Hill *et al.*, 2002), and, conversely, magmatic intrusions and hydrothermal pressurization can induce nearby tectonic seismicity (White and Power, 2001; Legrand *et al.*, 2002).

This chapter describes the fundamental matter of volcanoes – molten rock (magma), which is composed of melt (typically silicate melt), crystals (dominantly, feldspars, quartz, and Fe–Mg silicates), and volatiles (mostly water). The volatile species are dissolved or occur as bubbles in the melt. Although magma is generated in the upper mantle and lower crust, this chapter is concerned mainly with its ascent from relatively shallow crustal reservoirs to the surface, and associated pressurization of magma and surrounding groundwater. Volcanic unrest – volcanic seismicity, ground deformation, gas emission, and other changes – reflects intrusion, pressurization, and pre-eruption degassing of magma, and forms the basis for short-term eruption forecasts. Finally, the chapter notes briefly different eruptive phenomena, posteruption erosion of volcanoes, and future directions in volcanology.

Volcanic eruptions range from a few m^3 to thousands of km^3 of magma, minutes to decades in duration, and gently effusive to astoundingly explosive. One important control on the explosivity of an eruption appears to be the 'leakiness' of the system, that is, how much of its magma's gas has leaked out since the previous eruption and before soon-to-erupt magma reaches the surface.

Worldwide, more than 500 million people live within zones that could be seriously affected by volcanic eruptions (Tilling, 1991; Ewert and Harpel, 2004). Millions more fly over or downwind

from volcanoes every year. The average annual death toll from volcanic activity between AD 1800 and 2000 was $\sim$1000 yr^{-1}. That average is dominated by just a few particularly sad disasters, such as the eruptions of Montagne Pelée in 1902 and of Nevado del Ruiz in 1985. During the past four decades, roughly 1.5 million people on the ground have been evacuated out of harm's way, and hundreds of thousands of air passengers, possibly more, have been diverted away from dangerous ash clouds. Populations living on and near volcanoes are ever increasing, and volcanologists today are faced with rising expectations for accurate and timely eruption forecasts to keep these populations safe.

Although a review, this chapter shows some biases on the way volcanoes work. Volcanologists including the author are strongly influenced by the volcanoes we have known. This chapter should be considered as a set of working hypotheses, judged likely today but subject to revision tomorrow. What follows is by no means comprehensive. Sigurdsson *et al.* (2000), Francis and Oppenheimer (2004), and Schmincke (2004) are excellent references on many more aspects of volcanology, including volcanic rocks and features, and Dobran (2001) gives mathematical descriptions of many volcanic processes.

4.12.2 Magma and Gas

4.12.2.1 Melt

Molten silicate, or in rare cases molten carbonate, is generated in the upper mantle beneath spreading centers and hot spots, in mantle wedges of arcs, and within the crust. Most mantle melts are basaltic in composition, commonly distinguished by their variable alkali (Na and K) contents. They originate by variable degrees (usually <20%) partial melting of mantle peridotite. Crustal melts tend to be much more silicic, rhyolite or dacite, produced by <20% and often <10% partial melting of (mostly lower) crust or, in some cases, by fractionation of mantle-derived basalt. The percentage of partial melt present in a source region at any given time is considerably less, probably not more than a few percent. Mafic melts can have temperatures up to 1200°C; rare ultramafic melts can be up to 1400°C. Silicic melts can exist down to about 700°C.

The term magma (molten rock) encompasses not only the molten silicate but also any crystals that have formed or remain from incomplete melting, and volatiles that are either dissolved within the silicate melt or occur as bubbles in that melt (**Figure 1**).

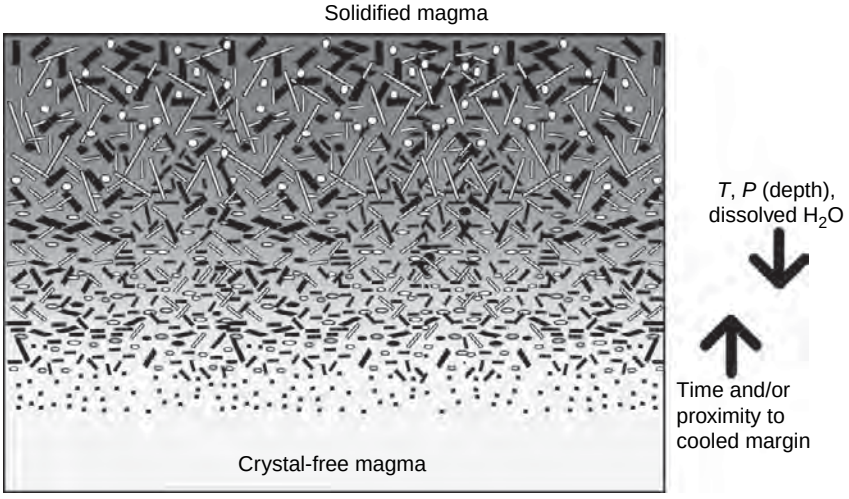


Figure 1 Components and evolution of magma, from melt to solid. Initially, at high temperature and pressure, only melt is present. As temperature declines, crystals will begin to form. As lithostatic pressure declines, water and other volatiles will also begin to exsolved into bubbles, inducing further crystallization. The sequence shown here from bottom to top, symbolizing change as magma moves from depth to the surface, is also repeated from center to margin of dikes and other intrusions. After Marsh BD (2000) Magma chambers. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 191–206. San Diego, CA: Academic Press.

Magmas can further differentiate by fractional crystallization in which early crystallizing phases and remaining melt are separated by crystal-settling or/or by filter-pressing melt out of crystal-rich mush, by assimilation (digestion) of country rock, and by magma mixing.

4.12.2.2 Crystals

Crystals form in melt as it cools or as it loses dissolved volatiles. Decreasing temperature and loss of dissolved water lead to polymerization of silica tetrahedra (SiO_4) in the melt that are the building blocks of common volcanic minerals like quartz (SiO_2), plagioclase feldspars (Ca/Na–Al silicates), alkali feldspars (Na–K–Al silicate), olivine (Mg/Fe silicate), pyroxenes (Ca/Mg/Fe silicates), hornblende and other amphiboles (hydrous Ca/Na/K–Mg/Fe/Al silicates), biotite and other micas (hydrous K/Na–Mg/Fe/Al/Ti silicates) (slash indicates solid solution between these end members, for example, between Ca and Na in plagioclase feldspar). Other important volcanic minerals, small in volume but very useful in tracking petrologic processes, are the Fe–Ti oxide minerals which can be used to determine the temperature and oxygen fugacity of magma.

Crystals occur in two broad size groups, phenocrysts (typically, 1–10 mm) and groundmass crystals (microlites, typically <<1 mm). Phenocrysts form while magma is still in its reservoir, whereas microlites grow during final ascent and eruption of magma.

Volcanic rocks rarely contain more than 60% phenocrysts (Marsh, 1981). The reason is that higher phenocryst contents make magmas too viscous to erupt. The balance, 40%, is melt in the reservoir but can grow a significant percentage of microlites shortly before and during eruptions.

4.12.2.3 Volatile Phases – Both Dissolved and Exsolved

4.12.2.3.1 Gas supply

Volatile components of magma, dominantly H_2O, CO_2, sulfur gases SO_2 and H_2S, halogen gases HCI and HF, hydrogen, and helium in decreasing order of abundance, are incorporated into magma during initial melting of mantle or crust. Water content in magma can range from <0.1 wt.% up to ~8 wt.%. Carbon dioxide content can be as high as 1 wt.% (10 000 ppm) but is usually a fraction of this. Sulfur concentrations (expressed as elemental S) are typically 10–1000 ppm, and occasionally several times higher. Beneath arcs, additional water, and other volatiles come from subducted sediment and from metamorphism of downgoing, hydrothermally altered oceanic crust to eclogite. That water is released upward into the mantle wedge, where it fluxes melting and gets incorporated into magmas.

4.12.2.3.2 Degassing

There are limits to volatile solubility in silicate melts. The limits depend mostly strongly on pressure; the composition, temperature, and oxygen fugacity of the silicate melt; and the specific volatile species in question. Volatiles will form bubbles if they are added to or concentrated in melt in excess of their solubility. One way to add volatiles to magma is by convective and diffusive transfer from fresh magma entering the base of a magma reservoir. Another very common scenario of volatile concentration is called 'second-boiling', in which volatiles, excluded from most crystal structures, are concentrated by magma crystallization into the remaining melt fraction, and eventually exceed their solubilities in that melt. Temperature may be constant or even falling, yet volatiles are 'boiling', exsolving from the melt into bubbles.

Bubbles will be of lower density than their host melt and can rise buoyantly through that melt. Large bubbles (sometimes, up to several meters in diameter) are seen to rise up through fluid magma in Hawaii, at Mount Erebus in Antarctica, and elsewhere. This process ('separated flow') is an effective way to degas fluid magmas (Vergniolle and Jaupart, 1986). An everyday analogy is the bottle of soda pop or champagne from which bubbles rise freely once the confining (lithostatic) pressure is released. Eventually, the soda pop and magma 'go flat' unless there is resupply of volatiles from below. Degassing in viscous magmas is not as well understood. Clearly gases do escape – we can measure them at the surface and we also know that silicic magmas can 'go flat'. However, in viscous magmas simple Stokes' law bubble rise is too slow to explain observed gas fluxes. Three alternate processes may be at play. One is convective transport of gas-rich magma up to a level at which it can foam, release its gas, and then sink in return flow as dense, now-degassed silicate melt and crystals. One can observe this process directly in lava lakes, and it is inferred at many other volcanoes for which the gas flux far exceeds what could have been dissolved in the volume of magma erupted over that same period (e.g., Allard, 1977; Allard *et al.*, 1994; Witter *et al.*, 2004; 2005). A second mechanism by which gas can escape from

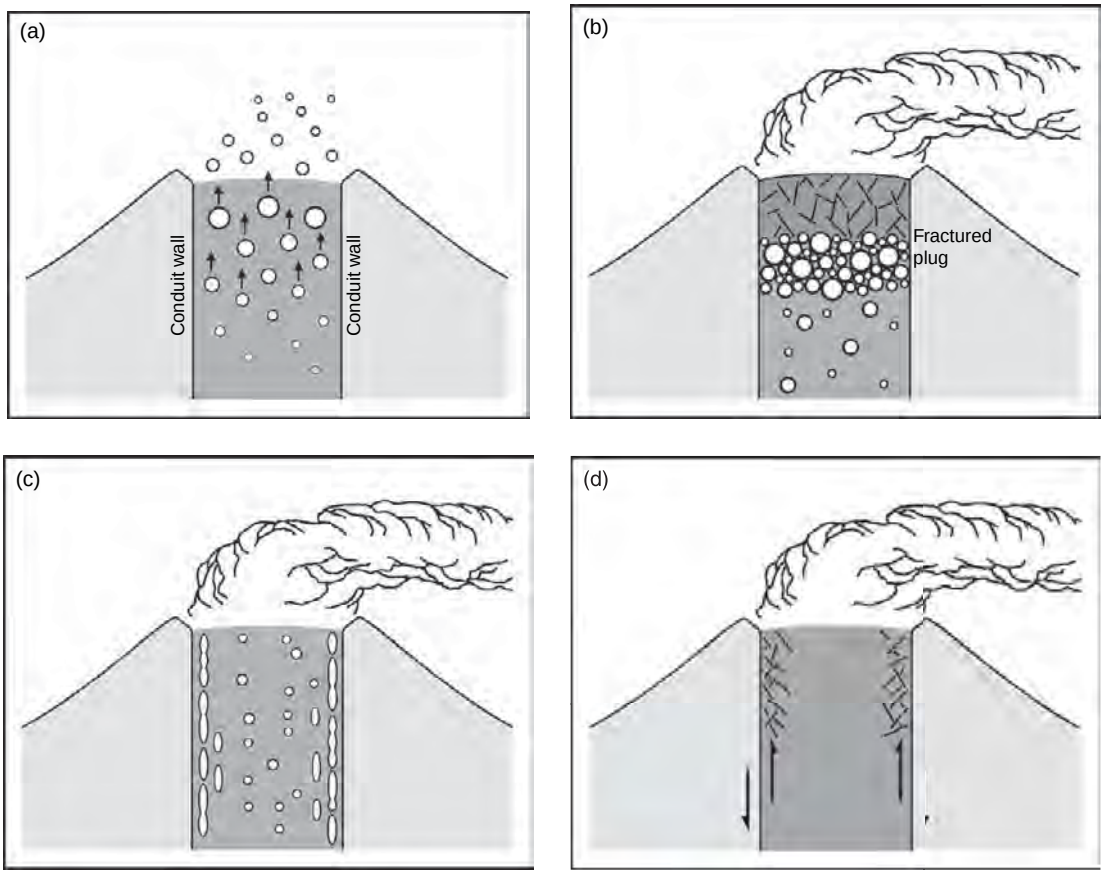


Figure 2 Various modes of gas permeability in magma: (a) separated flow of bubbles through melt; (b) development of a permeable foam at top of magma column; (c) dynamic shear-induced permeability (from bubbles and melt being stretched); and (d) shear-induced fractures along the margins of a volcanic conduit.

viscous magma is by diffusion or bubble transport over relatively short distances to the edge of magma conduits, followed by escape into high-permeability breccias that form along the margins of some conduits by shear (Jaupart and Allègre, 1991; Stasiuk *et al.*, 1996; Rust *et al.*, 2003; Burgisser and Gardner, 2005). Third, some dynamic bubble-to-bubble permeability may also develop within viscous magma as bubbles are brought into contact and interconnected by shear during magma ascent. Various modes of degassing are illustrated in **Figure 2**.

4.12.2.3.3 Gas budgets

The balance between (1) gas influx into crustal magma reservoirs (e.g., arriving dissolved in fresh basaltic magma), (2) continuous, passive degassing into the atmosphere and into hydrothermal systems, and (3) degassing during explosive eruptions is an important measure of the behavior of a volcano (**Figure 3**).

The gas influx term can be roughly estimated from a volume rate of magma supply (from the mantle and lower crust up into a shallow magma reservoir) times the concentration of each volatile species in that magma. A minimum estimate for volume rate of magma supply can come from the long-term eruption rate for that volcano; a better estimate that includes magma which does not erupt (intrusions) can come from long-term growth rate for a whole volcanic edifice or volcanic arc. If magma supply is relatively continuous, these long-term rates may also apply to shorter time intervals. Typical magma supply rates for arc volcanoes are $\sim 10^7$ $m^3\,yr^{-1}$, plus or minus an order of magnitude; rates for ocean island and hot-spot volcanoes can be $\sim 10^8$ $m^3\,yr^{-1}$ plus or minus.

If one measures gases that are not common in the atmosphere, for example, SO_2, one can quantify passive degassing of magma between times of eruptions. Unfortunately, passive gas flux has been estimated

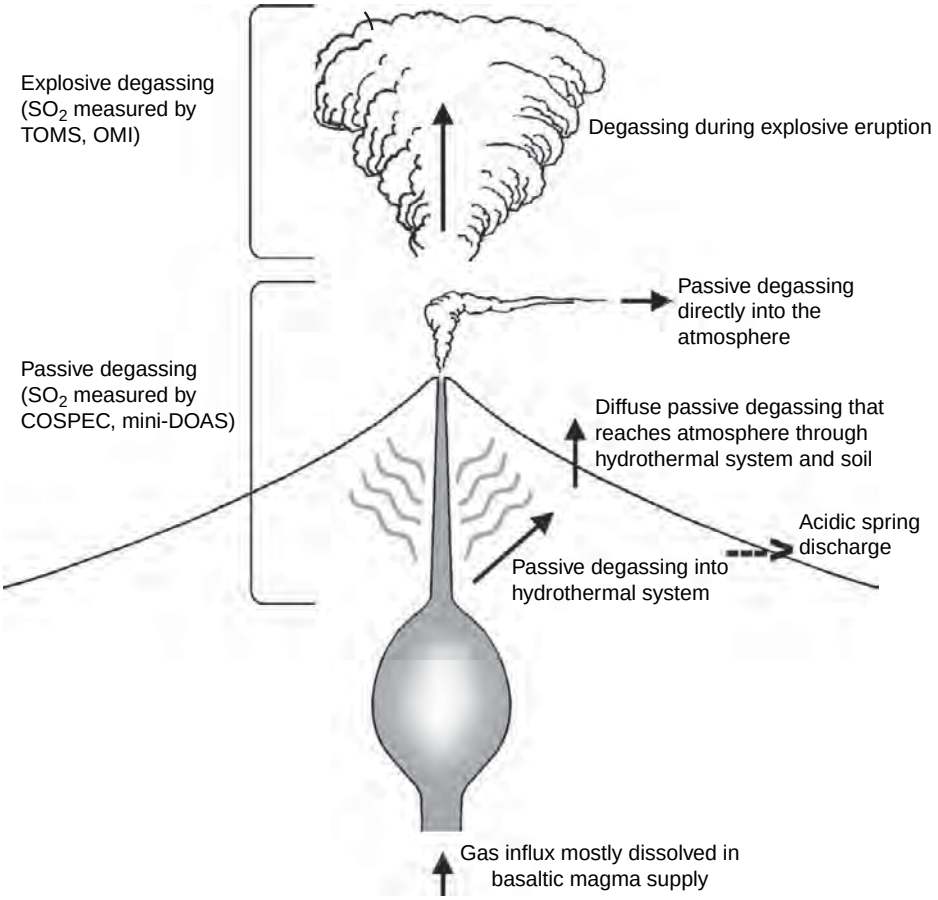


Figure 3 Elements of a simple volcanic gas budget, including input from depth, loss into a hydrothermal system, acidic spring discharge, passive loss to the atmosphere, and explosive loss to the atmosphere.

consistently at only a handful of volcanoes. Elsewhere, measurements have been too infrequent to give reliable estimates. A promising development for systematic, space-based monitoring of passive degassing of SO_2, down to levels of hundreds of tonnes/day, is the new ozone mapping instrument (OMI) which began collecting data in 2004 (Krotkov *et al.*, 2003).

A final term in a gas budget is the mass of gas released during eruptions. Again using SO_2 (and estimating other species in relation to SO_2), the best data come from NASA's total ozone mapping spectrometer (TOMS). Data from the largest eruptions have been gathered since 1977 (Bluth *et al.*, 1997; Carn *et al.*, 2003). OMI data will supplement those from TOMS.

Of the three main terms in a gas budget – input, passive degassing, and eruptive degassing – the input term is the most poorly known. As a result, it is difficult to know the degree to which magma has already degassed before eruption. Yet, pre-eruption degassing may be the key to estimating explosive potential of an impending eruption.

One important insight that is appearing from gas budgets is that most volcanoes lose most of their gas passively, between eruptions. Degassing during explosive eruptions is spectacular, but less noticed, steady leakage dominates discharge. If we do a thought exercise by assuming concentrations of 1000 and 2500 ppm S in 10^7 m^3 yr^{-1} of fresh mafic magma supply to typically explosive arc volcanoes, multiply by the number of years since the last gas-depleting VEI ≥ 3 explosive eruption, convert from S to SO_2, and compare to the SO_2 release of the next VEI ≥ 3 eruption (Carn *et al.*, 2003), we find that for most arc volcanoes explosive gas release is <10% of inferred gas input. The difference must be released passively. This has interesting implications for conduit dynamics between eruptions as noted in Section 4.12.3.3, and for explosive potential as noted in Section 4.12.6.5.2.

Volcanoes in Hawaii are classic examples of volcanoes that degas passively. Even within arcs, for the above-mentioned thought exercise, there are sharp contrasts between volcanoes that release almost all of their gas passively (e.g., Mayon, **Figure 4(a)**) and those that retain a substantial percentage of their incoming gas and then release it in large explosive eruptions (e.g., Pinatubo, **Figure 4(b)**) (Newhall, 2004).

Another important insight that has come from gas budgets is that some magmas appear to accumulate volatiles well in excess of saturation, with the excess volatiles accumulating in discrete fluid bubbles (Luhr *et al.*, 1984; Gerlach *et al.*, 1996; Keppler, 1999; Wallace and Anderson, 2000; Wallace 2001). Petrologic comparison of volatile contents in undegassed glass inclusions that were trapped long before eruptions in early formed phenocrysts, versus degassed matrix glass, gives a minimum measure of how much degassing has occurred (Anderson, 1974). For example, undegassed glass inclusions might contain a few hundred or even a few thousand ppm S, whereas matrix glass might contain only a few tens of ppm S; the difference in concentration is attributed to degassing before eruption and multiplying this concentration difference by the volume of magma erupted gives a minimum mass of pre-eruption degassing. When the amount of SO_2 released during large explosive eruptions (as measured by TOMS) is compared to that which would have been predicted from the petrologic method, the former may be 10–100× higher than the latter, and, indeed, much greater than could ever have been dissolved in the volume of magma erupted. Since this gas is erupted all at once it must be assumed to have resided within the magma, but it must have been in discrete fluid bubbles well in excess of saturation in the melt. (Using the petrologic method on products of magmas that had discrete pre-eruption volatile bubbles at depth, one often observes just a small difference between S in glass inclusions and S in the matrix glass, implying minimal pre-eruption degassing. In fact, though, a great deal of degassing from the melt already occurred at depth – 'into' the discrete bubble phase. As soon as CO_2-rich bubbles begin to form at depth, other volatiles including H_2O and SO_2 move preferentially into the volatile phase rather than remain dissolved in the melt.)

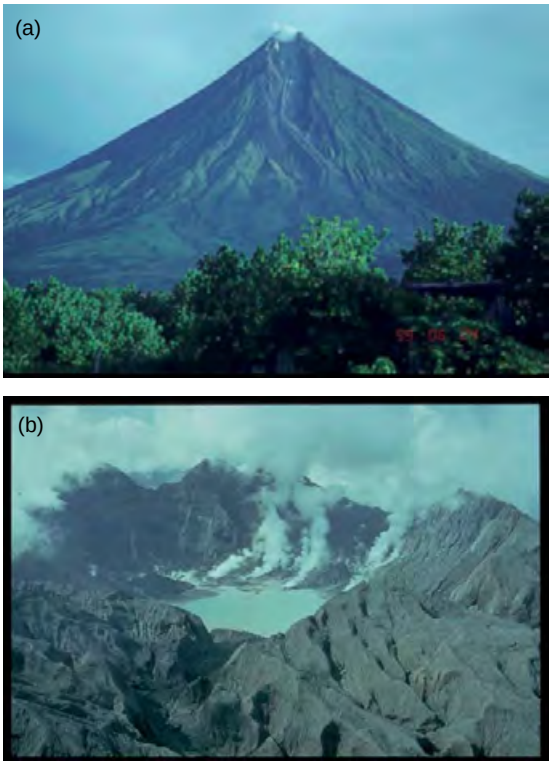


Figure 4 Two arc volcanoes with sharply different patterns of degassing. (a) Mayon, Philippines, which since 1978 has released ~95% of its gas passively and only ~5% explosively. Only 50 kt of SO_2 was recorded by the TOMS sensor, during the explosion of 1984. (b) Pinatubo, Philippines, whose conduit was plugged for approximately 500 years so ~17 Mt of SO_2 (~65% of incoming SO_2) accumulated in the magma and was released in a large explosive eruption in 1991; the remaining 35% was either released passively into the pre-1991 hydrothermal system or remains in the magma. (Assumptions for both volcanoes: continuous deep supply of 10^7 m^3 basaltic magma per year, 1000 ppm S. TOMS data are from Bluth *et al.* (1997) and Carn *et al.* (2003).

4.12.2.4 Viscosity, Density, and Volatile Content

Melt viscosity varies dramatically with magma composition and with cooling and degassing as magma rises from a crustal reservoir and approaches the surface. Higher-silica magmas show greater polymerization of silica tetrahedra, so for similar temperatures, a rhyolite might be 3–7 orders of magnitude more viscous than an alkali basalt (**Figure 5**) (Shaw (1972), Spera (2000) and others). At constant P_{H_2O}, the viscosity of those same magmas will increase by 0.5–2 orders of magnitude per 100°C decrease in temperature (faster in more silicic magmas). Viscosity will increase by 2–4 orders of magnitude as water content drops from 8 wt.% down to < 1wt.%, with changes slightly faster at lower temperatures than at higher temperatures, and especially fast in the last 1 wt.% of degassing. These are profound changes.

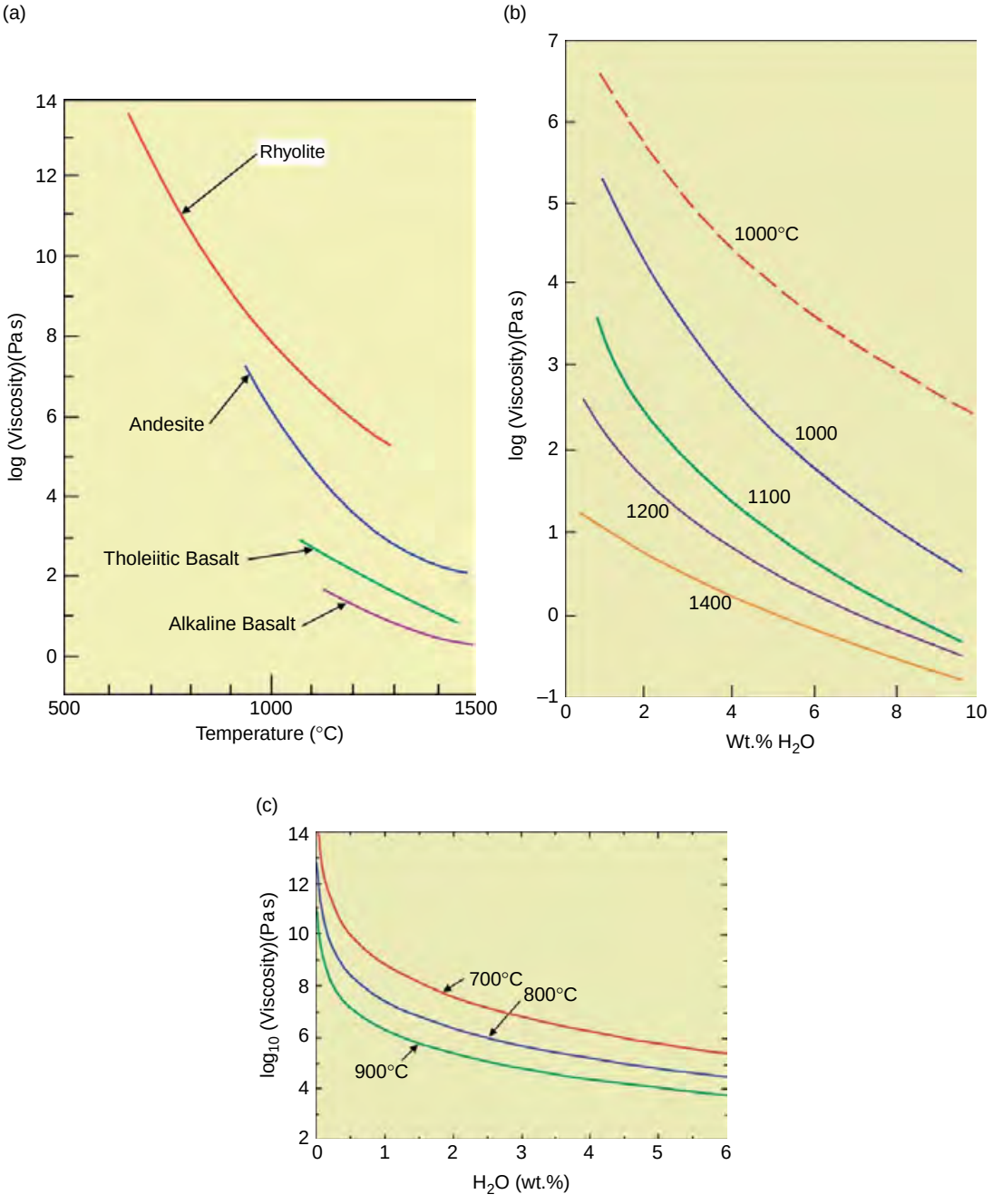


Figure 5 Viscosity of magma varies by orders of magnitude depending on SiO_2 and H_2O content, and can increase by orders of magnitude as it rises to the surface, losing H_2O and growing crystals. (a) Viscosity vs SiO_2 content and temperature; (b,c) viscosity vs H_2O content and temperature, for mafic and rhyolitic melts. Figures from Wallace and Anderson (2000), Dobran (2001), and Hess and Dingwell (1996).

Magma viscosity (i.e., viscosity of the melt + crystal + bubble mix) also increases by orders of magnitude as crystal contents rise from 0% to 60%, following an empirical relation (in lavas) of

$$\mu_{\text{magma}} = \mu_{\text{melt}}(1 - cX)^{-2.5}$$

where $c \sim 1.67$, $X =$ volume fraction phenocrysts (Marsh, 1981).

Magma viscosity can either increase or decrease with bubbles, depending on the 'capillary number' (the ratio of viscous and surface tension stresses in a bubble). Small capillary numbers inhibit stretching of bubbles so an increase in bubbles increases magma viscosity. Large capillary numbers allow bubbles to stretch so an increase in bubbles can lower magma viscosity (Manga and Loewenberg, 2001). In either case, change is usually by less than a factor of 5 (Pal, 2003).

Magma density is a minor function of melt composition (silicic, e.g., rhyolitic magmas are slightly less dense than mafic, e.g., basaltic magmas; 2.4 g cm^{-1} vs 2.6 g cm^{-1}). Even volatiles in bubble form, if above their critical points (for water, 374°C, ~22 MPa; for carbon dioxide, 31°C, ~7 MPa), exert relatively minor influence on magma density. A much stronger influence comes from volatiles that are already in a gas state, in magma at shallow depth at pressures less than a few tens of MPa. Recalling the ideal gas law, $PV = nRT$, pressure and volume are inversely related when temperature and number of gas molecules are constant. If depth and confining pressure decrease, bubbles will expand and the density of magma can decrease dramatically. We can visualize this with the same bottle of soda pop that is considered earlier. The density of soda pop with only dissolved gases is close to 1.0 g cm^{-1}. Once it foams, the density can drop to a small fraction of that.

Volatile contents play into both viscosity and density. Exsolution of water from melt has the effect of increasing viscosity of the remaining melt. Because degassing of melt also induces microlite crystallization, another major contributor to increased viscosity, degassing is especially effective in sharply increasing 'magma' viscosity. To some extent, this is a self-limiting process because the viscosity of the degassed outer carapace of magma will increase so much that it acts to block further degassing, but that carapace is subject to cracking. New bubbles, if they remain in the magma, will increase or decrease viscosity slightly, and have a major effect on its density.

4.12.3 Intrusions and Convection

4.12.3.1 Orientation, Geometry, Dimensions of Intrusions

The shapes and orientations of magma conduits are functions of principal tectonic stresses and pre-existing structures. In cases of intrusion into a homogeneous medium within a tectonic stress field, intrusions will be as roughly vertical planar bodies ('dikes') and perpendicular to the minimum principal stress direction σ_3. As intrusions reach closer to the surface, a vertical axis often becomes the minimum principal external stress direction and there is a tendency for intrusions to become vertical cylinders. However, in regions of extensional stress, σ_3 remains horizontal and intrusions may persist as dikes all the way to the surface. Where magma has limited buoyancy, or encounters an impenetrable horizontal layer during its ascent, it can spread laterally as broadly horizontal planar sheet ('sill').

Basaltic dikes range from a few tens of centimeters to several tens of meters in width, most less than a few meters. Composite dikes (dikes consisting of several intrusions, each one wedging into the previous one) can be wider, as can 'dike swarms' (sets of many subparallel dikes that have intruded the same area over time, each one solidifying before the next). Silicic dikes can be several meters to several tens of meters in width, rarely up to ~100 m wide. Cylindrical intrusions of magma termed plugs, often silicic, can be tens to hundreds of meters in diameter. Sills can vary from several meters to several tens (rarely, hundreds) of meters in thickness. Like composite dikes, sills may inflate by repeated, closely spaced intrusions.

4.12.3.2 Drivers of One-Way Magma Ascent

Conventional wisdom is that magma follows a one-way path upward to eruption. Forces that drive magma upward are (1) buoyancy, (2) lithostatic pressure, and (3) magmatic overpressure. Buoyancy is effective only in fluid–fluid interaction where downward-directed return flow can occur. Lithostatic pressure constantly squeezes the sides of a magma intrusion and that magma will try to move up to a zone of lower lithostatic pressure. The tops of many crustal magma reservoirs are at 3–6 km, where lithostatic pressures are between 100 and 200 MPa; speculative reasons for this fact are raised in Section 4.12.9. Magmatic overpressure is internal pressure in excess of lithostatic pressure, caused by volatile expansion at shallow depths or by addition of magma from greater depth. Like lithostatic pressure, it acts in all directions and not just upward, but the easiest relief of this pressure is usually upward.

In a few cases, lithostatic pressure may be augmented by compressive tectonic strain (literally, squeezing magma upward from cracks or from

ductile rock at depth). Compressive strain of about 0.1 MPa beneath Pinatubo as a result of the 1990 Luzon earthquake might have squeezed basaltic melt upward – from either the lower crust or, perhaps more likely, throughout the whole length of a deep magma feeder system – until it intruded a residual dacitic reservoir (Bautista *et al.*, 1996; Newhall *et al.*, 2002).

4.12.3.3 Drivers of Convective Magma Ascent with Return Flow

The conventional wisdom of one-way transport is hard to reconcile with so-called 'excess degassing', (Andres *et al.*, 1991), a common occurrence in which volcanoes emit far more SO_2 and presumably other volcanic gases than could have been dissolved in modest volumes of magma that are erupted over the same period. A combination of field observations and modeling (Kazahaya *et al.*, 1994, 2002, 2004; Stevenson and Blake, 1998; Witter *et al.*, 2004, 2005) suggests that in many, perhaps most of these cases, the extra gas came from magma that is convecting within the conduit. Low confining pressure allows the top of a magma column to develop into a permeable foam, probably, just a few hundred meters below the surface. Gas escapes and the residual degassed, collapsed foam sinks back down through the conduit (**Figure 6**). Gravitational settling of these blebs of degassed magma is what drives convection, by displacing gas-rich magma upward (Bergantz and Ni, 1999). This is not a 'conveyor belt' convection cell with continuity between upwelling and downwelling limbs, where higher viscosities of the degassed magma would stop the convection. Instead it is a modified, 'lava lamp' style of convection in which the viscosity of the downgoing blebs is irrelevant. Sinking blebs can even be solid and the process will still operate.

Magma foams are known from their 'frozen' equivalents, including very low-density, fluffy 'reticulite' pumice, and are also inferred from relic textures in some silicic lava flows that suggest they were erupted as foam (Eichelberger *et al.*, 1986, Eichelberger (1995); for cautions, see Fink *et al.* (1992)). Dimensional modeling by Witter (2003) indicates that mafic magmas not only can but probably will convect in conduits 1–5 m diameter, and relatively hot, volatile-rich silicic magmas will do the same in conduits 5–20 m in diameter. Relatively, high alkali content of silicic magmas will also reduce viscosity and increase the chance for convection. Rates of

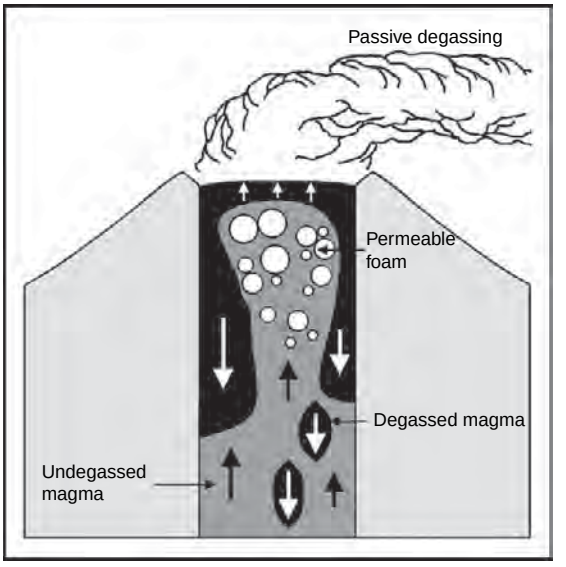


Figure 6 'Lava lamp' convection in a conduit. Gas-rich magma develops a head of foam near the surface. Gas escapes, and the degassing foam collapses. Now dense, it sinks back down through the column, displacing gas rich magma upward. Once the process begins, it will be driven primarily by sinking blebs, unless or until the supply of fresh gas rich magma into the conduit becomes too slow to sustain the process. Adapted from Kazahaya K, Shinohara H and Saito G (2002) Degassing process of Satsuma–Iwojima volcano, Japan: Supply of volatile components from a deep magma chamber. *Earth Planets Space* 54: 327–335.

SO_2 gas emission constrain rates of upward flow during routine, continuous convection to be in the order of mm to cm/sec.

This has important implications for both seismic and deformation precursors of eruptions. Where magma is convecting freely in a conduit, any seismic and deformation precursors to eruption will be subtle and occur only shortly before eruption. Little or no rock needs to be fractured before magma can erupt. Empirically, this is observed at frequently active, relatively open-vent volcanoes, often with mafic to intermediate magmas. Steady-state gas supply to the near-surface, balanced perfectly by gas leakage (degassing), can be tipped in favor of gas supply by a fresh intrusion of magma, gas, and heat at depth and a corresponding increase in the rate of convection. This imbalance, causing gas pressures to build slightly, will produce very subtle seismic and deformation precursors to eruptions – typically recordable only at or near a crater rim just minutes before an explosive eruption, for example, at Sakurajima Volcano (Kamo *et al.* 1994) or Merapi Volcano (Voight *et al.*, 2000). The same subtlety applies at all

open conduits, even where magma ascent appears to be one-way upward, for example, at Soufriere Hills, Montserrat (Voight *et al.*, 1999).

In contrast, seismic and deformation precursors where magma in the conduit is neither convecting nor otherwise open can be expected to be more pronounced, with significant magma pressurization and rock breakage as magma breaks a new path to the surface.

Aside from observed degassing, is there other evidence for convection of magma in conduits? Might there be seismic and/or electromagnetic evidence of magma moving through the 'lava lamp' convection 'cell'. The author is not aware if this has been found, but he expects this to be the case. Do we find large phenocrysts that show oscillatory compositional zonation to indicate that they rode a convection elevator up and down the conduit multiple times? Perhaps, as at Erebus volcano, Antarctica (Dunbar *et al.*, 1994), though most other cases of oscillatory zoning are equivocal and might alternatively be explained by fluctuations in magma water content or temperature without convection. We might expect that U–Th ages of plagioclase crystals in convecting magma might span a wide range of ages, and indeed such ages at Mount St. Helens range from 10^3 years to just a few years (Cooper and Reid, 2003), but neither gas emission nor other evidence points to convection in the conduit of Mount St. Helens. Perhaps the strongest evidence will be found in degassed groundmass glass. In products of the 2000 eruption of Mayon Volcano, Arpa *et al.* (2006) document abundant blebs of (apparently) degassed andesite with microcrystalline groundmass wrapped within darker, (apparently) undegassed lava of the same composition. Degassing was inferred from textures and still needs to be confirmed by direct analysis of the glasses. It would be interesting to reexamine more and less crystalline groundmasses of Merapi lavas described by Hammer *et al.* (2000), particularly in light of the findings of Allard *et al.* (1995) about excess degassing which, to this author, strongly suggests convection of magma in the conduit.

4.12.3.4 Magma Mixing – Conditions, Evidence, Effects

Mafic magmas arriving from the mantle and lower crust often encounter more silicic magmas enroute to the surface. This phenomenon is especially common at arc and continental volcanoes at which the long-term supply rates of mafic magma are sufficiently high (10^6–$10^8\,m^3\,yr^{-1}$) to generate and sustain silicic magmas by partial melting of crust and by fractional crystallization and other differentiation processes in long-lived magma reservoirs (Shaw, 1985). Such systems have a thermal memory of previous intrusions, and many have large silicic magma bodies (<5 to 1000's of km^3) at depths ~5–15 km beneath the surface.

In order for magmas to mingle or mix, their viscosities must be within approximately one order of magnitude of each other (Sparks and Marshall, 1986). A number of models of mingling and mixing have been proposed. One mechanism involves underplating of basaltic magma beneath a silicic reservoir, with thermally driven convection in one or both layers and mixing along a subhorizontal interface (Sparks *et al.*, 1977; Sparks and Marshall, 1986). Another mechanism involves intrusion of mafic magma into silicic magma. Rising mafic magma may have sufficient momentum to simply intrude right into overlying silicic magma (Bergantz and Breidenthal, 2001). In a third mechanism, cooling and partial crystallization of mafic magma when it encounters less dense silicic magma will induce vesiculation of the mafic magma, make it less dense than the silicic magma, and buoyancy will let mafic magma 'tunnel' vertically into the silicic body. Mingling will occur along a subvertical interface (Eichelberger, 1980; Bergantz and Breidenthal, 2001).

Mingling and mixing of two magmas is clearly evident in the hybrid products. Blebs of the intruding, vesiculated basalt may be quenched and preserved as inclusions in the hybrid magma, with distinctive open-lattice ('diktytaxitic') textures (Eichelberger, 1980; Pallister *et al.*, 1996). Further shear can cause quenched blebs to mechanically disintegrate. Incomplete mixing ('mingling') produces a marble-cake swirl of the two components. If mafic melt mixes completely with silicic melt, compositionally intermediate melts develop. Both mingled and mixed magmas show disequilibrium juxtaposition of otherwise incompatible minerals, for example, olivine and quartz, and reaction rims on minerals from the higher temperature end member, or resorption of the rims of plagioclase from the lower temperature end member.

Magma mixing is not known to have a direct geophysical signal. However, changes of temperature and oxygen fugacity associated with mixing can substantially change the solubilities of gases in both magmas. If solubilities are decreased enough, bubbles

will form and one should expect seismic and deformation signals associated with buoyant rise of magma and expansion or pressurization of bubbles as that magma rises and lithostatic pressure decreases. One of the best-documented instances of magma mixing began at Pinatubo Volcano several months before its eruption in June 1991 (Pallister *et al.*, 1996; M. Coombs, personal communication, 2002); the latest mixing occurred <1 week before eruption (Rutherford and Devine, 1996). Water-rich basalt intruded (tunneled into) and mixed with sluggish dacite magma, early in a chain of events that, by 15 June 1991, led to eruption of $\sim 5\,km^3$ of dacitic magma, the second largest eruption in the twentieth century.

4.12.3.5 Geospeedometers of Magma Ascent, Convection, and Mixing

Despite what we might expect from a dike propagating toward the surface, there are very few known cases in which earthquake hypocenters have migrated upward. A broad ~3 km upward migration of hypocenters may have occurred at Pinatubo (Philippines) in 6 days in June 1991 ($\sim 20\,m\,h^{-1}$) (Harlow *et al.*, 1996). That migration was apparent in fast playback of hypocenters in VOLQUAKE (Hoblitt *et al.*, 1996), though less obvious in relocated hypocenters (Rowe *et al.*, 2001). In 1975 at Plosky Tolbachik (Kamchatka), Fedotov *et al.* (1980) traced seismicity (and, by inference, basaltic dike intrusion) from 20 km to the surface at an average rate of about $0.1\,km\,h^{-1}$. In both cases, hypocenters spanned a significant part of the conduit length at any given time, but moved broadly upward. Other reported cases are Kilauea east rift basalt at $0.1–1.5\,km\,h^{-1}$ (Klein *et al.*, 1987), East of Izu basalt at $1\,km\,h^{-1}$ (Hayashi and Morita, 2003), and Unzen dacite at $\sim 3\,m\,h^{-1}$ (Nakada *et al.* 1999).

In principle, the tip of a dike would also be a point source of deformation (Mogi, 1958) and this, too, could be tracked upward. There are a few instances in which tiltmeters have recorded traces consistent with upward migration. A growing body of InSAR data will surely show some new rates of migration; one early result is that of Lu *et al.* (2003) for magma dynamics at Westdahl Volcano, Alaska.

A different geospeedometer – applicable during and perhaps before eruptions – considers the diameter of dense ultramafic or mafic xenoliths that get erupted, calculates their likely settling rate, and infers that rate as a minimum upward velocity of magma in order to bring these xenoliths to the surface (Spera, 1984). In the 1801 eruption of Hualalai, Hawaii, magma ascent was apparently $>30\,m\,h^{-1}$.

Several petrologic geospeedometers exist, all relying on the fact that minerals which are taken out of equilibrium will develop disequilibrium reaction rims (e.g., hornblende will develop pyroxene and Fe–Ti oxide rims), and the travel time for that magma from reservoir (equilibrium) to the surface (quenching) is modeled as the time required for the observed thickness of reaction rim to grow in lab experiments (Coombs *et al.*, 2000; Rutherford and Gardner, 2000). Sometimes, the absence of zoning or reaction rims can also be used to make a minimum estimate of ascent time – or maximum diameter of a conduit – on the premise that zoning or rims would have developed given enough time.

In discussion of convection, earlier, we noted that rates of ascent can be judged from gas flux rates and assumptions about conduit diameter and the concentration of gas in the magma. These estimates suggest rates of meters to tens of $m\,h^{-1}$ (Witter, 2003).

By similar arguments, a known volume rate of extrusion and an assumed conduit diameter will give an ascent rate. At Mount St. Helens in the 1980s, mass eruption rates during explosive events implied ascent rates of $1–2\,m\,s^{-1}$ ($3600–7200\,m\,h^{-1}$). During nonexplosive dome extrusion in the 1980s, and again in 2004–2005, ascent rates were tens to hundreds of $m\,h^{-1}$.

Finally, the time from magma mixing to eruption can be estimated by the thickness of reaction rims or by the preservation versus loss of disequilibrium mineral pairs that were juxtaposed by mixing. From Arenal Volcano, olivine reaction rims of at least 11 year growth are described by Coombs and Gardner (2004) and, at the opposite end of the time spectrum, preservation of disequilibrium Fe–Ti oxides at Mount St. Helens, Pinatubo, Unzen, and Soufrière Hills or cummingtonite (an amphibole mineral not stable above 800°C) at Pinatubo indicate maximum times from mixing to eruption of only a few days (Rutherford and Devine, 1996; Pallister *et al.*, 1996; Devine *et al.*, 1997, 2003; Pallister *et al.*, 2005). In another approach, Blake and Fink (2000) propose a conceptual framework from which timeframes of mixing might eventually be estimated from the relative sizes and degrees of deformation of mafic enclaves that get 'frozen' upon being mixed into cooler silicic magma.

4.12.4 Pressurization within and around Magma

4.12.4.1 Increase in Pressure

Pressures will be increased by the following.

4.12.4.1.1 Volatile accumulation

Volatile accumulation is to eruptive potential as accumulated strain is to tectonic earthquakes. To the extent that there are 'eruption cycles', they are driven by patterns of net volatile accumulation. Recall that under discussion of gas budgets, it is postulated that magma and gas supply to most volcanoes is relatively frequent and in small increments. At some volcanoes, it may be almost continuous at all timescales; at others, it may be intermittent over timescales of days to years but constant over timescales of decades. In either case, the interval of magma supply from depth, into a shallow crustal reservoir, is probably quite a bit shorter than the eruption interval.

Plots of cumulative volume of magma erupted versus time for various volcanoes can show time predictability, volume predictability, or no predictability, that is, completely Poisson behavior (Koyama and Yoshida, 1994). Most volcanoes that have open vents and leak gas and magma fairly freely are time predictable, as the threshold of internal pressure that needs to be exceeded is low and constant. Volcanoes whose conduits freeze up between eruptions are often volume predictable. At such volcanoes, explosivity correlates with volume erupted, both being controlled by the accumulated mass of volatiles that have not leaked out before magma reaches close to the surface. Most of the largest explosive eruptions – 'plinian eruptions' – are of magma that is not only saturated with volatiles but has also accumulated 10×–100× more volatiles (by weight) in bubble form (e.g., Pinatubo; Gerlach *et al.* (1996) and Wallace (2001)). This is the basis for the oft-repeated dictum that a volcano which has been dormant for hundreds of years is likely to produce a large explosive eruption. This dictum is true only if the conduit has been closed and gas has been accumulating, not leaking. Pinatubo was just such a volcano, and accumulated most of the incoming gas for ~500 years. Iwo-jima in Japan has had almost as long a repose period and plenty of magma supply but shows no signs of erupting magma – because it is exceptionally leaky (Newhall *et al.*, 1994, 2003; Newhall, 2004; Ukawa *et al.*, 2005).

If volatiles accumulate in excess of saturation, forming bubble phases, further accumulation will gradually increase pressure on the walls of the container. The effect is minimized at depths where volatiles are most likely accumulated and volatile phases are in a supercritical fluid. However, accumulation of volatile fluid bubbles increases the 'potential' for rapid pressure buildup at shallow depths where confining pressures are low.

4.12.4.1.2 Magma ascent and volatile expansion

As magma ascends and lithostatic pressure decreases, partial pressure of the gas increases and only the strength of viscous magma and country rock will hold it in check. The pressure of this contained gas will drive magma to intrude through country rock. Eventually, if the internal gas pressure exceeds the strength of the container, the container will fail and magma will erupt.

4.12.4.1.3 Heating and/or compression of groundwater

Confined groundwater, hot or cold, can add significantly to pressures in and around a magma body. If heated, it expands and increases pore water pressures (Reid, 2004), though slow diffusion of heat through groundwater will keep effects quite local or slow. Thermally induced hydraulic pressurization may be greatest at the tip of a rising body of magma – that is, a wedge driving open cracks.

If porous host rock undergoes sudden compression due to nearby magma intrusion, its pore water pressures will jump quickly (Sato *et al.*, 2001; Shibata and Akita, 2001; Newhall *et al.*, 2001; Matsumoto *et al.*, 2002; Roeloffs and Linde, 2007). Observations of research wells at Usu Volcano in Japan and Krafla Volcano in Iceland have shown up to 100 m immediate rise in water level, corresponding to a notable 1 MPa increase in pore water pressure. This pressure bleeds off in a matter of days, but can briefly trigger seismicity for 5–20 km around a volcano (White and Power, 2001), and/or sharply decrease slope stability as probably occurred at Mount St. Helens in 1980, Unzen in 1792, and many other volcanoes that have massive sector collapses (Day, 1996; Voight and Elsworth, 1997).

4.12.4.1.4 Self-sealing in hydrothermal systems

Hydrothermal waters often contain dissolved silica and other minerals, and precipitate them near their margins. The effect is to develop an impermeable

seal, trapping hot hydrothermal brines beneath (Fournier, 1999, 2007).

4.12.4.1.5 Lithostatic pressure

As described earlier, lithostatic pressure is the main source of pressure within magma bodies at depth, where pressures and temperatures are above the critical point for water. Lithostatic pressure squeezes the sides of a magma body at the same time it 'weighs heavily' on that magma. Squeezing of this body will be greater at its base than at its top, hence creating a driving force for upward intrusion. Furthermore, if a large body of magma feeds upward into a narrow feeder (e.g., a dike), magma in the conduit will act like a manometer (as in the straw of a slightly squeezed juice box), and stress will be concentrated at the tip of that dike that will help the magma break its way up through country rock.

4.12.4.1.6 Regional tectonics – quasistatic and dynamic strain

Any compression added by quasistatic or dynamic strain will enhance the squeezing mentioned above. Conversely, decompression, whether static or transient (dynamic), will tend to lower the solubility of gases in solution and/or allow expansion of bubbles in the magma. Lowering of solubility can increase internal volatile pressures more than the triggering decrease of confining pressure.

4.12.4.2 Decrease in Pressure

The pressure will be diminished by the following.

4.12.4.2.1 Gas leaks

The effect of gas leaks from magma is like that of gas leaks from soda pop: the magma goes flat. For details, see earlier discussion, Giggenbach (1996), and Jaupart (1998).

4.12.4.2.2 Eruptions

Volatiles tend to accumulate in the uppermost parts of magma conduits and reservoirs, and will be depleted by eruptions that, naturally, tap those parts. Eruptions are major 'gas leaks'.

4.12.4.2.3 Groundwater diffusion and dissipation of porewater pressures: Breakage of self-seals

A pressurized hydrothermal fluid surrounding magma will eventually bleed off its pressure through porous flow or fractures. If the cause of pressurization in the first instance was development of silica-rich self-seals, fracturing of those seals from within or by regional earthquakes will rapidly depressurize the source fluid but may briefly pump up pressures along connected, surrounding fractures. The concept is described by Fournier (1999, 2007).

4.12.4.3 Dynamic Balance between Pressurization and Depressurization

In summary, a variety of processes act to increase or decrease pressures within magma and surrounding hydrothermal systems. The increases and decreases may be in delicate balance. This delicate balance often appears during ascent of gas-rich magma from a crustal reservoir toward the surface. Volatile exsolution occurs as solubilities in melt decrease, and both newly exsolved and any bubbles already existing in the reservoir exert increasing internal pressure as lithostatic pressure decreases. If gas is able to leak out and decrease internal pressure as fast at that pressure is increased from ascent, there will be no net increase in gas pressure and magma may arrive at the surface with little or no explosive potential. Conversely, if magma ascent is rapid and pressures build faster than they can bleed off, an explosive eruption will result.

A parallel exists in hydrothermal systems. Confined groundwater can be pressurized by addition of insoluble gas, by heating (from that and from soluble gas), and, most effectively, by mechanical compression of the aquifer by nearby magma intrusion. If these pressurization processes are slow, porous and fracture outflow may prevent any rise of porewater pressure. On the other hand, there are clearly some instances in which water pressures rise sharply, whether at the tip of a rising dike or more broadly, and resulting hydrofracturing may facilitate magma ascent and/or phreatic explosions.

4.12.5 Volcanic Unrest

This section is a brief overview of the nature, magnitude, and frequency of 'observable' physical and chemical changes that occur between eruptions, and especially as magma intrudes toward the surface. Principal measuring techniques are also mentioned. Useful general references include UNESCO (1972), Tilling (1989), Carroll and Holloway (1994), McGuire *et al.* (1995), Scarpa and Tilling (1996), and Sigurdsson *et al.* (2000). Readers might note that

gas emissions are described ahead of seismicity and ground deformation, opposite the usual sequence. This is for two reasons. First, gas emissions are often the first potentially detectable sign of unrest. Second, increases of gas and hydrothermal pressures are what drive volcanic seismicity, ground deformation, and magma ascent.

4.12.5.1 Gas Emissions

In situ collection of gas samples from fumaroles, for subsequent analysis in a laboratory, has the advantage of collecting the whole gas and allowing use of a range of laboratory analytical instruments (Symonds *et al.*, 1994). However, for monitoring purposes it has the disadvantages of collections being infrequent and often dangerous. Remote measurements from a safer distance like the foot of a volcano, from aircraft, or even from satellites have the advantage of being more frequent and safer but are usually of just one or a few gas species.

The most easily and widely measured indicator of magma degassing is sulfur dioxide (SO_2). It is much more abundant in magmatic gas than in the atmosphere and can be measured *in situ* or remotely by now-small and relatively inexpensive UV spectrometers. Early work (Stoiber and Jepsen, 1973; Stoiber *et al.*, 1983) showed that SO_2 emission increased before eruptions as its solubility in rising magma decreased. Subsequent work has found that this pattern is sometimes complicated by path effects. One such effect is temporary blockage of SO_2 exit from magma by sharp increase in viscosity of magma and/or by quenching and solidification of a carapace on the magma. This effect is thought to explain brief decreases in SO_2 emission shortly before explosions of Galeras volcano (Fischer *et al.*, 1996; Zapata *et al.*, 1997) and perhaps at Pinatubo as well (Daag *et al.*, 1996). Another path effect is absorption of SO_2 into groundwater, in which it is highly soluble (Doukas and Gerlach, 1995; Symonds *et al.*, 2001). The apparent absence of SO_2 before eruptions, once thought to indicate absence of magma, is now understood in some cases to reflect absorption (scrubbing) of SO_2 into groundwater. Scrubbing persists as long as the gas must pass through water-saturated country rock; it ends and SO_2 emission can rise sharply after hot gas has boiled open a dry chimney through the groundwater.

Increasingly, CO_2 gas is also monitored – directly from fumaroles, by remote sensing (FTIR), and by measurement of CO_2 flux through soil. CO_2 has the lowest solubility of any of the major gases in silicate melt, so it is the first to be released – starting when magma is as deep as ~20 km depth). It is also much less soluble in groundwater than SO_2, so at many volcanoes, it leaks out of the slopes of the volcano (diffuse degassing) and can be measured in soil. At Usu Volcano, an increase in diffuse CO_2 outgassing appeared 6 months before seismic or geodetic changes were detected (Hernandez *et al.*, 2001). At Mammoth Mountain in California, diffuse CO_2 outgassing was first noticed in late 1989 and early 1990, months after seismicity in mid 1989 signaled a magma intrusion at ~2–10 km depth, and rose again during an increase in seismicity at 10–20 km depth throughout 1997 and at 4–8 km depth in late November 1997 (Gerlach *et al.*, 1998; McGee *et al.*, 2000; Hill *et al.*, 2003). The apparent delay in CO_2 in 1989–90 may simply reflect delayed sampling; there was no delay in 1997 (T Gerlach, oral communication).

Because the solubility of SO_2 in silicate melt is greater than that of CO_2 and less than that of HCl or HF, rising magma might produce CO_2 at first, then increasing SO_2/CO_2 ratios, and eventually increasing HCl/SO_2 or HF/SO_2 ratios (e.g., Noguchi and Kamiya, 1963). This simple solubility-related progression is a nice conceptual framework for monitoring and might even offer prospects of estimating magma depth from gas emissions. However, discovery that many magmas are already gas-saturated in their reservoirs, and that other gases remaining in melt will partition preferentially into bubble phases rather than silicate melt, means that there may already be a fairly complete suite of magmatic gases in bubbles at depth, and all can be released together.

4.12.5.2 Seismicity

Because this author is a geologist rather than a seismologist, and because (Chapter 4.13) give readers a modern view of volcanic seismology, I originally thought to skip this section entirely. However, it might be interesting for seismologists to have a glimpse of how this one nonseismologist struggles to understand their field, and to relate it to other topics in this review.

Most volcanic seismicity falls into categories described by Minakami (1960), Shimozuru (1972), McNutt (1996, 2000a, 2005), and (Chapter 4.13). Volcano-tectonic (VT) earthquakes result from brittle fracture (of country rock) beneath, within, and near the volcano, caused by rise and pressurization

of magma or induced by related increases in adjacent porewater pressures. Most VT events are normal, high-frequency events and are individually indistinguishable from small tectonic earthquakes. Indeed, some may be nothing more than tectonic earthquakes along faults beneath the volcano. What distinguishes VT events from strictly tectonic events is their common occurrence in swarms rather than in mainshock–aftershock sequences, their location beneath or near active volcanoes, and their common association with other indications of magma intrusion. *b*-values of VT earthquakes tend to be higher than for tectonic events (Sanchez *et al.*, 2004) because of heterogeneous structure and limited fault lengths, and often-elevated pore-pressures in hydrothermal fluids (Sanchez *et al.*, 2004; McNutt, 2005).

A subset of VT earthquakes that occurs around rather than beneath volcanoes has been called 'distal VTs' by White and Power (2001). Some volcanoes have earthquake swarms at distances of a few up to ~20 km from their summit, out beyond the margins of any magma body. Centered 17 km from Guagua Pichincha volcano, the North Quito swarm turned on and off as the hydrothermal system of the volcano pressurized and then erupted, respectively (Legrand *et al.*, 2002). Near Soufrière Hills and Pinatubo volcanoes, swarms 3 and 5 km from the respective summits occurred during early stages of their eruptions, as pressures in the magma reservoir and surrounding hydrothermal system increased. There is no implication that magma itself is present where distal VTs are occurring, nor even in hydraulic connection with the hypocentral area. The correlation can be through propagation of mechanical strain that acts on locally confined aquifers.

Decades ago, Minakami (1960) recognized a general progression from high-frequency, impulsive-onset VT (or, in his terminology, A-type) earthquakes at depths >1 km, upward to low-frequency (his B-type) earthquakes at <1 km depth that were often more emergent as well as having lower frequency content. Minakami's concept of a standard progression of earthquake types as magma rises remains useful, and was nicely updated into a generic earthquake swarm model by McNutt (1996).

Subsequent work suggests that 'B-type earthquakes' – shallow and of relatively low frequency – are actually of multiple origins. Some reflect source processes as discussed in the next paragraph, and others reflect attenuation of high frequencies and other path effects as emphasized by Minakami (1974) and Malone (1983). Ishihara and Iguchi (1989), Iguchi (1989), and Tsuruga *et al.* (1997) distinguish two to three subcategories of 'B-type' earthquakes at Sakurajima, representing both fracturing and subsurface explosions.

Low-frequency earthquakes at volcanoes have several different origins. Some seem to occur by rupture of relatively ductile material, either hydrothermally altered rock or of viscous, nearly solid magma (Ramos *et al.*, 1999; Tuffen and Dingwell, 2004; Moore *et al.*, 2005). If strain rate is high enough, even ductile materials can rupture. A subset of these low-frequency earthquakes ('hybrids') has a high-frequency onset followed by a low-frequency tail, and these commonly occur within or just beneath growing lava domes. Rupture may begin in the brittle carapace and propagate inward to softer material. Alternatively, hybrids might reflect occurrence of an earthquake in brittle rock that sets nearby fluid into oscillation (McNutt, 2000a), or result from complex paths rather than complex sources (Kedar *et al.*, 1996, 1998).

Another class of low-frequency earthquakes is sometimes termed a long-period (LP) earthquake to convey an added character – a relatively monochromatic character, dominantly 1–3 Hz. Some of these events are strikingly monochromatic, such as so-called 'tornillo' earthquakes of Galeras Volcano and elsewhere (Gil Cruz and Chouet, 1997; Gómez and Torres, 1997). Chouet (1992, 1996a, 1996b) modeled these as the result of sudden jetting of pressurized gas through cracks, followed by resonance of the crack that just opened and shut. In some but not all cases, there is visual corroboration of steam jetting shortly after LP events. Initially, occurrence of LP events in the buildup toward explosive eruptions (e.g., at Redoubt, 1989–90) led to an inference that these reflected gas pressures within the magma, and explosive potential (Chouet, 1996a). Subsequent observation of many such events in the aftermath of eruptions, when conduits are relatively open and streaming gas, suggests jetting of gas through cracks but not necessarily under great pressure. Hellweg (2003) suggests an interesting alternative, yet to be fully tested, that the monochromatic or harmonic character of some such events might be a path effect, resulting from scattering polarization of signals.

Some low-frequency (LP) events occur deep beneath volcanoes. At Pinatubo, deep LPs (DLPs) were approximately 35 km deep and were observed in late May and early June, only a week before magma erupted (White, 1996), but evidence for

earlier magma mixing suggests that other LPs might have occurred before seismic monitoring was established. Beneath Kilauea, deep LPs are 30–60 km deep (Koyanagi *et al.*, 1987), and beneath Fuji, 10–20 km deep (Nakamichi *et al.*, 2004; Ukawa, 2005). A number of examples are noted by Power *et al.* (2004). DLPs are thought to reflect resupply of magma from depth into shallower reservoirs. They do not by themselves indicate imminent eruption.

With increasing use of broadband seismometers at volcanoes, there are increasing reports of very-long-period (VLP) earthquakes with periods of up to several tens of seconds (e.g., Kumagai *et al.*, 2003; Chapter 4.13). Some are thought to occur when slugs of fluid (magma, water, or gas) pass through constrictions. The very long period reflects the time it takes for fluid head to pass through the constriction.

The last major category of volcanic seismicity is volcanic tremor – periods of continuous vibration lasting from minutes to years (McNutt, 1992; Konstantinou and Schlindwein, 2002; Instituto Geofisico, 2006). The energy of tremor normalized for station distance, magnification, and signal spreading is reported as reduced displacement (Aki *et al.*, 1977; Fehler, 1983). For surface waves, reduced displacement $(\mathrm{RD}) = \left(A\sqrt{\lambda r}/2\sqrt{2}M\right)$, where A is their peak-to-peak amplitude, λ is the wavelength, r is the distance in km from source to seismic station, $2\sqrt{2}$ is the root mean square (rms) amplitude correction, and M is the instrument magnification. Some tremor evolves from or devolves into a succession of many small earthquakes, for example, at Mount St. Helens (Malone and Qamar, 1984) or Soufrière Hills, Montserrat (Neuberg, 2000), and is apparently just overlapping of those earthquakes, be they high or low frequency.

Other tremor begins as a low-level continuous signal, grows and is sustained, and eventually fades away as it began. Some is broad spectrum; other tremor can be of remarkably narrow spectrum with harmonics. One suggested origin, in addition to overlapping of discrete events, is nonlinear oscillation induced by flow of magma (or hydrothermal fluid) through conduits (cracks) (Julian, 1994, 2000; Balmforth *et al.*, 2005). A second mechanism is excitation and resonance of a fluid-filled crack, irrespective of whether fluid is flowing (Chouet, 1992, 1996a, 1996b). Both origins are widely inferred.

Generation of tremor by a third mechanism, hydrothermal boiling, gets less attention than it probably deserves. Leet (1988) described experiments on 'subcooled boiling', that is, boiling in which bubbles form where groundwater is in contact with magma or hot gas, but then implode within cold surrounding water. Subcooled boiling appeared to be capable of generating strong tremor. At least some instances of strong tremor with reduced displacement $>30\,\mathrm{cm}^2$ are associated with phreatic eruptions and times when we know that magma is intruding and probably boiling groundwater, for example, at Mount St. Helens in April–May 1980 and a during a short but strong episode on 2 October 2004. Some such tremor is also 'banded', turning on for a few hours, off for a few hours, and back on and off for days, strongly suggestive of geyser activity (e.g., at Karkar Volcano; McKee *et al.* (1981)). Elegant studies of tremor at Yellowstone were published by Kieffer (1984) and Kedar *et al.* (1996, 1998).

4.12.5.3 Ground Deformation

Ground deforms around magma in two ways – elastically in response to pressure (magmatic or hydrothermal), and inelastically as it is fractured and physically shoved aside by magma or if the flank of a volcano is collapsing (e.g., Johnson, 1987). Dzurisin (2003, 2007) offers excellent reviews of methods and interpretation. Elastic deformation is a normal consequence of magma intrusion and pressurization at depth, and can be aseismic. Radar interferometric (InSAR) monitoring of volcanoes has shown a number of volcanoes that are being deformed by relatively deep sources (probably, resupply of basaltic magma) without accompanying seismicity, for example, at Westdahl Volcano, Alaska (Lu *et al.*, 2003; Dzurisin and Lu, 2007). Another good example is in the southwest corner of the Three Sisters volcanic complex in Oregon, where deformation began as early as 1996 and was not known to have induced any seismicity until a small, shallow swarm in March 2004 (Wicks *et al.*, 2002; Dzurisin *et al.*, 2004).

The shape and gradients of the deformation field indicate the depth and shape of the deformation source (Mogi, 1958; Okada, 1992; Cervelli, 2000). It is beyond the scope of this chapter to describe the relationship between sources and deformation fields in detail, but a few general relations can be noted. Uplift from a point source (such as the rising tip of a magma column) forms concentric contours. The half-width of uplift, defined as the diameter of the contour of half the maximum uplift, is 3/2 times the depth of the point source. Tilt resulting from a point source shows its maximum value at a radial horizontal distance 0.5 times depth of the point source. Different

patterns result from dikes. These models, developed by Mogi and Okada on theoretical grounds, are remarkably robust and consistent with seismic and other evidence for magma depths and shapes.

Inelastic deformation immediately around a magma intrusion is just extreme strain, exceeding the tensile strength of the rock. For example, grabens typically form above a near-surface dike, and the north flank of Mount St. Helens bulged irreversibly northward as a viscous dacite cryptodome (hidden dome) intruded into its north flank.

Iwo-jima Volcano (Japan) shows an interesting combination of elastic and inelastic deformation. Over the past 400 years at least, and perhaps the past 2600 years, the floor of the submarine Iwo-jima caldera has risen more than 120 m above sea level (Kaizuka *et al.*, 1985; Oyagi and Inokuchi, 1985; Ukawa *et al.*, 2005). Episodic magma intrusion of approximately $10^7\,m^3\,yr^{-1}$ is the driver. Some of the deformation is elastic, in which episodes of inflation once every decade or two are followed (partially recovered) over succeeding years by deflation of the central part of the island but, at least recently, by continued inflation outboard of the center (Ukawa *et al.*, 2005). But net deformation since the time of caldera formation is inelastic, and includes broad uplift plus numerous faults with offsets of several meters that cut post-WWII era runways (Oyagi and Inokuchi, 1985). Iwo-Jima is an actively resurging dome, and is also remarkable for its efficient degassing of magma that (at least for now) seems to prevent magmatic eruption (Newhall *et al.*, 2003). Some of the deformation clearly involves intrusion, possible pressurization, and degassing of magma; other deformation probably involves pressurization and depressurization of the surrounding hydrothermal system.

Because some nonvolcanic landslides are known to show accelerating creep shortly before failure, those monitoring the bulge at Mount St. Helens in 1980 were hoping for a similar indication of imminent failure. It did not occur. If anything, creep slowed slightly in the 3 days immediately before the failure (Lipman *et al.*, 1981). Apparently, resistance along the basal plane of the landslide block(s) was overcome suddenly.

4.12.5.4 Other Changes – Gravity, Magnetic, Electrical, Thermal, Hydrologic

Ascent of magma and release of hot gases into surrounding groundwater causes a variety of other changes. The strength of the gravitational field around volcanoes (typically, known at just one or a few monitoring sites) can increase or decrease by tens or even hundreds of microgals (10^{-1}–10^{+1} gu) (Rymer, 1996). An increase could be interpreted as subsidence, intrusion of magma that is denser than the rock into which it intrudes, constant-volume densification of magma by collapse of vesicles and coupled inflow, and/or a compression-induced rise in water table (increasing density just beneath the gravimeter). A decrease could be interpreted as uplift, drainback of magma (rare), vesiculation of magma, or dilatation-induced drop in water table. Independent geodetic measurements and modeling of the combined deformation-gravity field are needed to discriminate between these possible contributors (Rymer and Williams-Jones, 2000). Frequent or, better yet, continuous measurements are needed to capture short-lived, transient events. With GPS and other geodetic instruments we can have continuous measures of deformation, but only a very few volcanoes have had continuously recording gravimeters (Etna, Vesuvius, Merapi, Asama, Iwate, Miyake-jima, and Kilauea), and even fewer still have them.

Magnetic fields are known to change by as much as ten or more nanoteslas at some volcanoes and during some intrusions (Johnston, 1989; Sasai *et al.*, 1990; Zlotnicki, 1995). Heating can slowly demagnetize rock near an intrusion. Stress associated with intrusions can induce piezomagnetic effects, in the order of 1 nT per 10 MPa of overpressure (Zlotnicki, 1995). Still other changes may be of electrokinetic origin, when a conductive fluid (hydrothermal or magma) is moving (Mizutani *et al.*, 1976; Fitterman, 1978).

Magma intrusions can change the distribution of resistivity within a volcano, by changing porosity and permeability of a hydrothermal reservoir or conductivity of a hydrothermal fluid. Lenat (1995) gives a useful review of these changes and ways to measure them. One particularly interesting example is at Kilauea, where active-source electromagnetic monitoring saw change in phase and amplitude of signals during passage of a dike intrusion, and re-measurements of self-potential across a rift zone also showed change after a dike intrusion (Jackson *et al.*, 1985; Jackson, 1988). Similar changes have been recorded at Piton de la Fournaise and Etna volcanoes (Lenat, 1995), and at Miyake-jima (Zlotnicki *et al.*, 2003).

Minor shifts in the locus of thermal activity occur frequently at volcanoes, probably related to

self-sealing and changing patterns of permeability within the hydrothermal system. Most of these are insignificant. Increases in ground, water, or gas temperatures by several degrees or even tens of degrees are more significant, potentially caused by increased flux of hot gas. Crater lakes are particularly useful integrators of thermal and chemical flux from hot gas streaming up into the lake (Varekamp and Rowe, 2000), and Bernard (2004) is developing a useful application of ASTER satellite imagery to monitor changes in crater lake temperatures. Fumaroles at or near the boiling temperature of water rarely change temperature because they are buffered by boiling of groundwater. High-temperature, drier fumaroles (some, as hot as 950°C) can have many-degree swings in temperature with increases or decreases in the flux of gas from magma. Large increases in volume of steaming, as at Mount Baker (1975), Pinatubo (1991), Turrialba (1997–present), or Garbuna (2004) can be another indication of hot gas from a fresh magma intrusion heating water of a hydrothermal system.

Finally, changes in water level or spring discharge occur at some volcanoes, principally in response to mechanical compression or dilatation of confined aquifers (Newhall *et al.*, 2001; Roeloffs and Linde, 2007). Most reports of change in water level are anecdotal but good measurements have been obtained in Japan and in Long Valley caldera. As mentioned earlier, increases in porewater pressure can induce seismicity at or near a volcano, cause modest inflation, and destabilize the flanks of volcanoes if pressurization occurs in a widespread, outward-dipping weak layer such as a layer of weathered ash (Day, 1996; Voight and Elsworth, 1997; Reid, 2004).

4.12.6 Volcanic Eruptions

Historically, volcanologists believed that nearly all eruptions were triggered internally, that is, when pressures within magma exceeded the tensile strength of country rock. No doubt this is true for many eruptions. But, especially since the 1992 Landers earthquake, evidence has built that some eruptions are triggered externally (Hill *et al.*, 2002; Schmincke, 2004; Manga and Brodsky, 2006). Here is a list of potential eruption triggers.

4.12.6.1 Eruption Triggering Mechanisms

4.12.6.1.1 Internal

4.12.6.1.1.(i) Intrusion of fresh magma (and its gas) from depth, driven by partial melting and by buoyant forces and/or lithostatic squeezing These drivers exert modest upward and lateral pressures that can overcome tensile strength of country rock and lead to eruption. Also, gas dissolved within that incoming magma is added to pre-existing magma, potentially increasing buoyancy and internal gas pressure to the point of failure.

4.12.6.1.1.(ii) Buildup of gas pressure as magma rises The solubility of gases decreases with decreasing lithostatic pressure, and if gas has no room to expand, its internal pressure will rise. The effect holds true regardless of whether magma is on one-way ascent or in the upwelling limb of two-way convection. Intrusion of fresh magma (Section 4.12.6.1.1.(i)) and any associated increase of gas pressure are probably the most common triggers of eruptions.

4.12.6.1.1.(iii) In situ buildup of gas pressure from crystallization, as in 'second boiling' Crystallization concentrates volatile phases in the residual, interstitial silicate melt. When those concentrations rise above gas solubility, bubbles will form and increase in either volume or pressure. This process, called 'second boiling', may trigger relatively small eruptions.

4.12.6.1.1.(iv) Mixing of fresh magma with pre-existing magma Mixing may add gas to the pre-existing magma, and can also change the oxygen fugacity (f_{O_2}). Sulfur gases in particular have a solubility minimum in temperature-f_{O_2} space, so if mixing pushes magma toward this minimum in gas solubility, more gas will exsolve (Kress, 1997). Addition of fresh mafic magma usually raises the temperature of the pre-existing magma, in turn decreasing the solubility of most gas in that pre-existing magma. Also, thermally or gas-induced density changes in the pre-existing magma cause gas-bearing magma from the base of the reservoir to rise and exsolve still more gas. Any combination of these effects can trigger an eruption.

4.12.6.1.2 External

4.12.6.1.2.(i) Regional earthquakes Following the 1992 Landers earthquake and a number of more recent earthquakes, local seismicity and strain increased at Long Valley and other large

calderas. Possible mechanisms for such response are discussed in Hill *et al.* (2002), Manga and Brodsky (2006), and (Chapter 4.09). Many involve geothermal waters.

Although most regional earthquakes do not trigger eruptions, nor are most eruptions triggered by earthquakes, there are several dozen known cases in which a strong earthquake was followed 'shortly' by an eruption, for example, eruptions of Cordon Caulle (Puyehue) in Chile within hours of major earthquakes in 1921 and 1960 (Katsui and Katz, 1967). 'Shortly' is relative to the normal repose periods (interevent times) of the volcano, hours versus decades in the case of Cordon Caulle. In 1917, San Salvador volcano began steaming immediately and erupted for several months after an M 6.5 earthquake occurred just west of the capitol city. Another apparent case of earthquake-induced eruption is that of Pinatubo, Philippines, which became restless within weeks and erupted months after a nearby M 7.8 earthquake, after (as best as is known) being quiet for the preceding 500 years.

How might regional earthquakes trigger eruptions? Immediate triggering, as in Chile, may involve local faulting that relieves confining pressure (Lara *et al.*, 2004), shaking-induced bubble nucleation (Manga and Brodsky, 2006), or in low viscosity hydrothermal water, shaking-induced rise and pressurization of bubbles (Linde *et al.*, 1994). Eruptions that are delayed by weeks or months may be induced if a dense mat of crystals is dislodged from the roof or walls of a body of magma, sinks, and displaces an equal volume of gas rich magma upward (Bergantz and Ni, 1999; Hill *et al.*, 2002; Manga and Brodsky, 2006). An alternate explanation, invoked in the case of Pinatubo (Bautista *et al.*, 1996), is that volumetric compression (or elsewhere, dilatation) as a result of a nearby earthquake could squeeze (or draw) magma upward. In the case of Pinatubo, +0.1 MPa static compression throughout the crust, ~35 km-deep long period earthquakes and clear evidence of fresh basalt intrusion into dacite beginning ~6 months after the earthquake and 3 months before magmatic eruptions, suggest that basaltic magma was squeezed upward. It might have risen as a diapir (discrete batch) all the way from the lower crust or it might have been squeezed and coalesced from interconnected magma-filled cracks throughout the thickness of the crust.

4.12.6.1.2.(ii) Unloading, rapidly as by a massive landslide off a volcano's flank, or slowly, as by glacial retreat Decrease of confining pressure induces exsolution of volatiles and the related chain of increased buoyancy and/or increased internal pressure. A classic example of an eruption induced by a large landslide is that of Mount St. Helens in 1980. A statistical increase in eruptions following glacial retreat has been postulated in Iceland, but some caution is needed to distinguish between real postglacial increases in volcanism versus better postglacial preservation of deposits (Maclennan *et al.*, 2002; Van Vliet *et al.*, 2005) (see also Jellinek *et al.* (2004)).

Mastin (1994) made a related proposal: that rainfall-induced fracturing of hot rock can induce phreatic and eventually magmatic eruptions, working from the top down.

4.12.6.1.2.(iii) Earth tides Some volcanoes show a slight statistical preference to erupt on or around fortnightly tidal maxima (or, in a few cases, fortnightly minima), and a few show a preference to erupt on or near semidiurnal tidal maxima (or minima) (Dzurisin, 1980; McNutt and Beavan, 1981, 1987; Emter, 1997). Most volcanoes whose eruptions can be triggered by earth tides have relatively open conduits, relatively low-viscosity magma, and frequent, small eruptions, for example, Kilauea, Pavlof, and Mayon. The exact mechanisms by which earth tides trigger eruptions are not known, but possibilities include slightly enhanced exsolution during semidiurnal tidal decompression and/or a slightly increased rate of magma ascent. Hydrofracturing might also be involved, as seismicity in areas of high stress and high hydrothermal pore pressures can also correlate with earth tides (Glasby and Kasahara, 2001).

Statistical preference to erupt on tidal maxima (or minima) does not indicate a significant increase of eruption probability at each tidal cycle. Volcanoes that are not ready to erupt will not be affected by tidal cycles. Only those volcanoes that are ready to erupt, and that require only the slightest additional push, might show an influence from the tides.

4.12.6.2 Explosive versus Effusive Eruptions: Fragmentation of Magma

When magma finally nears the surface, its gas content and viscosity will exert important controls over the style of eruption. An overly simplistic view is that

magma with high gas content will explode, producing ash and larger pyroclasts, while magma with low gas content will erupt gently (effusively), producing lava flows. A simple analogy is the initial opening of soda pop or champagne, charged with gas and explosive if opened suddenly. Once open, it starts to 'go flat' and, soon, can produce only a gentle overflow. Viscosity in its simplest role limits how easily gas can escape from magma, that is, how easily it can go flat. While gas is certainly a necessary component of explosive eruptions, and while viscous magma tends to erupt more explosively than fluid magma, neither all gas-rich magmas nor all viscous magmas will explode.

The explosivity of an eruption is an expression of the AMOUNT and RATES of exsolution and bubble growth (Sparks, 1978, 2003). As magma is rising up a conduit, these are controlled by the following:

- Its initial gas concentration and enroute leakage, the latter being controlled in large part by viscosity and the ability of bubbles to separate from melt. Wallrock permeability may also be a factor.
- Rate of confining pressure decrease, determined largely by ascent rate. Since ascent rate is not readily measured, it is inferred from the volume rate of extrusion. At Soufrière Hills volcano, explosivity was directly related to ascent and extrusion rate; faster ascent brought gas-rich but viscous magma to the near-surface faster than it could degas, so explosive eruptions ensued (Sparks and Young, 2002).

When magma reaches within several tens to hundreds of meters of the surface, the dominant controls on rates of exsolution and bubble growth become:

- the remaining volatile content, in melt and as discrete bubbles;
- any sudden rupture and release of confining pressure;
- bubble nucleation rate;
- melt viscosity, which limits
  - how fast dissolved volatiles can diffuse through melt into bubbles. Water in silicic to mafic melts diffuses slowly, from 10^{-1} to 10^{3} μm^2 s^{-1} (Zhang and Behrens, 2000; Zhang *et al.*, 2003; Behrens *et al.*, 2004);
  - how fast bubbles can expand;
  - whether glass as it is stretched by bubble expansion will fragment or continue to stretch; and
- permeability of any foam that develops.

As magma expands into foam, the melt walls of bubbles are stretched, usually faster than the melt can stretch. Two things can happen: either the magma blows itself apart, or it develops new shear permeability that bleeds off gas pressure before the magma can explode. In laboratory experiments with rapid decompression (intentionally rupturing a membrane in a pressure vessel), fragmentation occurs when bubbles expand rapidly and stretch bubble walls to their breaking point (Eichelberger *et al.*, 1986; Cashman and Mangan, 1994; Klug and Cashman, 1996; Dingwell, 1998, 2003; Cashman *et al.*, 2000; Spieler *et al.*, 2004). A typical vol.% of bubbles before fragmentation is ~70%, reachable just by expansion of 0.1 wt.% gas from 100 MPa (3 km) to 0.1 MPa (surface), and far exceeded with higher wt.%'s of gas. If this vol.% of bubbles is exceeded rapidly, the stretched molten glass walls of those bubbles break quickly and the magma fragments (blows itself to pieces). However, if the decompression rate is slower, equivalent to an ascent rate of magma of less than 0.35 m s^{-1}, shearing-induced permeability may prevent explosions (Burgisser and Gardner, 2005).

Nucleation and diffusion rates are a minor issue if most of the gas is already in separate bubble phase (in extreme cases, magma can have 100 times more volatiles in pre-ascent bubbles than dissolved in the melt). Where there is not already a separate bubble phase, the presence or absence of other impurities like crystals is important. With impurities as bubble nucleation sites, exsolution can begin immediately upon saturation, so there is a long, slow, weakly explosive process of 'heterogeneous' bubble nucleation and growth. Without impurities, magma becomes oversaturated with gas and then bubbles nucleate 'homogeneously', suddenly and explosively (Mangan and Sisson, 2000).

Thus, the explosivity and eruption – reflecting the rates of bubble nucleation, growth, and magma fragmentation – is a function of initial gas content and melt and magma viscosity; presence or absence of tiny microlites to cause heterogeneous bubble nucleation; the presence or absence of a pre-existing discrete bubble phase; and the effective permeability of the magma, magma foam, and

wallrock, and hence the efficiency with which magma can bleed off excess gas pressure shortly before and during eruptions.

4.12.6.3 Measures of Magnitude, Intensity (Volume DRE, Mass Discharge Rate, VEI)

The concepts of magnitude ('bigness') and intensity of eruptions are different than magnitude and intensity of earthquakes. Neither term has a precise definition in volcanology, but 'magnitude' refers broadly to the volume of products erupted and 'intensity' is the rate at which products are erupted. To be more precise, volume is usually expressed in terms of m^3 or km^3, with a qualifier to indicate whether that is the volume of deposit ('bulk volume', thickness times areal extent, including bubble and intergranular space), or the volume of magma that erupted ('DRE', or dense rock equivalent, corrected to a bubble free or prevesiculation volume of magma). Tsuya (1955) proposed the first volume-based magnitude scale for eruptions, and applied it to the sum of lavas and pyroclastic deposits.

Conversion from bulk volume back to DRE volume is made by multiplying bulk volume by the ratio of average density of deposit to the density of bubble-free magma. Typically, lavas have densities close to that of magma and pyroclastic deposits have densities around half that of magma.

Mass discharge rate is typically expressed in mass erupted/unit time, for example, $kg\,s^{-1}$. It is estimated by multiplying erupted volume times density and dividing by the duration of the eruption. A theoretical and empirical relation has been demonstrated in explosive eruptions between column height and mass discharge rate (Wilson, 1976; Wilson *et al.*, 1980; Carey and Sigurdsson, 1989), so the instantaneous intensity of an eruption can also be estimated from column height. Another empirical relation, between reduced displacement of eruption tremor and column height, was found to give another estimate of column height during eruptions (McNutt, 2000b).

To estimate volume and mass discharge rate requires mapping of deposits and/or detailed visual, satellite, or seismic observations during an eruption. This is possible for many modern eruptions but difficult to impossible for many older eruptions. To describe the explosive magnitude of older eruptions semiquantitatively, Newhall and Self (1982) proposed a volcanic explosivity index (VEI) that is based (in decreasing priority) on bulk volume of pyroclastic deposit, maximum column height, duration of eruption, and qualitative adjectives. Over all but the smallest eruptions, VEI rises one step with each order of magnitude increase in volume of deposit. As with earthquake and many other magnitude–frequency relations, there is a power-law decrease in frequency with increasing VEI (Simkin, 1993; Simkin and Siebert, 1994, 2000) (**Figure 7**). Not all eruptions that occur after a long repose (e.g., after hundreds to thousands of years) are of high VEI, but nearly all high VEI eruptions follow long reposes. Not counting short intervals between the start and climax of large explosive eruptions, it appears that a long repose is a necessary but not sufficient condition for a large explosive eruption, probably because it allows enough time for gas to accumulate in excess of saturation if the volcano is not leaky.

However handy VEI is as a shorthand among volcanologists, it is not a quantitative parameter, so bulk volume, mass discharge rate, and other details should be reported directly when known. Walker (1973) proposed a more quantitative classification of

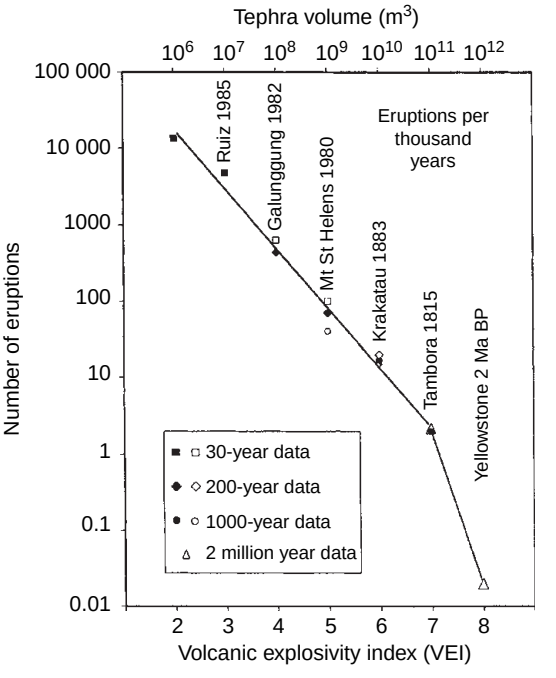


Figure 7 VEI magnitude–frequency plot. Over the range from VEI 2–8, each successively higher VEI is about 6 times less frequent than the next smaller category. A similar magnitude–frequency relation for earthquakes is well known. From Simkin T and Siebert L (1994) *Volcanoes of the World*, 349 p. Tucson, AZ: Geoscience Press.

explosive eruptions, based on two characteristics of their fall deposits:

1. fragmentation (F) = wt.% of pyroclasts <1 mm along dispersal axis, where thickness = 0.1 maximum thickness of deposit ($0.1\,T_{max}$).
2. dispersal (D) = area enclosed by ($0.01\,T_{max}$) isopach line of the tephra fall deposit.

Walker's fields of various eruption types on plots of (F) versus (D), as modified by Cas and Wright (1987) and reproduced here in **Figure 8**, correspond roughly to terminology that had already developed over preceding decades (plinian, strombolian, etc.), but define those eruption types more precisely.

Rutherford and Gardner (2000) and Cashman (2006) discuss the controlling effects of magma ascent rate, vesiculation, and degassing on explosivity. Ascent rates of ~0.5 m s^{-1} or higher inhibit degassing enroute to the surface and result in higher explosivity; those of ~0.05 m s^{-1} or lower encourage enroute degassing and result in relatively low explosivity. Cashman (2006) suggests correlation of magma supply rate to column height and dispersal (D), and proportion of erupted magma that is erupted explosively ('%tephra') to explosivity and fragmentation (F).

Further discussion of various measures of eruption 'bigness' may be found in Walker (1980), Carey and Sigurdsson (1989), and Pyle (2000).

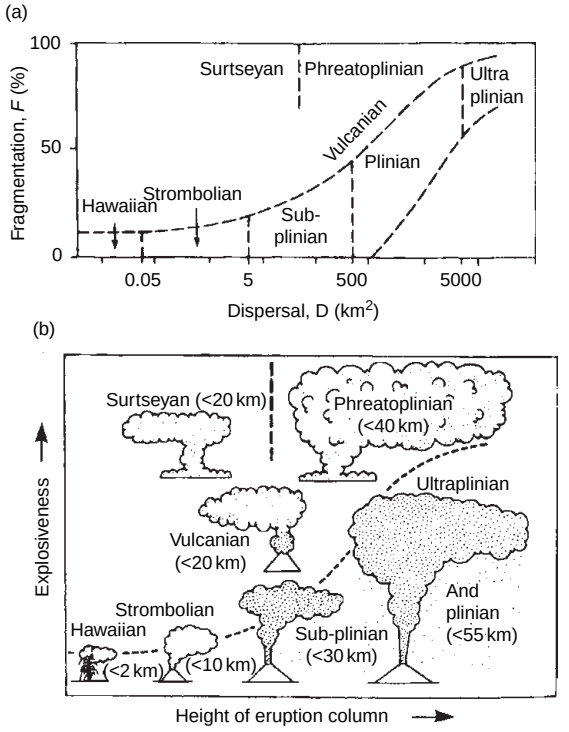


Figure 8 (a) Types of volcanic eruptions plotted in *F–D* space. F = fragmentation = wt.% of tephra fall that is <1 mm, along dispersal axis where $t = 0.1T_{max}$; D = dispersal = area enclosed by ($0.01T_{max}$) isopach line. T = thickness of tephra fall deposit. (from Cas and Wright 1987, after Walker 1973 and Wright *et al.* 1980). (b) Types of volcanic eruptions as functions of column height (itself, a proxy for mass discharge rate) and explosivity (as indicated by degree of fragmentation) (from Cas and Wright, 1987). Plot (a) is strictly descriptive; plot (b) is interpretive, though based on measurable proxies.

4.12.6.4 Eruptive Phenomena

4.12.6.4.1 Lava flows and lava domes

When magma pours or oozes from the ground, it is renamed 'lava'. Given a slope and a viscosity low enough to flow, lava will form lava flows. Most lava flows are elongate ribbons of lava with thickesses of 10^0–10^1 m, widths of 10^1–10^2 m, and lengths of 10^2–10^4 m. Flows an order of magnitude large are not uncommon. Flood basalts such as the Columbia River Basalt or the Deccan Traps are erupted from dikes but form broad, thick sheets up to 10^6 m long and wide. Basaltic and other low-silica, low viscosity flows can travel considerable distances; more silicic and thus more viscous lava might travel only a few kilometers from its vent.

A number of different terms are used to describe the surface features and flow details of lavas. Two of the most common are Hawaiian terms, *'aa'* and *'pahoehoe'*. *Aa* lava has a blocky, rough surface while *pahoehoe* has a smoother, ropy surface. Various factors contribute to the difference but the main factors are viscosity and shear rate. Higher viscosities and higher shear rates favor *aa*.

Extremely viscous lava – too viscous to flow more than a few tens or hundreds of meters from the vent – mounds up and forms 'lava domes'. Growth by extrusion of one or a succession of viscous flows is called exogenous; growth by inflation of an existing dome is called endogenous. An endogenous dome that grows just beneath and dramatically lifts the ground surface – soil, trees and all – is called a cryptodome ('hidden dome').

Nearly all lava flows travel slowly enough that people can walk out of harm's way. Buildings and other immovable objects are generally destroyed, either by fire or by mechanical bulldozing.

Thick, extensive flows of obsidian (volcanic glass) bear special mention. None has been observed closely during formation, but they pose an interesting

paradox. Glass forms by rapid quenching of melt upon eruption – usually, within seconds or minutes of eruption. The glassy character of obsidian implies rapid quenching. But many obsidian flows are so thick that they would have been expected to cool slowly, crystallizing rather than being quenched. Also, most obsidian flows are highly silicic and would have been expected to be more viscous than their broad dimensions suggest. Eichelberger *et al.* (1986) suggested the most likely explanation: that magma was erupted as a permeable foam, not unlike a good foam on a beer. As gas escaped, the foam collapsed into a quenched, dense glass that spread just a short distance farther. This gives the deceptive appearance of obsidian having flowed as lava over the entire distance from its vent, whereas in fact it was a much lower-viscosity foam over most of that travel distance.

4.12.6.4.2 Tephra fall

When an eruption is explosive, magma is blown out as fragments ranging in size from fine dust or sand ('ash') to large blocks or volcanic bombs. These fragments are called 'pyroclasts' (from the Greek for fire and broken) or 'tephra' (from the Greek for volcanic ash). Large fragments behave as ballistics, following parabolic trajectories and landing within a few hundred meters (rarely, up to a few kilometers) from the vent. Smaller fragments can be carried farther aloft by vigorous thermal convection. Eruption clouds from the largest explosive eruptions convect up to 50 km above the vent; small and moderate size eruption clouds (from VEI 1–3 eruptions) rise 0.1–20 km above the vent.

At altitude, ash or tephra-bearing clouds are typically sheared by prevailing winds and tephra rains out as a curtain elongate in the direction of the prevailing wind. For the most part, it rains out in order of decreasing particle size and density, following Stokes' law, but, if moist or electrically charged, the finest grain sizes clot together and fall out earlier than they otherwise would.

Light tephra fall (up to a few centimeters) does not directly threaten the lives of people on the ground, but can cause serious damage to computers and other electronics, electrical distribution grids, heating-ventilation-air conditioning (HVAC) systems, storm drains, and vulnerable crops. Tephra fall in excess of 10 cm, especially if wet, is the threshold beyond which structural damage and roof collapse is possible. Ash clouds pose a special hazard to modern jet aircraft, whose engines run at temperatures high enough to melt ingested ash. In the cooler parts of the turbine, molten ash glazes over fuel and air nozzles, engines shut down, and planes become gliders. Serious encounters, including near crashes of jumbo jets, have led to a new, global 24/7 (around-the-clock) network of Volcanic Ash Advisory Centers (Casadevall 1994, Romero 2004, NOAA 2005).

4.12.6.4.3 Pyroclastic flows and surges

Pyroclastic flows are avalanches of hot fragments and gas down the slopes of volcanoes, typically at 20–50 m s^{-1}. Most form by collapse of an eruption column, collapse of a dense slug of debris erupted just a few hundred meters above a vent, or collapse of the toe of a lava flow or dome growing on a steep slope. Pyroclastic flows have a relatively dense basal avalanche that exhibits laminar flow, and an accompanying, expanding turbulent ash and steam cloud. Ash from the basal avalanche is elutriated just like the dust cloud from a normal rock avalanche, and is almost as hot as the basal avalanche. Basal avalanches tend to follow valleys but can climb over some ridges; the overriding ash clouds are less constrained by topography. Runout distances are typically 1–20 km but in extreme cases can reach >100 km. Because of their high speed and high temperature, pyroclastic flows are generally lethal. The worst tragedy from a pyroclastic flow occurred in 1902 on the island of Martinique, where nearly all 29 000 residents of St. Pierre died in one short-lived but deadly flow.

Pyroclastic surges are relatively dilute cousins of pyroclastic flows, and indeed very similar to the upper, turbulent facies described above. The blast that leveled forest at Mount St. Helens was, in most senses, a pyroclastic surge. Increasingly, the term pyroclastic density current (PDC) is being applied to pyroclastic flows and surges.

4.12.6.5 Eruption Forecasts

Some volcanologists distinguish between forecasts (relatively general, often probabilistic statements based on past history) and predictions (more specific deterministic statements taking special account of monitoring and interpretation of current process) (e.g., Swanson *et al.*, 1983). In contrast, we see a spectrum of statements that are all probabilistic to some degree and thus 'forecasts', based on monitoring, past history, and interpretation of current process.

4.12.6.5.1 Long-range forecasts

Forecasts for eruptive activity within coming months to decades are typically based on eruptive history as seen in the geologic record and in historical documents. The past is the key to the future. In a few cases there might be robust enough clustering or other temporal patterns to infer something other than a Poisson distribution, but these are rare. From geologic and human history, we can outline the types, frequency, and probable extent of future eruptions. Such forecasts are heavily probabilistic. Hazard maps are one form of long-range forecasts, as are estimates of the probability that a particular site (e.g., a critical facility) will be hit by eruptive events within a specified time. Long-range forecasts are used for land-use planning, emergency response planning, facility siting, and other planning decisions.

4.12.6.5.2 Short-range forecasts

Forecasts for eruptive activity within minutes to months are based on a combination of geophysical, geochemical, and visual monitoring plus background information from the long-range forecast. Specific patterns of seismicity, ground deformation, gas emission, and other changes can indicate specific steps along the way toward eruption. In some cases, progress toward eruption gives a monotonic or exponential increase in monitored parameters; in other cases, values of a parameter might suddenly decrease, for example, seismic quiescence shortly before some eruptions. Patterns vary from volcano to volcano and, in some cases, from eruption to eruption, and the best short-range forecasts will be those that weigh empirical experience and real-time interpretation of processes together. Short-range eruption forecasts are analogous to weather forecasts.

When a volcano that is preparing to erupt is known to have exhibited a wide range of explosive scales in the past, can the explosive magnitude (VEI) of its impending eruption be forecast? At present, there is no tried and true method for forecasting VEI. Indeed, precursors of the giant 1991 eruption of Pinatubo were indistinguishable from precursors to much smaller eruptions until well into the eruption, just 24 h before its climax (Harlow *et al.*, 1996; Power *et al.*, 2001; Newhall *et al.*, 2002).

One approach to estimating explosive 'potential' (i.e., maximum possible VEI) of an impending eruption is to calculate a gas budget since the last time gas-rich magma was reamed out of the volcano (e.g., since the time of the previous VEI ≥ 3 eruption). If during the latest repose magma has degassed apace with inferred fresh volatile supply, the author's unpublished estimates suggest that the maximum VEI will be 4 and most probably less, but if volatiles have accumulated into a discrete volatile phase a VEI 3 or higher is almost certain and the maximum potential VEI could be 5 or higher. This approach is being tested as new eruptions permit, but it has not yet been validated. Within the bounds of maximum possible VEI, the actual VEI will depend greatly on magma ascent rate so any indication of rapid magma ascent, including rapid escalation of unrest, will favor higher rather than lower VEI.

Short-range forecasts can be of narrative form (what is expected, when, where, etc.), in the form of alert levels (e.g., color-coded alert levels; Gardner and Guffanti, 2006), or in the form of forecast windows that might be weeks long in an early or intermediate stage of unrest and days to hours long if seismicity and other parameters point to an imminent eruption (Swanson *et al.*, 1983).

One way to integrate short- and long-range forecasts is through use of event trees (Newhall and Hoblitt, 2002; Aspinall *et al.*, 2002; Marzocchi *et al.*, 2004, 2005; Hincks *et al.*, 2005). The thought process to estimate probabilities at each node of an event tree, and make Bayesian updates of these with each new piece of information, can be very helpful in guiding scientific discussion and further work.

4.12.6.5.3 Successes and false alarms

We have already seen that volcanoes give signals that magma is rising or otherwise preparing to erupt. Some particularly gratifying successes, such as at Pinatubo in 1991, have raised expectations. On the other hand, some eruptions still occur without being forecast, from volcanoes that are not monitored at all or at which the precursors were too subtle for current monitoring. Relatively fewer cases occur in which warning signs are detected but underestimated or otherwise misinterpreted.

A bigger challenge for volcanologists is an overabundance of precursors, and how to distinguish those that will lead to eruption from those which will not. Signs that magma is rising toward the surface are similar regardless of eventual outcome. At an early stage of unrest, we may see unmistakable signs of magma rising but not know yet whether it will stall or eventually erupt, and intrusion may continue until practically the 'last minute' before stalling, as occurred at Soufrière Guadeloupe in 1976 (Feuillard *et al.*, 1983; Fiske, 1984). Empirically, magma that is rising rapidly, with corresponding

accelerating unrest, is more likely to erupt. Nishimura (2006) modeled accelerating ground deformation from undegassed magma, whereas degassed magma produced constant-rate deformation. There may be other differences between successful versus failed eruptions, also related to magma ascent rate and gas overpressures, but these are still in the research arena.

In 1999, precursors warned of an approaching eruption of Tungurahua volcano in Ecuador. The town of Baños, extremely vulnerable, was evacuated. The volcano did erupt and still erupts, but until 2006 the eruption was more spectacular than dangerous. The townspeople returned home after only a few months of evacuation and were loathe to ever evacuate again. It is very hard for a lay public to understand how degassing will increase magma viscosity and explosive potential, and that it will be very hard for the scientific monitoring team to know which of the daily explosions will be bigger than the rest and produce a pyroclastic flow. It is equally hard for the public to visualize that the deposits on which their town is built are from pyroclastic flows not much older than the town! Fortunately, stronger explosive activity in 2006 had clear precursors, clear warnings were issued, and those at highest risk (just outside Baños) did evacuate in time.

A stark illustration of another false alarm dilemma comes from Vesuvius and towns around its base. The population is so large (approximately 550 000 in the 'red' danger zone) that an evacuation will require at least 1 week. Yet, it is rarely possible to be sure that a volcano will erupt until just hours to days before an eruption. The earlier the evacuation, the better the chances that it will be successful, but also the greater the chances of false alarm.

4.12.7 Erosion of Volcanoes

Here is a brief introduction to hydrothermal alteration, sector collapse, and other processes that erode volcanoes.

4.12.7.1 Hydrothermal Alteration

Every volcano, to one degree or another, has a hydrothermal system – a system of groundwater heated by magma and circulating convectively through fractures and porous rock. Hydrothermal fluid that is heated directly by volcanic gases is acidic, and reacts with country rock to form clays and other alteration minerals plus a more neutral hydrothermal fluid. Largely hidden from view, hydrothermal alteration softens and weakens a volcanic edifice and makes it prone to large-scale collapse. Some volcanoes, for example, Yellowstone and Rainier have unusually extensive hydrothermal alteration. At Rainier, hydrothermal alteration weakens the steep slopes and has been partly responsible for some giant landslides (flank collapse) that evolved downslope into clay-rich debris flows.

4.12.7.2 Sector Collapses

Two distinctive geomorphic features – horseshoe-shaped craters and fields of small hills ('hummocks') at their feet – were recognized at a handful of volcanoes before 1980. Some were thought to be the result of large, possibly slow-moving landslides; others, including small hills north of Mount Shasta, were not understood at all. Immediately upon the giant landslide of Mount St Helens in 1980, hundreds of such pairs were found and understood at once to have formed in the same, catastrophic way (Ui, 1983; Siebert *et al.*, 1987; Siebert, 1996). Subsequent work at Mount St. Helens confirmed at least three previous collapses, two from the north flank about 3 ka BP and another of the south flank ~20 ka BP (Hausback, 2000).

Although the first parallels were noticed from steep-sided arc stratovolcanoes, workers in Hawaii soon realized that even gently sloping shield volcanoes have also collapsed, in landslides from hundreds and even up to 5000 km^3 (Moore *et al.*, 1994) Just one 10 min long event – the landslide of Mount St. Helens – changed the paradigm of thinking of volcanoes as fundamentally stable with rare exceptions to fundamentally unstable edifices that usually fail at least once in their lifetime.

4.12.7.3 Concentrated Sediment-Water Flows (Lahars)

Concentrated flows of volcanic sediment and water, called by the Indonesian term 'lahars', form by the following:

- eruption through crater lake (e.g., Kelut, Indonesia), followed by scouring of downstream deposits;
- sector collapse of water-saturated cone, followed by either downstream evolution of avalanche into

debris flow (e.g., Ontake, Japan; Mount Rainier, USA), or halting and compaction dewatering of the debris avalanche deposit (e.g., Mount St. Helens, USA);
- rapid melting of snow and ice by pyroclastic flows and incorporation of sediment from the pyroclastic flows and also from scoured downstream deposits (e.g., Nevado del Ruiz, Colombia);
- overtopping and rapid scouring failure of natural dams (e.g., Ruapehu, New Zealand; Pinatubo, Philippines); and
- heavy rainfall on fresh pyroclastic deposits (e.g., Pinatubo, Philippines). Fine ash on the surface reduces infiltration rates by an order of magnitude and watersheds become prone to sediment-laden flash floods, just as in arid climates or following forest fires.

The first three originate only during eruptions. The latter two can be syneruptive or posteruptive.

Lahars behave in ways quite unlike normal streamflow. Bulking up from clear water streamflow to >20 vol.% sediment creates hyperconcentrated flows, and further bulking up to >60 vol.% sediment creates debris flows, both termed lahars. Hyperconcentrated flow is dilute enough that it has no appreciable yield strength, yet it transports considerable volumes of sediment in short periods. Hyperconcentrated flow is turbulent and deposition from that flow is progressive (grain by grain) though fast enough that classical stream deposit features such as cross-bedding are absent. In contrast, a debris flow exhibits significant yield strength and laminar flow, the latter helping it to flow fast and far. Debris flows are analogous to fast-moving slurries of concrete, and the combination of their density and speed allows them to plow right through (or lift and float) buildings, bridges, and anything else in the way. Deposition from debris flows can range from progressive to sudden, *en masse* deposition, and resulting deposits are massive and without any internal stratification.

Lahars generate relatively high-frequency tremor and dense seismic networks can detect and track lahars. Inexpensive 10–300 Hz geophones installed on riverbanks (LaHusen, 2005) were used with good success to detect and quantify lahars at Pinatubo (Marcial *et al.*, 1996; Tungol and Regalado, 1996), and are presently deployed at a number of volcanoes including Rainier and Ruapehu to warn of approaching lahars.

4.12.7.4 Normal Streamflow, Sediment Issues, and Flooding

Lahars can transport very large amounts of sediment in the years following a large explosive eruption. In the case of Pinatubo, >3 km^3 was eroded from the roughly 5–6 km^3 of 1991 pyroclastic debris on the slopes of the volcano (Tungol, 2002). Most of this sediment was transported as lahars in the first 5 posteruptive years, but by 1997, 25% of declining annual sediment transport was by normal streamflow (Hayes *et al.*, 2002). Except for one large lake-breakout lahar in 2002, nearly all sediment transport since 2000 has been by normal streamflow. High sediment yields were also observed in the decades following the 1980 eruption of Mount St. Helens (Major, 2004), where torrential rains are rare and most sediment was carried by normal streamflow.

Copious loads of sediment are deposited in channels on the surrounding lowlands, sharply reducing their capacity to carry even modest normal runoff. Lowland flooding is a serious by-product of lahars and other heavy sediment transport after an explosive eruption.

4.12.8 People and Volcanoes

Demographers have long-recognized dense populations around tropical and temperate volcanoes. Reasons include fertile soils, healthy cooler climates above tropical lowlands, and esthetic and recreation appeal. The result is an unhealthy juxtaposition of large populations with dangerous volcanic phenomena, and an extra challenge for those who must forecast eruptions and avoid false alarms (Ewert and Harpel, 2004).

4.12.8.1 Risk Mitigation

The principal tools for long-range risk mitigation are hazard and risk maps (the latter, integrating hazard, vulnerability, and exposure), land use management, public education and dialog about volcanic risks, and early development and testing of emergency plans. The principal tools for immediate risk mitigation are short-range eruption forecasts, evacuations, aircraft reroutings, and other emergency measures such as installing ash filters and covering storm drains.

Can people, buildings, and infrastructure be protected against ashfall? In general, yes, though at some cost. Can flow hazards be stopped by engineering measures? The answer is a qualified 'sometimes'. If topography is favorable, lava flows can be slowed or

diverted away from critical structures. If topography is favorable and considerable time and money is invested in advance, lahars can be slowed and sediment can be sequestered by dams and related structures. Even bunker-like shelters from pyroclastic flows have been tried. However, the success rate of such efforts to control volcanic flows is low and costs are high. Usually, it makes more sense simply to avoid building in the path of volcanic flows and to move out of the way if threatened.

4.12.8.2 Probabilities and Acceptable Risk

Communities and individuals generally resist evacuation and other costly mitigation measures, especially if eruptions of that volcano are infrequent or if forecasts are conspicuously uncertain. In some cases, this stems from ignorance or denial in the face of clear and present danger, but in other cases it stems from a calculated decision to accept the risk from the volcano rather than take socially and economically costly mitigation measures. Quantification of volcanic risk, even with high uncertainties, usually helps citizens and officials to decide whether that risk is acceptable to them. Further discussion of this can be found in Newhall and Hoblitt (2002) and Aspinall *et al.* (2002), Hincks *et al.* (2005).

4.12.9 Future Directions

Brodsky (2004) offered interesting thoughts on outstanding challenges in volcanology. The author concurs, and adds these additional thoughts and questions.

What causes magmas to pond and accumulate in reservoirs with tops between 3 and 6 km?

This has been termed a region of 'neutral buoyancy' by Ryan (1987). Is this the level at which decompression degassing increases viscosity sharply and causes magmas to stall? Or is 3–6 km a region in which rising magma encounters significant groundwater and is thus quenched and slowed and/or can spread by hydrofracturing of country rock? Is this a level at which lithostatic pressure that is squeezing magma upward becomes too low to force open new fractures toward the surface? Or is it a level where exsolution of volatiles forms a barrier of low-density magma that blocks ascent of following magma? Understanding when and where each of these processes is dominant may help us to understand periodicity of shallow intrusions and eruptions, and to understand how long-dormant volcanoes reawaken.

Is magma supply from the mantle and lower crust continuous and are eruptions thus controlled by buffering in shallow reservoirs or control by near-surface valves? Or, alternatively, is supply into volcanic roots episodic and, itself, a major control on when and how volcanoes erupt?

One way to address this question is by InSAR monitoring of broad deformation, indicating resupply of magma into volcanic roots. With C-band radar, only a limited number of unvegetated high-latitude or high-altitude volcanoes could be monitored. However, now that L-band radar is once again available from Japan's ALOS/PALSAR satellite, we will see good data from many volcanoes, even in the tropics.

Supply of magma from depth is probably episodic if viewed to the resolution of hours to years. But as the viewing resolution blurs to decades or centuries, supply might be more or less steady. One hypothesis to be tested is that if supply is episodic and less frequent than the time required for magma in the conduit to degas, cool, and solidify after an eruption, then eruptions will occur only after fresh magma resupply and perhaps only after repeated episodes of resupply. If supply is more or less steady on the timescale of solidification of magma in conduits, then solidification will occur only when the rate of magma supply is relatively low. When supply rates are higher, conduits will stay open and filled with convecting magma, and eruptions can occur at anytime or with just the slightest fresh resupply.

How, in detail, does magma degas, and how does this control eruptive style and magnitude?

This topic was introduced earlier in two contexts: leakage during magma ascent from depth if that ascent is slow enough, and leakage during the moment of eruption (modulated expansion of magma foam) if the decompression is slow enough to allow development of shear permeability before the magma explodes. Both of these topics invite further laboratory experimentation and field confirmation. There are great opportunities for more interaction between those who measure gas emissions in the field, those who study degassing and fragmentation in the laboratory, and those who reconstruct it from the petrology of erupted products.

Specifically, field and lab investigators can test the working hypothesis that volcanoes which are plugged tight (i.e., with minimal inter-eruption passive degassing) accumulate gas in excess of saturation and have

high explosive potential, especially if their eventual ascent to the surface is rapid. Did all magmas that fed plinian eruptions have an excess volatile phase? Data with which the author is familiar suggest yes, but any exceptions will be even more interesting! At volcanoes with open conduits, what causes the occasional explosions that are much larger than normal? Is it simply that the convection is speeded up, or is homogeneous nucleation also involved?

Do early bubbles in gas-charged magma, acting as potential sites for further heterogeneous nucleation, ever reduce explosivity? In relatively viscous magmas, what is the relative importance of foaming versus shear permeability as a means to efficiently degas magma? How was strong degassing like that of the past decade at Popocatépetl sustained even when little or no magma was erupting? How does Iwo-jima degas so efficiently and, in particular, might groundwater and hydrofracturing play a role? What independent evidence can be found to support or reject lava-lamp style conduit convection, and associated degassing at the top of an inferred 'convection cell?' Combined geologic, geophysical, and geochemical studies are needed to improve forecasts of explosive magnitude (VEI) so that evacuations and other measures are of appropriate scale. For example, at Tungurahua volcano in Ecuador in 1999, the geologic record warned of pyroclastic flows and the population of Baños was briefly evacuated, but no major pyroclastic flows occurred until 2006. What can be monitored in small, strombolian explosions that will foretell of whether and when the eruption will escalate into a more explosive one? The same general issue of forecasting VEI arises during unrest at any volcano that exhibits a range in the VEIs of its eruptions. Can estimation of a gas budget, as in **Figure 3**, help in estimating the VEI of an impending eruption?

At what timescales are eruptions stochastic versus predictable?

Eruptions may seem to occur randomly or predictably, depend on the timescale. Many variables influence whether a volcano will erupt in a given time period (e.g., history of magma supply, gas leakage, local stress–strain relations), so on longer timescales, eruptions may indeed be stochastic. At the same time, systematic buildup of precursors before eruptions argues for predictability. On timescales of weeks to months, volcanoes with open conduits, erupting frequently, may appear to be stochastic while those with closed conduits, erupting infrequently, appear predictable. On the other hand, if one looks closely enough at those with open vents, they, too, exhibit precursors and predictability on timescales of seconds to minutes. For forecast purposes, volcano monitoring must be scaled to the openness of the vent.

How can we best estimate critical intrinsic parameters (e.g., viscosity, water content, overpressure) from observable parameters (seismicity, deformation, gas leaks, etc.). Can active probes help?

A conundrum facing volcanologists is that the parameters which directly indicate explosive potential (volatile content of magma, gas overpressures, magma viscosity) cannot be measured before eruptions. What can be measured – seismicity, deformation of the ground surface, volumetric strain, gas emissions, ground or fumarole temperatures, and a few other parameters – are at best indirect indicators of the state of the magma.

Two general approaches might help. One is to track the volcano's changes (seismic, ground deformation, volumetric strain, gravity, gas) in response to known, repeating natural signals such as earth tides. In response to tidal decompression, we might expect greater expansion response if shallow magma is charged with a separate volatile phase (high explosive potential) than if it is gas poor. Less viscous magmas might respond more quickly than more viscous magmas. Microseismic signals and hour-to-hour gas emissions might vary with volatile concentration, state, and magma permeability.

As a second approach, we might try artificial sources. As a magma becomes charged with bubbles, and especially as those expand as gas, seismic velocities through the magma will slow down, but the general problem is how to focus a repeating source through relatively small target bodies of magma? Stacking of repeating artificial signals, such as in the Japanese ACROSS project (Ikuta *et al.*, 2002; Ikuta and Yamaoka, 2004), is promising but has not been proved at volcanoes yet. Other ideas are needed!

How can we know if an impending eruption from a large volcanic system will be just a small aliquot or a colossal eruption?

Large magnitude eruptions are thought to grow from small ones by runaway vesiculation (bubble nucleation and growth), just as large earthquakes may grow from small ones by cascading rupture. Volcanic gas is concentrated in the uppermost parts of magma reservoirs. An eruption might tap just a small volume of gas-charged magma from the top of a reservoir, if the gas content of underlying magma is too low and/or the vent and conduit get clogged quickly. On the other hand, if the gas content is high,

and the bleed rate is too low, an explosive eruption may escalate. Runaway vesiculation may increase the volume and mass discharge rate of an eruption to exceptionally high values. Hoblitt *et al.* (1996) interpreted the pattern of more and more closely spaced, smaller and smaller eruptions of Pinatubo in 1991 leading up to the runaway climactic phase as a result of tapping magma with increasing gas concentrations, which systematically reduced the time between explosive pulses. At the same time, the vent was also getting reamed out, also decreasing resistance to eruption. Eventually, the pulse that began at 1342 h on June 15 tapped magma that was rich enough in gas to sustain and escalate the discharge. Harlow *et al.* (1996) reported a parallel escalation in the energy release of shallow long-period earthquakes. The runaway eruption continued until the eruption tapped into deeper, gas-depleted magma, and/or tapped so deep in the conduit that overlying foam and debris shut off vesiculation, or the conduit was squeezed shut with the same effect.

We do not know if the patterns observed by Harlow *et al.* (1996) and Hoblitt *et al.* (1996) are characteristic of the buildup to large (plinian) explosive eruptions. If they are, then we might have hours of forewarning of larger explosive events. From a practical point of view, populations must already be evacuated on the basis of any geologic evidence of past plinian eruptions, as the confirming indications from monitoring might come too late for evacuations.

4.12.10 Conclusion

This review began with analogies between volcanoes and fault zones. Both are subject to stress and strain, both produce distinctive patterns of seismicity, and both exhibit 'stick-slip' behavior with similar power-law magnitude–frequency relations. In the course of this review we have also seen a further analogy: that the variable degassing pattern of volcanoes, some open and freely degassing versus others plugged tight, is analogous to creeping versus locked segments of faults. Gas accumulation in magma is analogous to strain accumulation on faults, steadily accumulating unless released by passive leakage or creep. Estimation of explosive potential demands an accounting of gas accumulation versus passive degassing, over the timeframe since the previous explosive eruption (typically, years to decades) and also over the timeframe of current magma ascent (typically, hours to weeks).

References

Aki K, Fehler M, and Das S (1977) Source mechanism of volcanic tremor: Fluid-driven crack models and their application to the 1963 Kilauea eruption. *Journal of Volcanology and Geothermal Research* 2: 259–287.

Allard P (1977) Endogenous magma degassing and storage at Mount Etna. *Geophysical Research Letters* 24(17): 2219–2222.

Allard P, Carbonnelle J, Dajlevic D, Metrich N, and Sabroux J (1995) The volatile source and magma degassing budget of Merapi volcano: Evidence from high temperature gas emissions and crystal melt inclusions, 66 p. Merapi Decade Volcano Workshop, Yogyakarta, Indonesia.

Allard P, Carbonnelle J, Metrich N, Loyer H, and Zettwoog P (1994) Sulphur output and magma degassing budget of Stromboli Volcano. *Nature* 368: 326–330.

Anderson AT, Jr. (1974) The before-eruption water content of some high-alumina magmas. *Bulletin of Volcanology* 37: 530–552.

Andres RJ, Rose WI, Kyle PR, *et al.* (1991) Excessive sulfur dioxide emissions from Chilean volcanoes. *Journal of Volcanology and Geothermal Research* 46: 323–329.

Arpa MCB, Bornas MAV, Abigania MIT, *et al.* (2006) Characterization of lava flow and pyroclasts from the February to March 2000 eruption of Mayon Volcano, Philippines. *Journal of the Geological Society of the Philippines* 61: 1–19.

Aspinall WP, Loughlin SC, Michael FV, *et al.* (2002) The montserrat volcano observatory: Its evolution, organization, role and activities. In: Druitt TH and Kokelaar BP (eds.) *Geological Society of London, Memoir 21: The Eruption of Soufriere Hills Volcano, Montserrat from 1995 to 1999*, pp. 71–91. Bath, UK: Geological Society Publishing House.

Balmforth NJ, Craster RV, and Rust AC (2005) Instability in flow through elastic conduits and volcanic tremor. *Journal of Fluid Mechanics* 527: 353–377.

Bautista BC, Bautista MLP, Stein RS, *et al.* (1996) Relationship of regional and local structures to Mount Pinatubo activity. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, pp. 351–370. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.

Behrens H, Zhang Y, and Xu Z (2004) H_2O diffusion in dacitic and andesitic melts. In: Xue X, Kanzaki M, Neuville DR, and Kawamoto T (eds.) Structure and properties of silicate melts and fluids. *Geochemica et Cosmochimica Acta* 68: 5139–5150.

Bergantz GW and Breidenthal RE (2001) Non-stationary entrainment and tunneling eruptions: A dynamic link between eruption processes and magma mixing. *Geophysical Research Letters* 28(16): 3075–3078.

Bergantz GW and Ni J (1999) Numerical study of sedimentation by dripping instabilities in viscous fluids. *International Journal of Multipase Flow* 25: 307–320.

Bernard A (2004) Monitoring the temperature of Taal crater lake using ASTER TIR images (abs.) IAVCEI 2004, Pucon, Chile, Session 6c, http://iavcei2004.free.cl (accessed Feb 2007).

Blake S and Fink JH (2000) On the deformation and freezing of enclaves during magma mixing. *Journal of Volcanology and Geothermal Research* 95: 1–8.

Bluth GJS, Rose WI, Sprod IE, and Krueger AJ (1997) Stratospheric loading of sulfur from explosive volcanic eruptions. *Journal of Geology* 105: 671–683.

Brodsky EE (2004) International Association of Volcanology and Chemistry of the Earth's Interior. In: IUGG Working Group *Geosciences: The Future*. International Union of Geodesy and Geophysics http://www.iugg.org/publications/reports/geosciences.pdf, pp. 23–27.

Burgisser A and Gardner JE (2005) Experimental constraints on degassing and permeability in volcanic conduit flow. *Bulletin of Volcanology* 67: 42–56.

Carey SN and Sigurdsson H (1989) The intensity of plinian eruptions. *Bulletin of Volcanology* 51: 28–40.

Carn SA, Krueger AJ, Bluth GJS, *et al.* (2003) Volcanic eruption detection by the Total Ozone Mapping Spectrometer (TOMS) instruments: A 22-year record of sulphur dioxide and ash emissions. In: Oppenheimer C, Pyle DM, and Barclay J (eds.) *Geological Society of London, Special Publication 213*: *Volcanic Degassing*, pp. 177–202. London: Geological Society of London.

Carroll MR and Holloway JR (eds.) (1994) *Reviews in Mineralogy, vol. 30: Volatiles in Magmas*, 517 p. Washington, DC: Mineralogical Society of America.

Cas RAF and Wright JV (1987) *Volcanic Successions: Ancient and Modern*. London: Allen and Unwin.

Casadevall TJ (ed) (1994) Volcanic ash and aviation safety. Proceedings of the First International Symposium on Volcanic Ash and Aviation Safety, US Geological Survey Bulletin 2047: 450.

Cashman KV (2006) Bowen Lecture: Bubbles, crystals, and melts – 'the three-phase world of volcanic eruptions' (abs), Eos Transactions AGU Fall Meeting Supplement 87(52): V21E–02.

Cashman KV and Mangan MT (1994) Physical aspects of magmatic degassing. Part II: Constraints on vesiculation processes from textural studies of eruptive products. In: Carroll MR and Holloway JR (eds.) *Reviews in Mineralogy 30: Volatiles in Magma*, pp. 447–478. Washington, DC: Mineralogical Society of America.

Cashman KV, Sturtevant B, Papale P, and Navon O (2000) Magmatic fragmentation. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 421–430. San Diego, CA: Academic Press.

Cervelli P (2000) DISLOC3D, Displacements and their derivatives from a dislocation in an elastic half-space (software based on Okada 1992). Version 1.0, September 22 2000, http://www.avo.alaska.edu/Software/Kilauea/disloc3d.html (accessed Feb 2007).

Chouet B (1992) A seismic model for the source of long-period events and harmonic tremor. In: Gasparini P, Scarpa R, and Aki K (eds.) *Volcanic seismology*, pp. 133–156. Berlin: Springer.

Chouet B (1996a) Long-period seismicity: Its source and use in eruption forecasting. *Nature* 380: 309–316.

Chouet BA (1996b) New methods and future trends in seismological volcano monitoring. In: Scarpa R and Tilling RI (eds.) *Monitoring and Mitigation of Volcano Hazards*, pp. 23–97. Berlin: Springer.

Coombs ML, Eichelberger JC, and Rutherford MJ (2000) Magma storage and mixing conditions for the 1953–1974 eruptions of Southwest Trident volcano, Katmai National Park, Alaska. *Contributions to Mineralogy and Petrology* 140: 99–118.

Coombs ML and Gardner JE (2004) Reaction rim growth on olivine in silicic melts; implications for magma mixing. *American Mineralogist* 89: 748–759.

Cooper KM and Reid MR (2003) Re-examination of crystal ages in recent Mount St. Helens Lavas: Implications for magma reservoir processes. *Earth and Planetary Science Letters* 213: 149–167.

Daag AS, Tubianosa BS, Newhall CG, *et al.* (1996) Monitoring sulfur dioxide emission at Mount Pinatubo. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 409–414. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.

Day SJ (1996) Hydrothermal pore fluid pressure and the stability of porous, permeable volcanoes. In: McGuire WJ, Jones AP, and Neuberg J (eds.) *Geological Society of London Special Publication 110: Volcano Instability on the Earth and other Planets*, pp. 77–93. London: Geological Society of London.

Devine JD, Rutherford MJ, and Gardner JE (1997) Petrologic determination of ascent rates for the 1995–1997 Soufriere Hills Volcano andesitic magma. *Geophysical Research Letters* 25: 3673–3676.

Devine JD, Rutherford MJ, Norton GE, and Young SR (2003) Magma storage region processes inferred from geochemistry of Fe–Ti oxides in andesitic magma, Soufriere Hills Volcano, Montserrat, W.I. *Journal of Petrology* 44: 1375–1400.

Dingwell DB (1998) Recent experimental progress in the physical description of silicic magma relevant to explosive volcanism. In: Gilbert JS and Sparks RSJ (eds.) *Geological Society of London, Special Publication 145: The Physics of Explosive Eruption*, pp. 9–26. London: Geological Society of London.

Dingwell DB (2003) Liquid to glass: Quantifying properties and structures of melts across the glass transition. In: De Vivo B and Bodnar RJ (eds.) *IAVCEI Developments in Volcanology 5: Melt Inclusions in Volcanic Systems: Methods, Applications, and Problems*, pp. 45–63. Amsterdam: Elsevier.

Dobran F (2001) *Volcanic Processes: Mechanisms in Material Transport*, 590 p. Kluwer New York: Academic/Plenum.

Doukas MP and Gerlach TM (1995) Sulfur dioxide scrubbing during the 1992 eruptions of Crater Peak, Mount Spurr Volcano, Alaska. In: Keith TEC (ed.) *US Geological Survey Bulletin 2139: The 1992 Eruptions of Crater Peak Vent, Mount Spurr Volcano, Alaska*, pp. 47–57. Denver, CO: US Geological Survey.

Dunbar NW, Cashman KV, and Dupré R (1994) Crystallization processes of anorthoclase phenocrysts in the Mount Erebus magmatic system: Evidence from crystal composition, crystal size distributions, and volatile contents of melt inclusions. In: Kyle PR (ed.) *America Geophysical Union Antarctic Research Series: Volcanological and Environmental Studies of Mount Erebus, Antarctica*, pp. 129–146. Washington, DC: American Geophysical Union.

Dzurisin D (1980) Influence of fortnightly earth tides at Kilauea Volcano, Hawaii. *Geophysical research Letters* 7: 925–928.

Dzurisin D (2003) A comprehensive approach to monitoring volcano deformation as a window on the eruption cycle. (corrected) *Reviews of Geophysics* 41(2): 1009 (doi:10.1029/2003RG000134).

Dzurisin D (ed.) (2007) *Volcano Deformation: Geodetic Monitoring Techniques*, 441 p. Berlin: Springer-Praxis.

Dzurisin D, Lisowski M, Moran SC, Wicks CW, Poland MP, and Endo ET (2004) An ongoing episode of magmatic inflation at the Three Sisters Volcanic Center, central Oregon Cascade Range: Inferences from recent geodetic and seismic observations (abs.) Eos Transactions AGU, Fall Meeting Supplement Abstract 85(47): G42A–04.

Dzurisin D and Lu Z (2007) Interferometric synthetic-aperture radar (InSAR). In: Dzurisin D (ed.) *Volcano Deformation: Geodetic Monitoring Techniques*, pp. 153–194. Berlin: Springer-Praxis.

Eichelberger JC (1980) Vesiculation of mafic magma during replenishment of silicic magma reservoirs. *Nature* 288(5790): 446–450.

Eichelberger JC (1995) Silicic volcanism; ascent of viscous magmas from crustal reservoirs. *Annual Reviews of Earth and Planetary Sciences* 23: 41–63.

Eichelberger JC, Carrigan CR, Westrich HR, and Price RH (1986) Non-explosive silicic volcanism. *Nature* 323(6089): 598–602.

Emter D (1997) Tidal triggering of earthquakes and volcanic events. In: Wilhelm H, Zürn W, and Wenzel H-G (eds.) *Tidal Phenomena*, pp. 293–309. Berlin: Springer-Verlag.
Ewert JW and Harpel CJ (2004) In harm's way: Population and volcanic risk. *Geotimes* 49: 14–17.
Fedotov SA, Gorelchik VI, and Stepanov VV (1980) Seismological studies on the mechanism of the Large Tolbachik fissure eruption, 1975–1976. *Bulletin of Volcanology* 43(1): 73–84.
Fehler M (1983) Observations of volcanic tremor at Mount St. Helens volcano. *Journal of Geophysical Research* 88: 3476–3484.
Feuillard M, Allegre CJ, Brandeis G, *et al.* (1983) The 1975–1977 crisis of La Soufriere de Guadeloupe (F.W.I.): A still-born magmatic eruption. *Journal of Volcanology and Geothermal Research* 16: 317–334.
Fink JH, Anderson SW, and Manley CR (1992) Textural constraints on effusive silicic volcanism: Beyond the permeable foam model. *Journal of Geophysical Research* 97: 9073–9083.
Fink JH and Manley CR (1987) Origin of pumiceous and glassy textures in rhyolite flows and domes. In: Fink JH (ed.) *Geological Society of America Special Paper 212: The Emplacement of Silicic Domes and Lava Flows*, pp. 77–88. Boulder, CO: Geological Society of America.
Fischer TB, Arehart GB, Sturchio NC, and Williams SN (1996) The relationship between fumarole gas composition and eruptive activity at Galeras Volcano, Colombia. *Geology* 24: 531–534.
Fiske RS (1984) Volcanologists, journalists, and the concerned local public: A tale of two crises in the eastern Caribbean. In: National Research Council, Geophysics Study Committee (ed.) *Explosive Volcanism*, pp. 110–121 Washington, DC: National Academy Press.
Fitterman DV (1978) Electrokinetic and magnetic anomalies associated with dilatent regions in a layered earth. *Journal of Geophysical Research* 83: 5923–5928.
Fournier RO (1999) Hydrothermal processes related to movement of fluid from plastic into brittle rock in the magmatic-epithermal environment. *Economic Geology* 94: 1193–1211.
Fournier RO (2007) Hydrothermal systems and volcano geochemistry. In: Dzurisin D (ed.) *Volcano Deformation: Geodetic Monitoring Techniques*, 323–341. Berlin: Springer-Praxis.
Francis P and Oppenheimer C (2004) *Volcanoes,* 2nd ed, 521 p. Oxford, New York: Oxford Press.
Gardner CA and Guffanti MC (2006) US Geological Survey's alert notification system for volcanic activity. US Geological Survey Fact Sheet 2006–3139, 4 p. (http://pubs.usgs.gov./fs/2006/3139/).
Gerlach TM, Doukas MP, McGee KA, and Kessler R (1998) Three-year decline of magmatic CO_2 emissions from soils of a Mammoth Mountain tree kill, Horseshoe Lake, California, 1995–1997. *Geophysical Research Letters* 25: 1947–1950.
Gerlach TM, Westrich HR, and Symonds RB (1996) Preeruption vapor in magma of the climactic Mount Pinatubo eruption: Source of the giant stratospheric sulfur dioxide cloud. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 415–433. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.
Giggenbach WF (1996) Chemical composition of volcanic gases. In: Scarpa R and Tilling RI (eds.) *Monitoring and Mitigation of Volcano Hazards*, pp. 221–256. Berlin: Springer.
Gil Cruz F and Chouet BA (1997) Long-period events, the most characteristic seismicity accompanying the emplacement and extrusion of a lava dome in Galeras Volcano, Colombia, in 1991. *Journal of Volcanology and Geothermal Research* 77: 121–158.
Glasby GP and Kasahara J (2001) Influence of tidal effects on the periodicity of earthquake activity in diverse geological settings with particular emphasis on submarine hydrothermal systems. *Earth-Science Reviews* 52(4): 261–297.
Gómez DM and Torres RA (1997) Unusual low-frequency volcanic seismic events with slowly decaying coda waves observed at Galeras and other volcanoes. *Journal of Volcanology and Geothermal Research* 77: 173–193.
Hammer JE, Cashman KV, and Voight B (2000) Magmatic processes revealed by textural and compositional trends in Merapi dome lavas. *Journal of Volcanology and Geothermal Research* 100: 165–192.
Harlow DH, Power JA, Laguerta EP, Ambubuyog G, White RA, and Hoblitt RP (1996) Precursory seismicity and forecasting of the June 15, 1991 eruption of Mount Pinatubo, Philippines. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, pp. 285–305. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.
Hausback BP (2000) Geologic map of the Sasquatch Steps area, north flank of Mount St. Helens, Washington. U S Geol Surv, Geologic Investigations Series, Report: I–2463, 1 sheet http://geopubs.wr.usgs.gov.offcampus.lib.washington.edu/i-map/i2463/ (acessed on Feb 2007).
Hayashi Y and Morita Y (2003) An image of a magma intrusion process inferred from precise hypocentral migrations of the earthquake swarm east of the Izu Peninsula. *Geophysical Journal International* 153: 159–174.
Hayes SK, Montgomery DR, and Newhall CG (2002) Fluvial sediment transport and deposition following the 1991 eruption of Mount Pinatubo, Philippines. *Geomorphology* 45: 211–224.
Hellweg M (2003) The polarization of volcanic seismic signals: Medium or source? *Journal of Volcanology and Geothermal Research* 128: 159–176.
Hernandez PA, Notsu K, Salazar JM, *et al.* (2001) Carbon dioxide degassing by advective flow from Usu Volcano, Japan. *Science* 292: 83–86.
Hess K-U and Dingwell DB (1996) Viscosities of hydrous leucogranitic melts: A non-Arrhenian model. *American Mineralogist* 81: 1297–1300.
Hill DP, Langbein JO, and Prejean S (2003) Relations between seismicity and deformation during unrest in Long Valley Caldera, California, from 1995 through 1999. *Journal of Volcanology and Geothermal Research* 127: 175–193.
Hill DP, Pollitz F, and Newhall C (2002) Earthquake–volcano interactions. *Physics Today* 55(11): 41–47.
Hill DP, Prejean S, this volume, Dynamic stresses and Remote Triggering
Hincks TH, Aspinall W, and Woo G (2005) Decision analysis tools for volcano observatories (abs.) Eos, Transactions AGU, Fall Meeting Supplement, Abstract 86(52): V32A–05.
Hoblitt RP, Mori J, and Power JA (1996) Computer visualization of earthquake hypocenters. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, pp. 383–385. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle + disk with Volquake program.
Hoblitt RP, Wolfe EW, Scott WE, Couchman MR, Pallister JS, and Javier D (1996) The preclimactic eruptions of Mount Pinatubo, June 1991. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 457–511. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.
Iguchi M (1989) Distribution of the initial motions of volcanic microearthquakes (B-type) at Sakurajima Volcano. *Bulletin of the Disaster Prevention Research Institute*, *Kyoto University* 32(B-1): 13–22.

Ikuta R and Yamaoka K (2004) Temporal variation in the shear wave anisotropy detected using the Accurately Controlled Routinely Operated Signal System (ACROSS). *Journal of Geophysical Research* 109(B9): 14.

Ikuta R, Yamaoka K, Miyakawa K, Kunitomo T, and Kumazawa M (2002) Continuous monitoring of propagation velocity of seismic waves using ACROSS. *Geophysical Research Letters* 29(13): 4.

Instituto Geofisico, Escuela Politecnica Nacional (2006) Volcán Tungurahua. http://www.igepn.edu.ec/vulcanologia/tungurahua/actividad/actividad.htm (accessed Feb 2007).

Ishihara K and Iguchi M (1989) The relationship between microearthquake swarms and volcanic activity at Sakurajima Volcano. *Bulletin of the Disaster Prevention Research Institute, Kyoto University* 32(B-1): 1–12.

Jackson DB (1988) Geoelectric observations (including the September 1982 summit eruption). In: Wolfe EW (ed.) *US Geological Survey Professional Paper 1463: The Puu Oo Eruption of Kilauea Volcano, Hawaii: Episodes 1 through 20, January 3, 1983, Through June 8, 1984*, pp. 237–251. Washington, DC: US Geological Survey.

Jackson DB, Kauahikaua J, and Zablocki CJ (1985) Resistivity monitoring of an active volcano using the controlled-source electromagnetic technique, Kilauea, Hawaii. *Journal of Geophysical Research* 90: 12545–12555.

Jaupart C (1998) Gas loss from magmas through conduit walls. In: Gilbert JS and Sparks RSJ (eds.) *Geological Society, London, Special Publication 145: The Physics of Explosive Volcanic Eruptions*, pp. 73–90. London: Geological Society of London.

Jaupart C and Allègre CJ (1991) Gas content, eruption rate and instabilities of eruption regime in silicic volcanoes. *Earth and Planetary Science Letters* 102: 413–429.

Jellinek AM, Manga M, and Saar MO (2004) Did melting glaciers cause volcanic eruptions in eastern California? Probing the mechanics of dike formation. *Journal of Geophysical Research* 109: B09206 (doi:10.1029/2004JB002978).

Johnson DJ (1987) Elastic and inelastic magma storage at Kilauea Volcano. In: Decker RW, Wright TL, and Stauffer PH (eds.) *US Geological Survey Professional Paper 1350: Volcanism in Hawaii*, pp. 1297–1306. Washington, DC: US Government Printing Office.

Johnston MJS (1989) Review of magnetic and electric field effects near active faults and volcanoes in the USA. *Physics of the Earth and Planetary Interiors* 57: 47–63.

Julian BR (1994) Volcanic tremor: nonlinear excitation by fluid flow. *Journal of Geophysical Research* 99: 11 859–11 877.

Julian BR (2000) Periodic doubling and other nonlinear phenomena in volcanic earthquakes and tremor. *Journal of Volcanology and Geothermal Research* 101: 19–26.

Kaizuka S, Kato S, Nagaoka S, and Miyauchi T (1985) Geomorphology of Iwo-jima and surrounding sea floor. *Journal of Geography* (Tokyo) 94: 424–436.

Kamo K, Ishihara K, and Tahira M (1994) Infrasonic and seismic detection of explosive eruptions at Sakurajima Volcano, Japan, and the PEGASAS-VE early-warning system. In: Casadevall TJ (ed.) *US Geological Survey Bulletin 2047: Volcanic Ash and Aviation Safety*, pp. 357–365. Denever, CO: US Geological Survey.

Katsui Y and Katz HR (1967) Lateral fissure eruptions in the southern Andes of Chile. *Journal of Faculty of Science Hokkaido University* 13(4): 433–448.

Kawakatsu H and Yamamoto M, this volume, Volcano Seismology.

Kazahaya K, Shinohara H, and Saito G (1994) Excessive degassing of Izu–Oshima volcano; magma convection in a conduit. *Bulletin of Volcanology* 56(3): 207–216.

Kazahaya K, Shinohara H, and Saito G (2002) Degassing process of Satsuma–Iwojima volcano, Japan: Supply of volatile components from a deep magma chamber. *Earth Planets Space* 54: 327–335.

Kazahaya K, Shinohara H, Uto K, *et al.* (2004) Gigantic SO_2 emission from Miyakejima Volcano, Japan, caused by caldera collapse. *Geology* 32: 425–428.

Kedar S, Kanamori H, and Sturtevant B (1998) Bubble collapse as the source of tremor at Old Faithful Geyser. *Journal of Geophysical Research* 103(B10): 24283–24299.

Kedar S, Sturtevant B, and Kanamori H (1996) The origin of harmonic tremor at Old Faithful Geyser. *Nature* 379: 708–711.

Keppler H (1999) Experimental evidence for the source of excess sulfur in explosive volcanic eruptions. *Science* 284: 1652–1654.

Kieffer S (1984) Seismicity of Old Faithful Geyser: An isolated source of geothermal noise and possible analogue of volcanic tremor. *Journal of Volcanology and Geothermal Research* 22: 59–96.

Klein FW, Koyanagi RY, Nakata JS, and Tanigawa WR (1987) The seismicity of Kilauea's magma system. In: Decker RW, Wright TL, and Stauffer PH (eds.) *US Geological Survey Professional Paper 1350: Volcanism in Hawaii*, pp. 1019–1185. Washington, DC: US Government Printing Office.

Klug C and Cashman KV (1996) Permeability development in vesiculating magmas: Implications for fragmentation. *Bulletin of Volcanology* 58: 87–100.

Konstantinou KI and Schlindwein V (2002) Nature, wavefield properties and source mechanism of volcanic tremor: A review. *Journal of Volcanology and Geothermal Research* 119: 161–187.

Koyama M and Yoshida Y (1994) Relationship between eruptive history and crustal stress field based on cumulative changes in magma discharge from volcanoes. *Bulletin of the Volcanological Society of Japan* 39: 177–190 (in Japanese with English abstract).

Koyanagi RY, Chouet B, and Aki K (1987) Origin of volcanic tremor in Hawaii. Part 1: Data from the Hawaiian Volcano Observatory 1969–1985. In: Decker RW, Wright TL, and Stauffer PH (eds.) *US Geological Survey Professional Paper 1350: Volcanism in Hawaii*, pp. 1221–1258. Washington, DC: US Government Printing Office.

Kress V (1997) Magma mixing as a source for Pinatubo sulphur. *Nature* 389(6651): 591–593.

Krotkov N, Datta S, Krueger AJ, *et al.* (2003) Measuring sulfur dioxide from space: The promise of Ozone Monitoring Instrument (OMI) on EOS-AURA platform (abs.) Eos Transactions AGU, Fall Meeting Supplement 84: 985.

Kumagai H, Miyakawa K, Negishi H, Inoue H, Obara K, and Suetsugu D (2003) Magmatic dike resonances inferred from very-long-period seismic signals. *Science* 299: 2058–2061.

LaHusen RG (1996) Detecting Debris Flows Using Ground Vibrations: US Geol Surv Fact Sheet 236–96, http://vulcan.wr.usgs.gov/Projects/AFM/Publications/FS236-96/ (accessed Feb 2007).

LaHusen RG (2005) Acoustic Flow Monitor System – User Manual. US Geological Survey Open-File Report 02–429, 22 pp.

Lara LE, Naranjo JA, and Moreno H (2004) Rhyodacitic fissure eruption in Southern Andes (Cordon Caulle; 40.5°S) after the 1960 (Mw 9.5) Chilean earthquake: A structural interpretation. *Journal of Volcanology and Geothermal Research* 138: 127–138.

Leet RC (1988) Saturated and subcooled hydrothermal boiling in groundwater flow channels as a source of

harmonic tremor. *Journal of Geophysical Research* 93: 4835–4849.
Legrand D, Calahorrano A, Guillier B, *et al.* (2002) Stress tensor analysis of the 1998–1999 tectonic swarm of northern Quito related to the volcanic swarm of Guagua Pichincha volcano, Ecuador. *Tectonophysics* 344: 15–36.
Lenat J-F (1995) Geoelectrical methods in volcano monitoring. In: McGuire B, Kilburn CRJ, and Murray J (eds.) *Monitoring Active Volcanoes*, pp. 248–274. London: UCL Press.
Linde AT, Sacks IS, Johnston MJS, Hill DP, and Bilham RG (1994) Increased pressure from rising bubbles as a mechanism for remotely triggered seismicity. *Nature* 371: 408–410.
Lipman PW, Moore JG, and Swanson DA (1981) Bulging of the north flank before the May 18 eruption: Geodetic data. In: Lipman PW and Mullineaux DR (eds.) *US Geological Survey Professional Paper 1250: The 1980 eruptions of Mount St. Helens, Washington*, pp. 143–156. Denver, CO: US Geological Survey.
Lu Z, Masterlark T, Dzurisin D, Rykhus R, and Wicks C, Jr. (2003) Magma supply dynamics at Westdahl Volcano, Alaska, modeled from satellite radar interferometry. *Journal of Geophysical Research* 108: 2354 (doi:10.1029/2002JB002311).
Luhr JF, Carmichael IS, and Varekamp JC (1984) The 1982 eruption of El Chichón volcano, Chiapas, Mexico: Mineralogy and petrology of the anhydrite-bearing pumices. *Journal of Volcanology and Geothermal Research* 23: 69–108.
Maclennan J, Jull M, McKenzie D, Slater L, and Gronvold K (2002) The link between volcanism and deglaciation in Iceland. *Geochemistry Geophysics Geosystems* 3: 25.
Major JJ (2004) Posteruption suspended sediment transport at Mount St. Helens: Decadal-scale relationships with landscape adjustments and river discharges. *Journal of Geophysical Research* 109: (doi:10.1029/2002JF000010).
Malone SD (1983) Volcanic earthquakes: Examples from Mount St. Helens. In: Kanamori H and Boschi E (eds.) *Earthquakes: Observations, Theory and Interpretation*, pp. 436–455. Amsterdam: Elsevier.
Malone SD and Qamar A (1984) Repetitive microearthquakes as the source for volcanic tremor at Mount St. Helens. *Eos Transactions American Geophysical Union* 65: 1001.
Manga M and Brodsky E (2006) Seismic triggering of eruptions in the far field: Volcanoes and geysers. *Annual Reviews of Earth and Planetary Sciences* 34: 263–291.
Manga M and Loewenberg M (2001) Viscosity of magmas containing highly deformable bubbles. *Journal of Volcanology and Geothermal Research* 105: 19–24.
Mangan M and Sisson TW (2000) Delayed, disequilibrium degassing in rhyolite magma; decompression experiments and implications for explosive volcanism. *Earth and Planetary Science Letters* 183: 441–455.
Marcial S, Melosantos AA, Hadley KC, LaHusen RG, and Marso JN (1996) Instrumental lahar monitoring at Mount Pinatubo. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 1015–1022. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.
Marsh BD (1981) On the crystallinity, probability of occurrence, and rheology of lava and magma. *Contributions to Mineralogy and Petrology* 78: 85–98.
Marsh BD (2000) Magma chambers. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 191–206. San Diego, CA: Academic Press.
Marzocchi W, Sandri L, Gasparini P, Newhall C, and Boschi E (2004) Quantifying probabilities of volcanic events: The example of volcanic hazard at Mt. Vesuvius. *Journal of Geophysical Research* 109: B11201 (doi:10.1029/2004JB003155).
Marzocchi W, Selva J, and Sandri L (2005) BET-VH: A probabilistic tool for long- and short-range volcanic hazard assessment (abs). Eos, Transactions AGU, Fall Meeting Supplement, Abstract 86(52): V32A–04.
Mastin LG (1994) Explosive tephra emissions at Mount St. Helens, 1989–1991: The violent escape of magmatic gas following storms? *Geological Society of America Bulletin* 106: 175–185.
Matsumoto N, Sato T, Matsushima N, Akita F, Shibata T, and Suzuki A (2002) Hydrological anomalies associated with crustal deformation before the 2000 eruption of Usu Volcano, Japan. *Geophysical Research Letters* 29(5): 1057 (doi:10.1029/2001GLO13968).
McGee KA, Gerlach TM, Kessler R, and Doukas MP (2000) Geochemical evidence for a magmatic CO_2 degassing event at Mammoth Mountain, California, September–December 1997. *Journal of Geophysical Research* 105: 8447–8456.
McGuire B, Kilburn CRJ, and Murray J (eds.) (1995) *Monitoring Active Volcanoes*, 421 p. London: UCL Press.
McKee CO, Wallace DA, Almond RA, and Talai B (1981) Fatal hydro-eruption of Karkar volcano in 1979: Development of a maar-like crater. In: Johnson RW (ed.) *Geological Survey of Papua New Guinea Memoir 10: Cooke-Ravian Volume of Volcanological Papers*, pp. 63–84. Port Moresby: Geological Survey PNG.
McNutt SR (1992) Volcanic tremor. In: Nierenberg WA (ed.) *Encyclopedia of Earth System Science*, vol. 4, pp. 417–425. San Diego, CA: Academic Press.
McNutt SR (1996) Seismic monitoring and eruption forecasting of volcanoes. In: Scarpa R and Tilling RI (eds.) *Monitoring and Mitigation of Volcano Hazards*, pp. 99–146. Berlin: Springer.
McNutt SR (2000a) Volcanic Seismicity. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 1015–1033. San Diego, CA: Academic Press.
McNutt SR (2000b) Seismic monitoring. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 1095–1119. San Diego, CA: Academic Press.
McNutt SR (2005) Volcanic seismology. *Annual Reviews of Earth and Planetary Sciences* 33: 461–491.
McNutt SR and Beavan RJ (1981) Volcanic earthquakes at Pavlof Volcano correlated with the solid Earth tide. *Nature* 294(5842): 615–618.
McNutt SR and Beavan RJ (1987) Eruptions of Pavlof Volcano and their possible modulation by ocean load and tectonic stresses. *Journal of Geophysical Research* 92: 11509–11523.
Minakami T (1960) Fundamental research for predicting volcanic eruptions. Part I: Earthquakes and crustal deformations originating from volcanic activities. *Bulletin of Earthquake Research Institute* 38: 497–544.
Minakami T (1974) Seismology of volcanoes in Japan. In: Civetta L, Gasparini P, Luongo G, and Rapolla A (eds.) *Physical Volcanology*, pp. 1–27. Amsterdam: Elsevier.
Mizutani H, Ishido T, Yokokura T, and Ohnishi S (1976) Electrokinetic phenomena associated with earthquakes. *Geophysical Research Letters* 3: 365–368.
Mogi K (1958) Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them. *Bulletin of Earthquake Research Institute* 36: 99–134.
Moore PL, Iverson NR, and Iverson RM (2005) Frictional properties of gouge generated during the 2004–2005 lava dome extrusion at Mount St. Helens and implications for seismicity (abs) Eos Transactions AGU, Fall Meeting Supplement, Abstract V53D–1592.

Moore JG, Normark WR, and Holcomb RT (1994) Giant Hawaiian landslides. *Annual Reviews of Earth and Planetary Sciences* 22: 119–144.
Nakada S, Shimizu H, and Ohta K (1999) Overview of the 1990–1995 eruption at Unzen Volcano. *Journal of Volcanology and Geothermal Research* 89: 1–22.
Nakamichi H, Ukawa M, and Sakai S (2004) Precise hypocenter locations of midcrustal low-frequency earthquakes beneath Mt. Fuji, Japan. *Earth Planets Space* 56: e37–e40.
Neuberg J (2000) Characteristics and causes of shallow seismicity in andesitic volcanoes. *Philosophical Transactions of the Royal Society of London* A 358: 1533–1546.
Newhall CG (2004) Degassing of arc volcanoes: Generalizations from WOVOdat (abstract) IAVCEI General Assembly, Pucon Chile, November 2004.
Newhall CG, Albano SE, Matsumoto N, and Sandoval T (2001) Roles of groundwater in volcanic unrest. *Journal of the Geological Society of the Phillippines* 56: 69–84.
Newhall CG, Fujita E, Hill D, Kobayashi T, Lowenstern J, and Ukawa M (2003) Comparative seismicity, deformation, and magma supply at Iwo-jima, Long Valley, and Yellowstone calderas (abstract). Abstracts, IUGG, Sapporo, July 2003.
Newhall CG and Hoblitt RP (2002) Constructing event trees for volcanic crises. *Bulletin of Volcanology* 64: 3–20.
Newhall CG, Power JA, and Punongbayan RS (2002) Pinatubo: 'to make grow:' *Science* 295: 1241–1242.
Newhall CG, Punongbayan RS, and Gerlach TM (1994) Tight and leaky volcanoes: Implications for forecasting explosive eruptions. *Atti dei Convegni Lincei* 112: 13–21.
Newhall CG and Self S (1982) The Volcanic Explosivity Index (VEI): An estimate of explosive magnitude for historical volcanism. *Journal of Geophysical Research* 87: 1231–1238.
Nishimura T (2006) Magma ascent with and without degassing and ground deformation (abstract). Transactions, Fall Meeting Supplement, Abstract AGU 87(52): V32C–01.
NOAA Satellite and Information Service (2005) Volcanic Ash Advisory Centers. URL: http://www.ssd.noaa.gov/VAAC/vaac.html (accessed on Feb 2007).
Noguchi K and Kamiya H (1963) Prediction of volcanic eruption by measuring the chemical composition and amounts of gases. *Bulletin of Volcanology* 26: 367–378.
Okada Y (1992) Internal Deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America* 82: 1018–1049.
Oyagi N and Inokuchi T (1985) Geology of Iwo–jima. *Journal of Geography* (Tokyo) 94: 436–445.
Pal R (2003) Rheological behavior of bubble-bearing magmas. *Earth and Planetary Science Letters* 207: 165–179.
Pallister JS, Hoblitt RP, Meeker GP, Knight RJ, and Siems DF (1996) Magma mixing at Mount Pinatubo: Petrographic and chemical evidence from the 1991 deposits. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 687–731. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.
Pallister JS, Reagan M, and Cashman K (2005) A new eruptive cycle at Mount St. Helens? *Eos Transactions of the American Geophysical Union* 86: 499.
Power JA, Stihler SD, White RA, and Moran SC (2004) Observations of deep long-period (DLP) seismic events beneath Aleutian arc volcanoes; 1989–2002. *Journal of Volcanology and Geothermal Research* 138: 243–266.
Power JA, White RA, Murray TL *et al.* (2001) Seismicity and forecasting of the 1991 eruption of Mount Pinatubo: A ten-year retrospective (abs). Eos Transactions AGU, Fall Meeting, Abstract U31A–03.
Pyle DM (2000) Sizes of volcanic eruptions. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 263–269. San Diego, CA: Academic Press.
Ramos EG, Hamburger MW, Pavlis GL, and Laguerta EP (1999) The low-frequency earthquake swarms at Mount Pinatubo, Philippines; implications for magma dynamics. *Journal of Volcanology and Geothermal Research* 92: 295–320.
Reid ME (2004) Massive collapse of volcano edifices triggered by hydrothermal pressurization. *Geology* 32: 373–376.
Roeloffs EA and Linde AT (2007) Borehole observations of continuous strain and fluid pressure. In: Dzurisin D (ed.) *Volcano Deformation: Geodetic Monitoring Techniques*, ch. 9, pp. 305–322. Berlin: Springer-Praxis.
Romero R (2004) The International Airways Volcano Watch (IAVW). Proceedings Second International Conference on Volcanic Ash and Aviation Safety, Alexandria, VA June 2004, session 4, 10 p.
Rowe CA, Thurber CH, and Power JA (2001) Precise correlation-based phase repacking of Pinatubo pre-eruption seismicity illuminates families of repeating earthquakes and a refined image of precursory hypocenter locations (abs). Eos Transactions AGU Fall Meeting Supplement, Abstract 82(47): U32A–0002.
Rust AC, Manga M, and Cashman KV (2003) Determining flow type, shear rate and shear stress in magmas from bubble shapes and orientations. *Journal of Volcanology and Geothermal Research* 122: 111–132.
Rutherford MJ and Devine JD (1996) Preeruption pressure and temperature conditions and volatiles in the 1991 dacitic magma of Mount Pinatubo. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 751–766. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.
Rutherford MJ and Gardner JE (2000) Rates of magma ascent. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 207–217. San Diego, CA: Academic Press.
Ryan MP (1987) Neutral buoyancy and the mechanical evolution of magmatic systems. In: Mysen BO (ed.) *Special Publication, Geochemical Society, vol. 1: Magmatic Processes: Physicochemical Principles*, pp. 259–287. St Louis, MO: The Geochemical Society.
Rymer H (1996) Microgravity monitoring. In: Scarpa R and Tilling RI (eds.) 1996 *Monitoring and Mitigation of Volcano Hazards*, pp. 169–197. Berlin: Springer.
Rymer H and Williams-Jones G (2000) Volcanic eruption prediction; magma chamber physics from gravity and deformation measurements. *Geophysical Research Letters* 27(16): 2389–2392.
Sanchez JJ, McNutt SR, Power JA, and Wyss M (2004) Spatial variations in the frequency-magnitude distribution of earthquakes at Mount Pinatubo volcano. *Bulletin of the Seismological Society of America* 94: 430–438.
Sasai Y, Shimomura T, Hamano Y, *et al.* (1990) Volcano-magnetic effect observed during the 1986 eruption of Izu-Oshima volcano. *Journal of Geomagnetism and Geolectricity* 42: 291–317.
Sato T, Matsumoto N, Takahashi M, *et al.* (2001) Observation of water level, temperature and electric conductivity of groundwater around Usu Volcano. In: Tomiya A (ed.) *Bulletin of the Geological Society of Japan 52: The 2000 eruption of Usu Volcano*, pp. 143–148. Tokyo: Geological Survey of Japan.
Scarpa R and Tilling RI (eds.) (1996) *Monitoring and Mitigation of Volcano Hazards*, 841 p. Berlin: Springer.
Schmincke H-U (2004) *Volcanism*, 324 p. Berlin: Springer.
Shaw HR (1972) Viscosities of magmatic silicate liquids; an empirical method of prediction. *American Journal of Science* 272(9): 870–893.

Shaw HR (1985) Links between magma-tectonic rate balances, plutonism, and volcanism. *Journal of Geophysical Research* 90(B13): 11275–11288.

Shibata T and Akita F (2001) Precursory changes in well water prior to the March 2000 eruption of Usu volcano, Japan. *Geophysical Research Letters* 28(9): 1799–1802.

Shimozuru D (1972) A seismological approach to the prediction of volcanic eruptions. In: UNESCO (ed). *The Surveillance and Prediction of Volcanic Activity*, pp. 19–45. Paris: UNESCO.

Siebert L (1996) Hazards of large volcanic debris avalanches and associated eruptive phenomena. In: Scarpa R and Tilling RI (eds.) *Monitoring and Mitigation of Volcano Hazards*, pp. 541–572. Berlin: Springer.

Siebert L, Glicken H, and Ui T (1987) Volcanic hazards from Bezymianny- and Bandai-type eruptions. *Bulletin of Volcanology* 49: 435–459.

Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) (2000) *Encyclopedia of Volcanoes*, 1417 p. San Diego, CA: Academic Press.

Simkin T (1993) Terrestrial volcanism in space and time. *Annual Reviews of Earth and Planetary Sciences* 21: 427–452.

Simkin T and Siebert L (1994) *Volcanoes of the World*, 349 p. Tucson, AZ: Geoscience Press.

Simkin T and Siebert L (2000) Earth's volcanoes and eruptions: An overview. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 249–261. San Diego, CA: Academic Press.

Sparks RSJ (1978) The dynamics of bubble formation and growth in magmas: A review and analysis. *Journal of Volcanology and Geothermal Research* 3: 1–37.

Sparks RSJ (2003) Dynamics of magma degassing. In: Oppenheimer C, Pyle DM, and Barclay J (eds.) *Geological Society of London, Special Publication 213: Volcanic Degassing*, pp. 5–22. London: Geological Society of London.

Sparks RSJ and Marshall LA (1986) Thermal and mechanical constraints on mixing between mafic and silicic magmas. *Journal of Volcanology and Geothermal Research* 29: 99–124.

Sparks RSJ, Sigurdsson H, and Wilson L (1977) Magma mixing: a mechanism for triggering acid explosive eruptions. *Nature* 267: 315–318.

Sparks RSJ and Young SR (2002) The eruption of Soufrière Hills Volcano, Montserrat (1995–1999); overview of scientific results. In: Druitt TH and Kokelaar BP (eds.) *Geological Society of London, Memoir 21: The Eruption of Soufriere Hills Volcano, Montserrat from 1995 to 1999*, pp. 45–69. Bath, UK: Geological Society Publishing House.

Spera FJ (1984) Carbon dioxide in petrogenesis. Part III: Role of volatiles in the ascent of alkaline magma with special reference to xenolith-bearing magmas. *Contributions to Mineralogy and Petrology* 88: 217–232.

Spera FJ (2000) Physical properties of magmas. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 171–190. San Diego, CA: Academic Press.

Spieler O, Dingwell DB, and Allidibirov M (2004) Magma fragmentation speed: An experimental determination. *Journal of Volcanology and Geothermal Research* 129: 109–123.

Stasiuk VM, Barclay J, Carroll MR, *et al.* (1996) Degassing during magma ascent in the Mule Creek vent (USA). *Bulletin of Volcanology* 58: 117–130.

Stevenson DS and Blake S (1998) Modelling the dynamics and thermodynamics of volcanic degassing. *Bulletin of Volcanology* 60: 307–317.

Stoiber RE and Jepsen A (1973) Sulfur dioxide contributions to the atmosphere by volcanoes. *Science* 182: 577–579.

Stoiber RE, Malinconico LL, and Williams SN (1983) Use of the correlation spectrometer at volcanoes. In: Tazieff H and Sabroux J-C (eds.) *Forecasting Volcanic Events*, pp. 425–444. New York: Elsevier.

Swanson DA, Casadevall TJ, Dzurisin D, Malone SD, Newhall CG, and Weaver CS (1983) Predicting eruptions at Mount St Helens June 1980 through December 1982. *Science* 221: 1369–1376.

Symonds RB, Gerlach TM, and Reed MH (2001) Magmatic gas scrubbing: Implications for volcano monitoring. *Journal of Volcanology and Geothermal Research* 108: 303–341.

Symonds RB, Rose WI, Bluth GJS, and Gerlach TM (1994) Volcanic-gas studies: Methods, results, and applications. In: Carroll MR and Holloway JR (eds.) *Reviews in Mineralogy, vol. 30: Volatites in Magmas*, pp. 1–66. Washington, DC: Mineralogical Society of America.

Tilling RI (1989) *Volcanic Hazards, Short Course in Geology*, vol. 1., 123 pp. Washington, DC: American Geophysical Union.

Tilling RI (1991) Reducing volcanic risk: Are we winning some battles but losing the war? *Earthquakes and Volcanoes* 22(3): 133–137.

Tsuruga K, Yomogida K, Honda S, Ito H, Ohminato T, and Kawakatsu H (1997) Spatial and temporal variations of volcanic earthquakes at Sakurajima Volcano, Japan. *Journal of Volcanology and Geothermal Research* 75: 337–358.

Tsuya H (1955) Geological and petrological studies of volcano Fuji. Part 5: On the 1707 eruption of volcano Fuji. *Bulletin of Earthquake Research Institute, University of Tokyo* 33: 341–383.

Tuffen H and Dingwell DB (2004) Fault textures in volcanic conduits: Evidence for seismic trigger mechanisms during silicic eruptions. *Bulletin of Volcanology* 67: 370–387 (doi:10.1007/s00445-004-0383-5).

Tungol NM (2002) *Engineering Geology of the Pasig-Potrero Lahar Fan*, *Eastern Mount Pinatubo*, *Luzon Island*, *Philippines*. PhD Thesis, University of Canterbury, Christchurch.

Tungol NM and Regalado MTM (1996) Rainfall, acoustic flow monitor records, and observed lahars of the Sacobia River in 1992. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 1023–1032. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.

Ui T (1983) Volcanic dry avalanche deposits – identification and comparison with nonvolcanic debris stream deposits. *Journal of Volcanology and Geothermal Research* 18: 135–150.

Ukawa M (2005) Deep low-frequency earthquake swarm in the mid crust beneath Mount Fuji (Japan) in 2000 and 2001. *Bulletin of Volcanology* 68: 17 (doi:10.1007/s00445-005-0419-5).

Ukawa M, Fujita E, Ueda H, Kumagai T, Nakajima H, and Morita H (2005) Long-term geodetic measurements of large scale deformation at low-jima caldera, Japan. *Journal of Volcanology and Geothermal Research* 150: 21 (doi:10.1016/j.jvolgeores.2005.07.008).

UNESCO (1972) *The Surveillance and Prediction of Volcanic Activity*. Paris: UNESCO.

Van Vilet LB, Bourgeois O, Dauteuil O, Embry JC, Schneider JL, and Guillou H (2005) Deglaciation and volcano-seismic activity in northern Iceland: Holocene and early Eemian. *Geodinamica Acta* 18: 81–100.

Varekamp JC and Rowe GL (eds.) (2000) Crater Lakes. *Journal of Volcanology and Geothermal Research* 97(1–4): 508.

Vergniolle S and Jaupart C (1986) Separated two-phase flow and basaltic eruptions. *Journal of Geophysical Research* 91(12): 12842–12860.

Voight B and Elsworth D (1997) Failure of volcanic slopes. *Geotechnique* 47: 1–31.

Voight B, Sparks RSJ, Miller AD, *et al.* (1999) Magma flow instability and cyclic activity at Soufrière Hills Volcano, Montserrat, British West Indies. *Science* 283: 1138–1142.

Voight B, Young KD, Hidayat D, *et al.* (2000) Deformation and seismic precursors to dome-collapse and fountain-collapse nuées ardentes at Merapi Volcano, Java, Indonesia, 1994–1998. *Journal of Volcanology and Geothermal Research* 100: 261–287.

Walker GPL (1973) Explosive volcanic eruptions: A new classification scheme. *Geologische Rundschau* 62: 431–446.

Walker GPL (1980) The Taupo Pumice: Product of the most powerful known (ultraplinian) eruption? *Journal of Volcanology and Geothermal Research* 8: 69–94.

Wallace PJ (2001) Volcanic SO_2 emissions and the abundance and distribution of exsolved gas in magma bodies. *Journal of Volcanology and Geothermal Research* 108: 85–106.

Wallace P and Anderson AT, Jr. (2000) Volatiles in magmas. In: Sigurdsson H, Houghton BF, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes*, pp. 149–170. San Diego, CA: Academic Press.

White RA (1996) Precursory deep long-period earthquakes at Mount Pinatubo: Spatio-temporal link to a basalt trigger. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines,* pp. 751–766. Philippine Institute of Volcanology and Seismology, Quezon City, and University of Washington Press, Seattle.

White RA and Power JA (2001) Distal volcano-tectonic earthquakes (DVT's): Diagnosis and use in eruption forecasting (abs.). Eos Transactions AGU Fall Meeting Supplement, Abstract 82(47): U32A–0001.

Wicks CW, Jr., Dzurisin D, Ingebritsen S, Thatcher W, Lu Z, and Iverson J (2002) Magmatic activity beneath the quiescent Three Sisters volcanic center, central Oregon Cascade Range, USA. *Geophysical Research Letters* 29: 7 (doi:10.1029/2001GL014205).

Wilson L (1976) Explosive volcanic eruptions. Part III: Plinian eruption columns. *Geophysical Journal of the Royal Astronomical Society* 45: 543–556.

Wilson L, Sparks RSJ, and Walker GPL (1980) Explosive volcanic eruptions. Part IV: The control of magma properties and conduit geometry on eruption column behavior. *Geophysical Journal of the Royal Astronomical Society* 63: 117–148.

Witter JB (2003) *Convection of Magma in Volcanic Conduits as a Degassing Mechanism at Active Volcanoes*. PhD Thesis, University of Washington.

Witter JB, Kress VC, Delmelle P, and Stix J (2004) Volatile degassing, petrology, and magma dynamics of the Villarrica Lava Lake, Southern Chile. *Journal of Volcanology and Geothermal Research* 134: 303–337.

Witter JB, Kress VC, and Newhall CG (2005) Volcán Popocatépetl, Mexico. Petrology, magma mixing, and immediate sources of volatiles for the 1994-present eruption. *Journal of Petrology* 46: 2337–2366.

Wright JV, Smith AL, and Self S (1980) A working terminology of pyroclastic deposits. *Journal of Volcanology and Geothermal Research* 8: 315–336.

Zapata JA, Calvache ML, Cortes GP, *et al.* (1997) SO_2 fluxes from Galeras Volcano, Colombia, 1989–1995; progressive degassing and conduit obstruction of a Decade Volcano. *Journal of Volcanology and Geothermal Research* 77: 195–208.

Zhang Y and Behrens H (2000) H_2O diffusion in rhyolitic melts and glasses. *Chemical Geology* 169: 243–262.

Zhang Y, Xu Z, and Liu Y (2003) Viscosity of hydrous rhyolitic melts inferred from kinetic experiments, and a new viscosity model. *American Mineralogist* 88: 1741–1752.

Zlotnicki J (1995) Geomagnetic surveying methods. In: McGuire B, Kilburn CRJ, and Murray J (eds.) *Monitoring Active Volcanoes*, pp. 275–300. London: UCL Press.

Zlotnicki J, Sasai Y, Yvetot P, *et al.* (2003) Resistivity and self-potential changes associated with volcanic activity: The July 8, 2000 Miyake–jima eruption (Japan). *Earth and Planetary Science Letters* 205: 139–154.

4.13 Volcano Seismology

H. Kawakatsu, University of Tokyo, Tokyo, Japan

M. Yamamoto, Tohoku University, Sendai, Japan

4.13.1 Introduction

Volcano seismology is a field of volcanology in which seismological techniques are employed to help understanding the physical conditions and dynamic states of volcanic edifices and volcanic fluid systems to such a level that it eventually contributes to predictions of initiation and cessation of hazardous volcanic activities. Seismology has been a powerful tool for this purpose, especially in the past 10–15 years as new observational and analysis techniques in seismology have become available (e.g., Chouet, 1996a, 1996b; McNutt, 2005); they include digital and broadband seismometry, and array and moment tensor waveform analyses. While the subjects that should be covered under the title of volcano seismology are quite broad and diverse, this chapter, however, neither intends to nor is capable to cover all of its areas; instead we try to focus on some of seismological phenomena which appear to be specific to volcanoes. In a sense, volcano seismology here is a field of seismology which deals with problems specific to volcanic areas. Some of the subjects, such as seismic structure of volcanoes (tomography) and eruption monitoring (seismicity), are therefore intentionally omitted. The readers who are interested in these subjects are advised to refer to the excellent reviews by Chouet (1996a, 1996b, 2003) and McNutt (1996, 2002, 2005).

4.13.2 Volcanic Seismic Signals

For seismologists who have only observed seismograms of ordinary faulting earthquakes, the most intriguing aspect of volcano seismology may be the ample variety of waveforms recorded in volcanic environments. Such volcanic seismic signals range from totally chaotic continuous vibrations lasting for minutes, hours, or sometimes days to purely monotonic vibrations decaying with constant damping rates; they often show several spectral peaks whose frequency may vary during the course of volcanic activities, possibly reflecting the change of physical conditions of volcanic fluid systems; when eruptions occur, variety of long-period signals up to as long as a few hundred seconds are generated and sometimes observed at seismic stations around the world. In the following, some of such intriguing volcanic seismic signals will be presented.

Terminology. Based on their appearance on short-period seismometers, volcanic seismic signals have been traditionally classified into four types: high-frequency or A type, low-frequency or B type, explosion quakes, and volcanic tremors (e.g., Minakami, 1974; for the detail of different terminology, see table 1 of McNutt (1996)). In this chapter, following this tradition, we use high frequency for frequencies higher than 5 Hz, and low frequency for frequencies between 5 and 0.5 Hz. On the other hand, the usage of period remains a complicated issue in volcano seismology; the most widely accepted notation to describe volcanic seismic signals is that of Chouet (1996b), in which events of about 1 s are called 'long-period' (LP) events and those with longer periods are called 'very-long-period' (VLP) and 'ultra-long-period' (ULP) events. This definition, however, contradicts with the conventional usage of period in traditional earthquake seismology, where periods longer than the dominant period of the ambient microseismic noise (around 5–10 s) are called 'long-period' (Aki and Lee, 2003; McNutt, 1996, 2005). In this chapter, as we try to cover the subject in a wider context of seismology, we follow the convention of earthquake seismology, and use the term long-period for periods longer than 5 s and short period for periods shorter than or around 1 s; for events with a period longer than 50 s, we may also use very long-period. To avoid confusion, whenever appropriate, we will try to associate actual number in seconds to describe a period, for example long-period (10 s).

4.13.2.1 Signals Observed by Short-Period Seismometers

The conventional seismometry at active volcanoes, except for a few rare cases, has been conducted using short-period seismometers. Even with such band-restricted sensors, wide varieties of waveforms recorded in volcanoes have led scientists to wonder what may be the origins of those signals. **Figure 1** shows a classical yet the most fascinating one of those

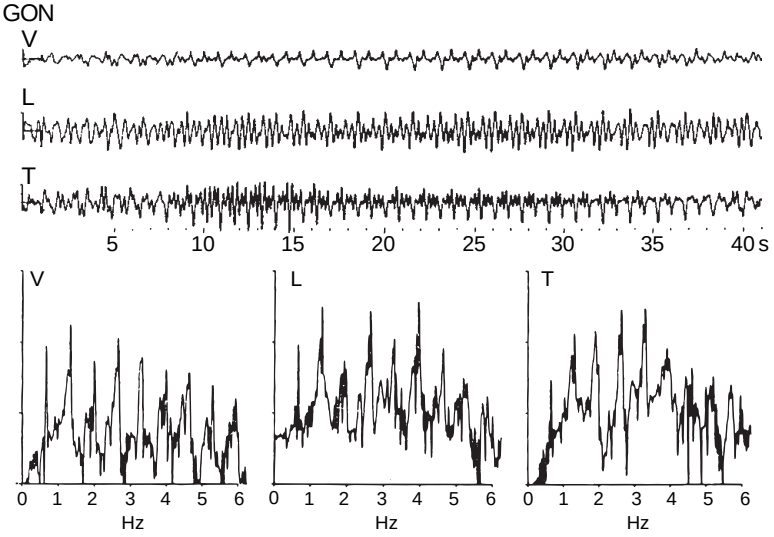


Figure 1 Volcanic tremor observed at Sakurajima. Top three traces show 40 s long seismograms for vertical, longitudinal, and transverse components. Corresponding spectra are shown below. Adapted from Kamo K, Furuzawa T, and Akamatsu J (1977) Some natures of Volcanic micro-tremors at the Sakurajima Volcano. *Bulletin of the Volcanological Society of Japan* 22: 41–58.

recordings of volcanic tremors observed at Sakurajima Volcano, Japan (Kamo *et al.*, 1977). Volcanic tremors are those sustained ground vibrations commonly observed at volcanoes, typically lasting for longer than 1 min, up to hours or days; they often show periodic spectral features as seen here with the lowest peak frequency around 1 Hz. During the course of activities, relative spectral content and peak frequencies may vary from time to time.

Figure 2 shows examples from Redoubt Volcano, Alaska, where different types of events are displayed: from the top, (A) high-frequency or volcano-tectonic (VT) earthquake, which most likely represents an ordinary brittle failure on a fault, (B) hybrid or mixed-frequency event, (C) low-frequency (LF) event with frequency around 1 Hz, and (D) volcanic tremor. Note that the volcanic tremor here appears more random compared to those in **Figure 1**, and also its frequency content seems similar to that of the low-frequency event (C). It is sometimes considered that tremors and low-frequency events share a common excitation mechanism.

Low-frequency events in some volcanoes show characteristic similar spindle-shape waveforms, suggesting a possible common origin for these events in different volcanoes.

Explosion quakes are observed when explosive eruptions occur, and often associated with a high-frequency arrival on seismograms due to a shock wave originated at the vents. **Figure 3** shows series

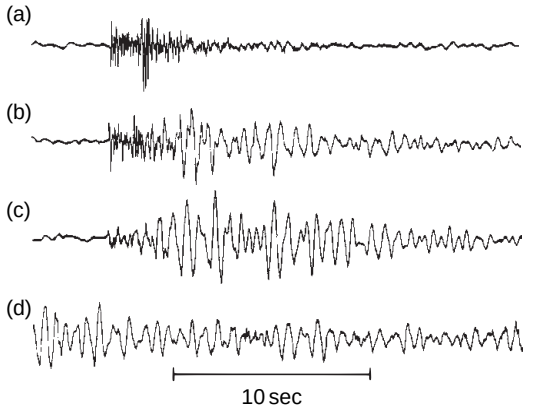


Figure 2 Typical waveforms observed at Redoubt. After McNutt SR (1996) Seismic monitoring and eruption forecasting of volcanoes: A review of the state-of-the-art and case histories. In: *Monitoring and Mitigation of Volcano Hazards*, pp. 99–146. New York: Springer.

of such events recorded in Langila Volcano, Papua New Guinea (Mori *et al.*, 1989), sorted by the ratio of ground-transmitted explosion-wave and air-wave amplitudes. As those without an air wave are classified as low-frequency events, this figure also indicates that explosion quakes and low-frequency events may share a common origin.

As noted earlier, the spectral content of these low-frequency volcanic seismic signals may vary throughout the course of activities. **Figure 4** shows such an example from Montserrat Volcano, West Indies; the temporal change of the system prior to an explosion is clearly manifested.

It appears clear that the low-frequency nature of these seismograms is the most characteristic of volcanic seismic signals. Although they are characterized here just in terms of the frequency content, there exist similarities and diversities among them; their origins may be different from a volcano to another, and may change from time to time even in a single volcano. Understanding of their origins and monitoring of their appearances to infer physical conditions and dynamic states of volcanic edifices and volcanic fluid systems are interesting and challenging tasks.

Deep (volcanic) low-frequency signals. Most of volcanic seismic signals originate within a shallow part of the crust (<5 km). There are, however, a certain type of volcanic seismic signals which originate either from the lower crust or from the uppermost mantle (15–50 km) right beneath active volcanoes. As they show similar low-frequency content as volcanic tremors, they are often called deep low-frequency volcanic tremors (or events). They are now reported in many volcanoes around the world (e.g., Aki and Koyanagi, 1981; Pitt and Hill, 1994; Hasegawa and Yamamoto, 1994; White, 1996; Nakamichi *et al.*, 2003). As deep low-frequency volcanic tremors may be related with deep magma-supplying system of volcanoes, understanding of their origin seems extremely important for the long-term prediction of the activities of volcanoes.

4.13.2.2 Broadband Signals

Availability of broadband seismic records has brought wider perspectives in our view of volcanic activities. **Figure 5** shows a very good example observed at Satsuma-Iwojima Volcano, Japan (Ohminato and Ereditato, 1997). The short-period sensor recording at ST3 shows amplitude modulation

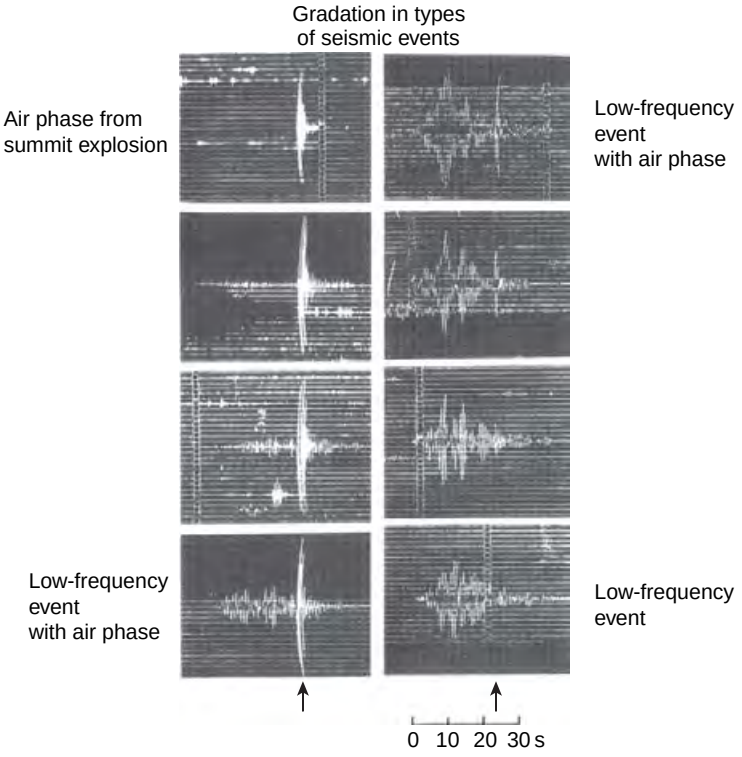


Figure 3 Range of volcanic seismic signals from the high-frequency air-wave phase to LF event observed at Langila. Arrows indicate timings of air-wave phases. Adapted from Mori J, Patia H, and McKee C, *et al*. (1989) Seismicity associated with eruptive activity at Langila Volcano, Papua New Guinea. *Journal of Volcanology Geothermal Research* 38: 243–255.

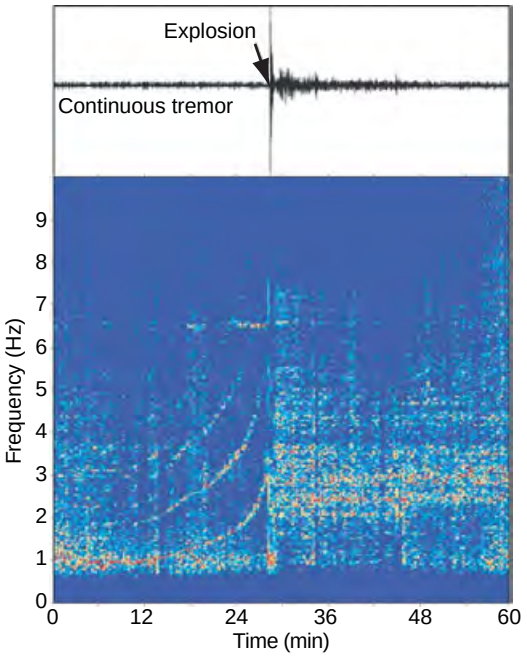


Figure 4 Example of tremor at Montserrat before an explosion. After Jousset P, Neuberg J, and Sturton S (2003) Modelling the time-dependent frequency content of low-frequency volcanic earthquakes. *Journal of Volcanology Geothermal Research* 128: 201–223.

of high-frequency signal with a periodicity of 46–50 min. When this amplitude becomes small, the lowpass-filtered broadband record at ST1 shows long-period (10 s) spike-like signals. This synchronicity of two different band signals continues, while the exact periodicity may fluctuate. Such an interplay of different band signals forces us to view the origin of these signals as a system (i.e., volcanic fluid system).

Figure 6 is another such example of a broadband record of a small phreatic eruption observed at Aso Volcano, Japan (Kaneshima *et al*., 1996; Kawakatsu *et al*., 2000). While the integrated displacement record exhibits a very long-period (>100 s) signal corresponding to the inflation and deflation stages of a crack-like conduit (Yamamoto *et al*., 1999) before and during the eruption, the raw velocity record exhibits short-period signals apparently due to the quick flow of volcanic fluids only seen in the deflation stage (i.e., eruption). Again signals seen in two different bands allow us to vividly image the process of the eruption of a volcanic fluid system. Different physics apply at different frequencies, and this makes volcanic seismic signals essentially broadband.

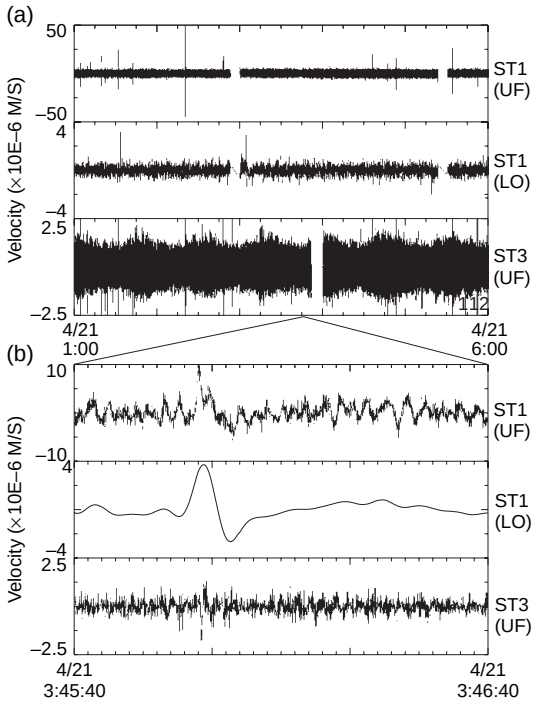


Figure 5 (a) Five-hour records of vertical components of velocity at Satsuma-Iwojima. A trace lowpass-filtered (LO) at 0.2 Hz is shown in between the two unfiltered records. (b) Expansion of the 1 min portion of the top panel. Adapted from Ohminato T and Ereditato D (1997) Broadband seismic observations at Satsuma-Iwojima Volcano, Japan. *Geophysical Research Letters* 24: 2845–2848.

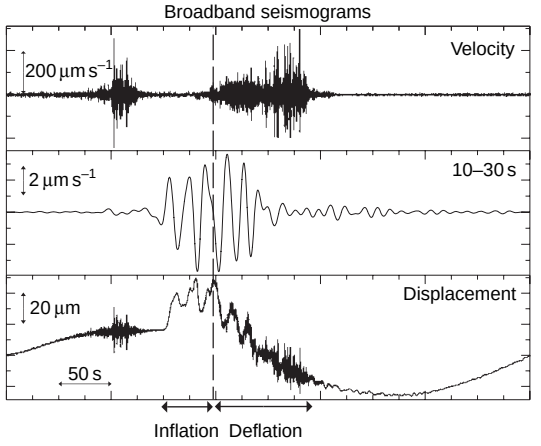


Figure 6 Broadband waveforms of a phreatic eruption observed at Aso. Adapted from Kawakatsu H, Kaneshima S, and Matsubayashi H, *et al*. (2000) Aso94: Aso seismic observation with broadband instruments. *Journal of Volcanology Geothermal Research* 101: 129–154.

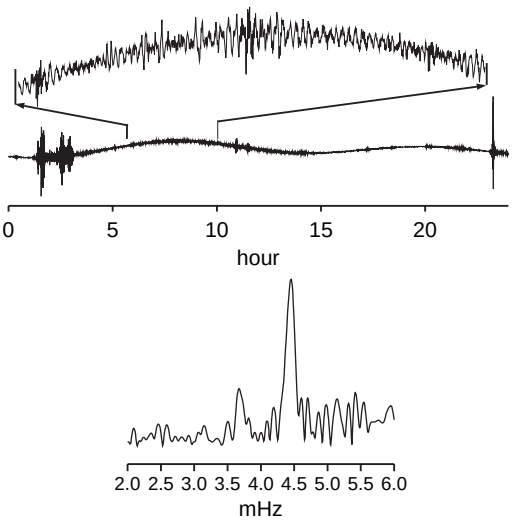


Figure 7 Broadband record of the Pinatubo eruption in 1991 observed in Japan. Adapted from Watada S (1995) *Part I: Near-source Acoustic Coupling Between the Atmosphere and the Soild Earth During Volcanic Eruptions*. PhD Thesis, California Institute of Technology.

Global observation. The last example (**Figure 7**) comes from the recordings of very long-period (200–400 s) seismic signals generated by the eruption of Pinatubo Volcano, Philippines, observed at broadband stations around the world (Kanamori and Mori, 1992; Widmer and Zürn, 1992). The long-period disturbances of the atmosphere generated by the eruption clouds are transmitted back to the solid Earth through the coupling between the two systems. Two spectral peaks observed in the seismic records are well explained by a theory which treats the solid Earth and the atmosphere as a single system (Watada, 1995); such a coupling between the solid and fluid parts of the Earth is also observed in so-called background free oscillations of the Earth (Nishida *et al.*, 2000). The broadband recordings of the Pinatubo eruption appear to have accelerated the emergence of a new class of seismology which treats the solid Earth, atmosphere, and ocean as a single system (cf. Chapter 4.14).

4.13.3 Description of Volcanic Seismic Sources

The diversity of observed volcanic seismic signals partly reflects the wide variety of seismic source geometries and dynamics present in active

volcanoes. A general kinematical description of seismic sources in terms of 'moment tensor' and 'single force' has become common in volcano seismology, and the waveform inversion solving for them has been becoming a general tool to understand the generation mechanisms of long-period signals obtained with broadband instruments.

4.13.3.1 Equivalent Forces of General Seismic Sources

The concept of the seismic moment tensor was first introduced by Gilbert (1970) as a description of seismic sources more general than the conventional double-couple description. Later Backus and Mulcahy (1976) gave a physical basis for such a general description within the framework of linear elastodynamics by introducing concepts of indigenous source and stress glut. An indigenous source is a seismic source which originates within the Earth, and its equivalent force system exerts neither total force nor total torque at any instance to the Earth. Seismologists, however, soon realized that there exist such indigenous seismic events which require single force equivalent force systems (e.g., Kanamori and Given, 1982; Kanamori *et al.*, 1984; Hasegawa and Kanamori, 1987; Kawakatsu, 1989); most of them happened to occur around active volcanoes. To circumvent this apparent conflict, Takei and Kumazawa (1994) extended Backus and Mulcahy's treatment to include fluid-dynamical description with a mass advection term, and introduced a concept of inertial glut; they further gave a clear definition of single force (and torque) source as a momentum (and angular momentum) exchange between the seismic source volume and the rest of the Earth. In the following, we briefly summarize their treatment.

To make the argument simple, we omit the Earth's self-gravity in the following; although gravity often takes the central role generating single forces in the actual Earth, this simplification does not alter the essence of the argument presented below. The true Eulerian equation of motion of the Earth without external force may be written as

$$\rho^{\mathrm{t}} D_t^2 u_j = \partial_i S_{ij} \qquad [1]$$

where $\rho^{\mathrm{t}}(\mathbf{x}, t)$ denotes the true density, $u_j(\mathbf{x}, t)$ is the displacement, $S_{ij}(\mathbf{x}, t)$ is the true stress field, and D_t is the particle derivative $D_t = \partial_t + v_k \partial_k$ with $v_k(\mathbf{x}, t)$ as the velocity. We compare this true equation of motion with the linear elastodynamic equation of motion,

$$\rho^{\mathrm{m}} \partial_t^2 u_j = \partial_i \phi_{ij} \qquad [2]$$

where $\rho^{\mathrm{m}}(\mathbf{x}, t)$ and $\phi_{ij}(\mathbf{x}, t)$ respectively denote the model density and model stress field due to the linear stress–strain constitution equation (Hooke's law), which we use to describe the wave propagation in the Earth. The equivalent body force $\gamma_j^V(\mathbf{x}, \mathrm{t})$ in the linear elastodynamic system arises as the difference of the two equations [1] and [2],

$$\rho^{\mathrm{m}} \partial_t^2 u_j = \partial_i \phi_{ij} + \gamma_j^V \qquad [3]$$

where

$$\gamma_j^V = \gamma_j^S + \gamma_j^D \qquad [4]$$

with

$$\gamma_j^S = -\partial_i \Gamma_{ij} = -\partial_i (\phi_{ij} - S_{ij}) \qquad [5]$$

$$\gamma_j^D = \rho^{\mathrm{m}} \partial_t^2 u_j - \rho^{\mathrm{t}} D_t^2 u_j \qquad [6]$$

$\Gamma_{ij}(\mathbf{x}, t) = \phi_{ij}(\mathbf{x}, t) - S_{ij}(\mathbf{x}, t)$ in [5] is the stress glut introduced by Backus and Mulcahy (1976) which represents failure of Hooke's law due to some nonlinear effect in the source region, and is zero outside of the source region. The second term in [4], $\gamma_j^D(\mathbf{x}, t)$, is named as inertial glut by Takei and Kumazawa (1994), and is the difference of the inertial forces in the actual value and in the model, which is again zero outside of the source region. It can be shown that total force and total torque due to $\gamma_j^S(\mathbf{x}, t)$ integrated over the volume is zero at any instance, while those due to the inertial glut $\gamma_j^D(\mathbf{x}, t)$ are necessarily not.

The spatial integral of [6] over the whole space

$$\gamma_j^{(0)D} = \int (\rho^{\mathrm{m}} \partial_t^2 u_j - \rho^{\mathrm{t}} D_t^2 u_j) \mathrm{d}V \qquad [7]$$

can be shown to be

$$\gamma_j^{(0)D} = -\partial_t^2 \int_{V_s} (\rho^{\mathrm{t}} - \rho^{\mathrm{m}}) u_j \mathrm{d}V \qquad [8]$$

which shows that $\gamma_j^{(0)D}(t)$ originates from a difference between ρ^{t} and ρ^{m} in the source region V_s and thus may not vanish. On the other hand, it is shown that the time integral of $\gamma_j^{(0)D}(t)$ during the event vanishes, that is,

$$\int_0^\infty \gamma_j^{(0)D} \mathrm{d}t = 0 \qquad [9]$$

the requirement that any indigenous source has to satisfy. The authors further defined single force and torque sources as linear and angular momenta

exchange between the source region and the rest of the Earth, and showed that, in the long-period limit, the single force source is equal to $\gamma_j^{(0)D}(t)$.

4.13.3.2 The Seismic Moment Tensor

The spatial integral of stress glut is the seismic moment tensor M_{ij} which describes the overall feature of the seismic source. When the wavelength of observed seismic waves is much longer than the spatial extent of the source, we may approximate the source by a point source moment tensor. Each component of M_{ij} corresponds to one set of opposing forces (dipole or force couple; **Figure 8**). The seismic moment tensor is a symmetric second-order tensor, and thus has six independent components, while a double-couple equivalent body force for a shear dislocation has only four degrees of freedom. These two extra degrees of freedom in the moment tensor are called non-double-couple components. For the general discussion on non-double-couple components, readers may refer to the following review articles (Julian *et al.*, 1998; Miller *et al.*, 1998). Seismic sources observed in the volcanic environments often show non-double-couple components. The end members of such non-double-couple sources are discussed below.

Spherical source. Let us consider an initial condition that a reservoir contains a certain amount of volcanic fluid (V) under a static pressure P. When a sudden volume increase of ΔV (measured under the same pressure condition) is introduced in the reservoir, either due to an injection of a new volcanic fluid or a thermal expansion, the reservoir should expand and act as a seismic source. This volumetric increase ΔV corresponds to the stress-free volumetric strain introduced by Eshelby (1957) (Aki and Richards, 2002, p. 53) and characterizes the amount of the increase of the volcanic fluid. On the other hand, due to the confining pressure of the surrounding elastic medium (matrix), the actual volumetric increase of the reservoir ΔV_m may be smaller than ΔV, the ratio $\Delta V_m/\Delta V$ depending on the geometry of the reservoir, with a resultant pressure increase ΔP_m.

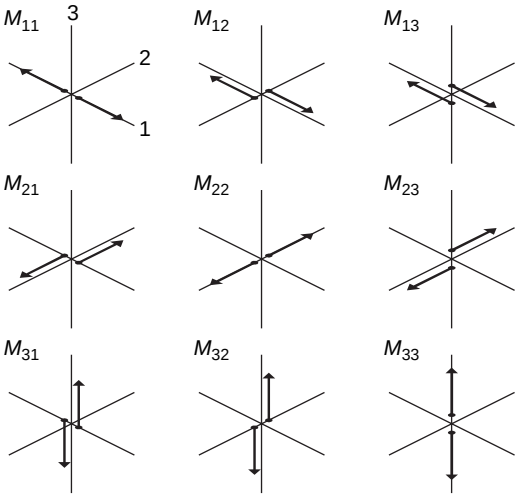


Figure 8 Nine force couples corresponding to moment tensor components.

The corresponding moment tensor for a spherical reservoir is given as

$$M = \Delta V \begin{pmatrix} \lambda + \frac{2}{3}\mu & 0 & 0 \\ 0 & \lambda + \frac{2}{3}\mu & 0 \\ 0 & 0 & \lambda + \frac{2}{3}\mu \end{pmatrix} \quad [10]$$

where the ratio

$$\frac{\Delta V_m}{\Delta V} = \frac{\lambda + (2/3)\mu}{\lambda + 2\mu} \quad [11]$$

and the actual volumetric strain and pressure increase may be related as

$$\Delta P_m = \frac{4}{3}\mu \frac{\Delta V_m}{V} \quad [12]$$

(Aki and Richards, 2002, p. 61). As the actual volumetric change ΔV_m is also the volume change often used in geodetic analysis (e.g., Mogi, 1958), we propose to call this 'Mogi volume' to distinguish it from the stress-free volume increase ΔV.

Cylinder. A moment tensor corresponding to the radial expansion of a cylinder whose symmetry axis is x_1 may be given as

$$M = \Delta V \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda + \mu & 0 \\ 0 & 0 & \lambda + \mu \end{pmatrix} \quad [13]$$

where

$$\frac{\Delta V_m}{\Delta V} = \frac{\lambda + \mu}{\lambda + 2\mu} \quad [14]$$

and

$$\Delta P_m = \mu \frac{\Delta V_m}{V} \quad [15]$$

Tensile crack. A moment tensor corresponding to an opening of a thin crack in the direction of x_1-axis may be given as

$$M = \Delta V \begin{pmatrix} \lambda + 2\mu & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \qquad [16]$$

In case of a thin crack, two volumetric changes turn out to be the same. The Mogi volume here is given as $\Delta V = \Delta V_m = S \cdot \Delta \bar{u}$, where S and $\Delta \bar{u}$ are the crack area and opening width, respectively, and it is often referred to as 'potency' in seismological literatures (e.g., Ben-Menahem and Singh, 1981).

Compensated linear vector dipole. The isotropic components of the moment tensors of these three volumetric sources, when expressed in terms of stress-free volumetric change ΔV, are the same:

$$I = \frac{1}{3}\mathrm{trace}(M) = \left(\lambda + \frac{2}{3}\mu\right)\Delta V = \kappa \Delta V \qquad [17]$$

where κ is the bulk modulus of the elastic matrix. When a movement of magma from a reservoir to another excites seismic waves, such a seismic source should be observed as a summation of two volumetric sources with opposite signs. Equation [17] indicates that the corresponding moment tensor has zero isotropic component called compensated linear vector dipole (CLVD) by Knopoff and Randall (1970) (**Figure 9**).

Determination and interpretation of a moment tensor. Due to structural complexities of volcanic edifices, the conventional first motion polarity analysis method to determine source mechanisms of volcanic events is unlikely to give reliable estimates of moment tensors. Although the amplitude of radiated seismic waves contains more information, often the absolute amplitude also suffers from structural complexities. Methods that utilize relative amplitudes of waves which share common ray paths, such as the spectral ratio method of Nishimura *et al.* (1995), the amplitude ratio method of Julian and Foulger (1996), and the relative moment tensor inversion method of Dahm (1996), appear to be better suited; among them, the linear-programming approach of Julian and Foulger (1996) has been extensively used to show the existence of many non-double-couple earthquakes in different volcanic/geothermal environments (e.g., Ross *et al.*, 1996; Foulger *et al.*, 2004).

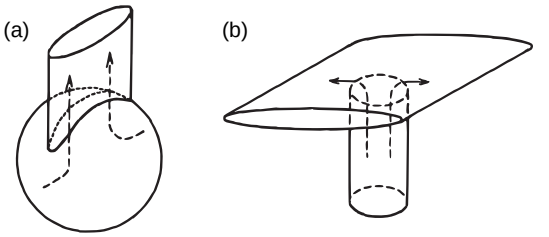


Figure 9 Geometry of selected transport models. Adapted from Chouet B (1996b) New methods and future trends in seismological volcano monitoring. In: *Monitoring and Mitigation of Volcano Hazards*, pp. 23–97 New York: Springer.

An arbitrary moment tensor observed at volcanoes may be decomposed as a linear combination of those four end-member moment tensors (plus a double couple). As such a decomposition is nonunique, there can be different decompositions for a given moment tensor, and thus there exists always nonuniqueness in interpreting moment tensor solutions in terms of actual phenomenon that is occurring at volcanoes. It is often a custom to choose the simplest possible one or to choose one which is most consistent with known source geometry of corresponding volcano (e.g., the presence of known crack-like conduit). Without such *a priori* information, on the other hand, one may wish to display a full moment tensor. The source-type plot introduced by Hudson *et al.* (1989) may be a useful tool to help interpreting different moment tensors observed in different volcanic environments (e.g., Foulger *et al.*, 2004; **Figure 10**). It displays moment tensors without regard to their orientations, and different types of moment tensors (different in relative magnitudes of principal moments) are projected on a rhombic plane in such a way that the area is proportional to the probability of the occurrence of different source types.

4.13.3.3 The Single Force

As described above, the single force equivalent body force is due to a momentum exchange between the source volume and the rest of the Earth. Whenever a part of the Earth is detached from the rest and gains a momentum, the counter force of this acceleration is felt by the rest of the Earth. As the detached mass eventually has to stop somehow, there must be a deceleration stage in this process, giving another counter force with the opposite direction. The total of these two counter forces must cancel out as required from [9].

The single force equivalence of long-period seismic waves excited by a landslide and explosions associated with the 1980 eruption of Mt. St. Helens was established by Kanamori and Given (1982) and

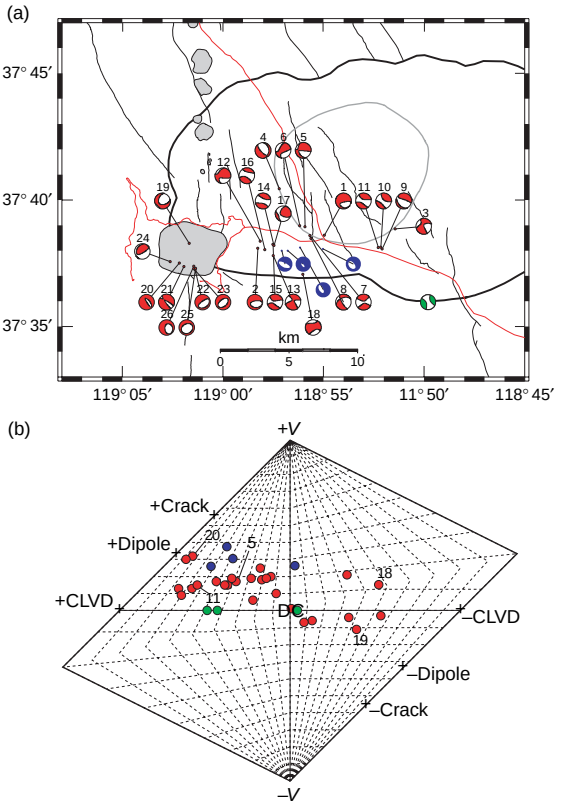


Figure 10 (a) Focal mechanism plot (upper focal hemisphere) of earthquakes in Long Valley. (b) Source-type plot of Hudson *et al*. (1989) for the same earthquakes. The vertical axis measures the size of volumetric component, and deviatoric moment tensors are located on the horizontal axis (a pure double couple at the origin). Different colors indicate mechanism solutions determined by different researchers. Adapted from Foulger GR, Julian BR, Hill DP, Pitt AM, Malin PE, and Shalev E (2004) Non-double-couple microearthquakes at Long Valley caldera, California, provide evidence for hydraulic fracturing. *Journal of Volcanology and Geothermal Research* 132: 45–71.

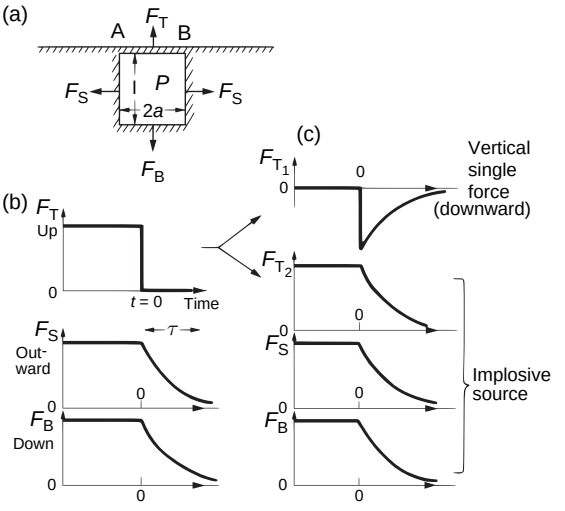


Figure 11 A force system equivalent to a volcanic eruption. (a) A pressure release model for a volcanic eruption. (b) Forces acting on the top, side, and bottom walls. (c) Decomposition of the force to a vertical single force and an implosive force. Adapted from Kanamori H, Given JW, and Lay T (1984) Analysis of seismic body waves excited by the Mount St. Helens eruption of May 18, 1980. *Journal of Geophysical Research* 89: 1856–1866.

Kanamori *et al.* (1984). At a long-period limit, a landslide may be viewed as a sliding of a box along a slope (Hasegawa and Kanamori, 1987; Kawakatsu, 1989), and its single force equivalence may be obvious. Kawakatsu (1989) developed a centroid single force (CSF) inversion method, an extension of the centroid moment tensor (CMT) inversion method of Dziewonski *et al.* (1981) to include single force components, and applied it successfully to several landslide/slump events recorded by global seismic networks. Ekström *et al.* (2003) applied the CSF inversion to 'glacial earthquakes' observed in southern Alaska, and showed that they are represented by stick-slip, downhill sliding of a glacial ice mass which takes about 30–60 s in duration.

Kanamori *et al.* (1984) also suggested that an equivalent force system for a volcanic eruption can be decomposed into a vertical downward single force and an isotropic moment tensor (**Figure 11**). A single force source may be also realized when high-viscous magma flows along a narrow conduit, as suggested by Ukawa and Ohtake (1987) to explain the deep monochromatic low-frequency (1 Hz) event observed beneath Izu-Oshima Volcano, Japan, about 1 year prior to the 1986 eruption. Chouet *et al.* (2003) also suggested that a piston-like rise of a slug of gas in the conduit produced observed single force components in the long-period signals associated with explosions in Stromboli Volcano, Italy.

4.13.3.4 Waveform Analysis of Volcanic Seismic Sources

The long-period part of volcanic seismic signals has been recorded by seismic networks of regional and/or global scales. One of the advantages of using such long-period signals is that waveform analysis to infer excitation mechanisms can be easily performed compared to short-period seismograms. As mentioned, Kanamori and Given (1982) modeled the long-period seismic signals accompanied with the eruption of the Mt. St. Helens in 1980 recorded at stations of the

global seismic networks. They showed that the single force mechanism can explain the radiation pattern of long-period surface waves, and attributed it to the gigantic mass movement of the volcano edifice due to a landslide. Kanamori *et al.* (1984) also analyzed the long-period body waves generated by the same eruption, and showed that vertical single forces due to the depressurization of the magma chamber explain the data. With these two single force mechanisms, they could vividly picture what had occurred during the whole sequence of the eruption. These two studies initiated the applications of 'long-period seismology' to study volcanic activities.

For long-period seismic signals, it is often appropriate to treat the source originating from a point in space. Then seismic wave field u_n ($\mathbf{x}$, t) may be written as

$$u_n(\mathbf{x}, t) = G_{np,q}(\mathbf{x}, \mathbf{x}_0, t) * M_{pq}(t) + G_{np}(\mathbf{x}, \mathbf{x}_0, t) * F_p(t)$$

where $M_{pq}(t)$ and $F_p(t)$ denote a moment tensor and single-force source time function at $\mathbf{x}_0$, respectively, and G_{np} ($\mathbf{x}$, $\mathbf{x}_0$, t) and $G_{np,q}$ ($\mathbf{x}$, $\mathbf{x}_0$, t) respectively represent Green's function and its spatial derivative (e.g., Aki and Richards, 2002); the asterisk denotes convolution, and the summation convention for repeated indices is assumed. Knowing the Earth structure (i.e., knowing the Green's function) and having enough waveform data, it is then possible to solve for the time history of general seismic sources, $M_{pq}(t)$ and $F_p(t)$.

Takeo *et al.* (1990) extended the waveform moment tensor inversion technique commonly used in the analysis of far-field records in earthquake seismology (e.g., Kikuchi and Kanamori, 1982) to volcano seismology by including the single force contribution. They applied it to the records observed in a lava drain-back stage of the 1987 eruption of Izu-Oshima Volcano, Japan, and showed that the resolved vertical single forces with opposing directions are due to a rapid collapse of overloading lava into the deeper vent. Ohminato *et al.* (1998) applied a similar technique to near-field broadband records observed at Kilauea Volcano, USA, and showed that long-period signals can be explained by a subhorizontal crack which periodically injects magma into a larger reservoir. Chouet *et al.* (2003, 2005) and Ohminato (2006) further incorporated the effect of complex topography of volcanic edifice that may significantly distort seismic signals observed in the near-field by calculating Green's functions using the finite difference scheme developed by Ohminato and Chouet (1997).

The general source analysis for long-period signals using moment tensor and single force has become common in seismological studies of volcanic activities. Examples of such studies include Mt. St. Helens (Kanamori and Given, 1982; Kanamori *et al.*, 1984; Kawakatsu, 1989); Long Valley (Julian, 1983; Aki, 1984; Wallace, 1985; Dreger *et al.*, 2000); Kilauea, Hawaii (Eissler and Kanamori, 1987; Ohminato *et al.*, 1998), USA; Asama (Takeo *et al.*, 1984; Ohminato *et al.*, 2006); Izu-Oshima (Takeo *et al.*, 1990); Sakurajima (Uhira and Takeo, 1994); Unzen (Yamasato *et al.*, 1993; Uhira *et al.*, 1994b); Ito-oki (Takeo, 1992); Tori-shima (Kanamori *et al.*, 1993), Japan; Iceland (Julian *et al.*, 1997; Nettles and Ekström, 1998); Stromboli, Italy (Chouet *et al.*, 2003); and Popocatépetl, Mexico (Chouet *et al.*, 2005). Applications of such analysis techniques for shorter-period (~1 s) near-field records obtained at volcanoes are also becoming available (e.g., Uhira *et al.*, 1994a; Nishimura *et al.*, 1995; Aoyama and Takeo, 2001; Nakamichi *et al.*, 2003; Nakano *et al.*, 2003; Ohminato, 2006; Kumagai *et al.*, 2005).

4.13.4 Physical Mechanisms for Volcanic Seismic Signals

The presence of ample fluid (gas, vapor, magma, or their mixtures) in volcanic edifice introduces an additional interesting class of vibration/wave phenomena to the well-studied elastic formulations. For example, the so-called 'crack wave', which is introduced by Chouet (1986) to explain the low-frequency nature of volcanic events, may be understood as a part of a class of waves due to solid/fluid coupling. Julian (1994) showed that the nonlinear effect of such solid/fluid coupling due to fluid flow transient can generate a wide variety of waveforms which have similar characteristics as volcano seismic signals. This section summarizes some such topics, relevant to generation of volcano-specific seismic signals.

4.13.4.1 Slow Waves in Solid/Fluid Composite

There exists a class of waves in the two-phase system of a solid–liquid composite that travel slower than any of the sound velocities of the pure material constituting the two-phase system (i.e., the wave speed

can be slower than that of the liquid phase). These waves include the tube wave (Biot, 1952), so-called Biot's slow wave in a porous medium (Biot, 1956), so-called crack wave (Chouet, 1986; Ferrazzini and Aki, 1987), and waves in solid–liquid alternating layers (Schoenberg and Sen, 1983) (Rayleigh wave and Stoneley wave may also be included in the same class, considering that in the high-frequency limit, all the above slow waves have the same characteristics as an interface wave).

The simplest (and the most relevant to volcano seismology) example of slow waves may be the plane wave propagation in a medium which consists of a fluid layer sandwiched by two elastic half-spaces studied by Ferrazzini and Aki (1987), who investigated the physical basis of the crack wave found by Chouet (1986). Here we follow Yamamura (1997) to formulate the problem.

Following Aki and Richards (2002, p. 263), we first express a plane-wave solution propagating parallel to the fluid layer as

$$\begin{pmatrix} u \\ w \end{pmatrix} = \begin{pmatrix} r_1(z) \\ ir_2(z) \end{pmatrix} e^{i(kx-\omega t)} \qquad [18]$$

where u and w are respectively the displacements for x- and z-directions. The frequency domain equation of motion in x-direction for the fluid layer may then be written as

$$-\omega^2 \rho_f r_1 = -k^2 \kappa_f r_1 - k\kappa_f \frac{\partial r_2}{\partial z} \qquad [19]$$

where ρ_f and κ_f denote density and bulk modulus of the fluid, respectively. Integrating [19] over the thickness of the fluid layer (i.e., z-direction) with appropriate boundary conditions (e.g., free-slip condition at the solid/fluid interface), we get the following expression:

$$-\omega^2 \rho_f \int_0^H r_1 \mathrm{d}z = -k^2 \kappa_f \int_0^H r_1 \mathrm{d}z - k\kappa_f r_2(H) \qquad [20]$$

where we assume a symmetric solution for the z-dependency (Ferrazzini and Aki, 1987) and the thickness of the fluid layer $2H$.

Equation [20] clearly demonstrates the role of the interface: without the last term $r_2(H)$, [20] indicates the averaged displacement in the fluid layer traveling with its own sound velocity; that is, the term represents the way the fluid layer is coupled with the surrounding elastic half-space. If we further rewrite [20] as

$$-\omega^2 \rho_f \int_0^H r_1 \mathrm{d}z = -k^2 \kappa_f \left[1 + \frac{1}{k} \frac{r_2(H)}{\int_0^H r_1 \mathrm{d}z} \right] \int_0^H r_1 \mathrm{d}z \qquad [21]$$

it is seen that

$$\kappa_f^e \equiv \kappa_f \left[1 + \frac{1}{k} \frac{r_2(H)}{\int_0^H r_1 \mathrm{d}z} \right]$$

represents the effective bulk modulus for the fluid layer; when $\kappa_f^e < \kappa_f$, that is,

$$\frac{r_2(H)}{\int_0^H r_1 \mathrm{d}z} < 0 \qquad [22]$$

the corresponding traveling wave has a smaller wave speed relative to the sound velocity; thus we have a slow wave. The condition [22] is satisfied if the denominator and numerator have an opposite sign to each other. This makes quite good sense intuitively; when the fluid 'tries' to contract (or expand) as a passage of the wave, due to the 'elastic coupling' along the solid/fluid boundary, there exists a solution in which the solid behaves out of phase to the fluid to expand (contract), deforming the boundary in such a way that fluid 'feels' easy to contract (expand), compared to the single-phase case; that is for a given amount of displacement in the propagation direction, the actual pressure increment of the fluid phase becomes lower than that of the pure fluid case (**Figure 12**). The net effect of this elastic coupling

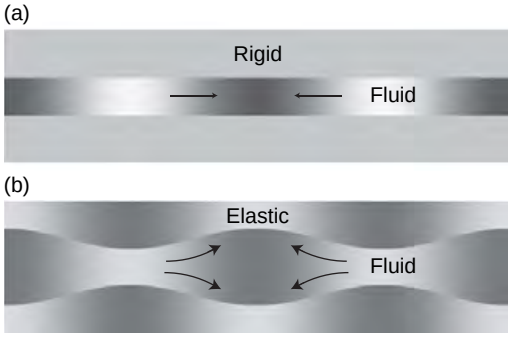


Figure 12 Conceptual image of the slow wave in solid/fluid composites. Out-of-phase motion of the solid (b) lowers the 'effective bulk modulus' of the fluid, and as a result the wave speed becomes slower than that in the pure fluid (a).

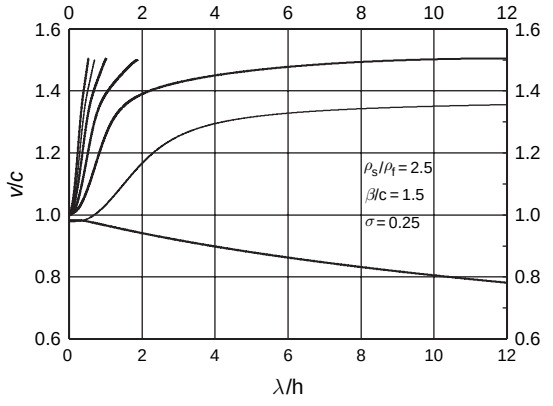


Figure 13 Phase velocity divided by the acoustic velocity of the fluid is plotted as a function of the wavelength divided by the fluid layer thickness. The fundamental branch having a phase velocity below 1 corresponds to the slow wave. Adapted from Ferrazzini V and Aki K (1987) Slow waves trapped in a fluid-filled infinite crack: Implication for volcanic tremor. *Journal of Geophysical Research* 92: 9215–9223.

is to reduce the 'effective bulk modulus' (i.e., the restoring force) of the fluid, and as a result the wave speed becomes slower than that of the pure fluid; stronger the elastic coupling, slower the wave speed (**Figure 13**).

This kind of the out-of-phase condition is satisfied by other slow waves such as Biot's slow wave in porous medium (Biot, 1956), and waves in solid/fluid alternating layers (Schoenberg and Sen, 1983). Indeed, using the argument presented above, Yamamura (1997) was able to derive a similar set of equations to describe the elastic coupling for the Schoenberg's problem to those Biot's equations of motion, and suggested that all these slow waves have a similar characteristic as an interface wave, and that in the short-wavelength limit they reduce to the Stoneley wave.

4.13.4.2 Resonating Sources: A Crack

The presence of regularly spaced spectral peaks often observed in volcanic seismic signals immediately reminds us of the presence of resonators within the volcanic edifice filled with volcanic fluid. To quantitatively model the volcanic tremor observed at Kilauea, Hawaii, Aki *et al.* (1977) considered fluid-filled tensile cracks as the most plausible resonating sources. Using two-dimensional finite difference method, they computed the dynamics of a series of fluid-filled cracks connected by narrow channels which are excited by jerky extension/opening of channels due to excess fluid pressure. Although their model offers both the driving force and the geometry adequate for the magma transport beneath active volcanoes, the fluid inside the crack acts passively like a cushion so that the stress at the crack wall depends only on local displacements of the crack wall; the fluid also dose not support the propagation of pressure perturbation caused by the deformation of the crack wall. To assess active roles of the fluid, Chouet and Julian (1985) extended Aki *et al.*'s model, and investigated dynamic interaction between fluid and elastic solid inside and outside a two-dimensional crack by solving equations for the elastic solid and the fluid simultaneously. The model was further extended to three-dimension by Chouet (1986).

Chouet's three-dimensional model consists of a single isolated fluid-filled crack in an infinite elastic solid body, and no external mass transfer into and/or out of the crack is taken into consideration. In the model, the thickness of the crack d is assumed to be much smaller than the wavelength of interest, and thus the motion of the fluid inside the crack is treated as two-dimensional. Assuming Poisson's ratio to be 0.25 (i.e., $\lambda = \mu$) and taking L and L/α as the characteristic length and time, where L and α are the length of the crack and the P wave velocity in the solid, the response of the crack is calculated by solving following dimensionless equations simultaneously with boundary conditions at the crack surface and perimeters:

$$\begin{aligned}
\text{Solid:}\quad & \frac{\partial \tilde{v}_i}{\partial \tilde{t}} = \frac{1}{3}\frac{\partial \tilde{\tau}_{ik}}{\partial \tilde{x}_k} \\
& \frac{\partial \tilde{\tau}_{ij}}{\partial \tilde{t}} = \frac{\partial \tilde{v}_k}{\partial \tilde{x}_k}\delta_{ij} + \frac{\partial \tilde{v}_i}{\partial \tilde{x}_j} + \frac{\partial \tilde{v}_j}{\partial \tilde{x}_i} \\
\text{Fluid:}\quad & \frac{\partial \tilde{U}_l}{\partial \tilde{t}} = -\frac{1}{3}\frac{\rho}{\rho_f}\frac{\partial \tilde{P}}{\partial \tilde{x}_l} \\
& \frac{\partial \tilde{P}}{\partial \tilde{t}} = -\frac{\kappa_f}{\mu}\frac{\partial \tilde{U}_m}{\partial \tilde{x}_m} - 2\frac{\kappa_f}{\mu}\frac{L}{d}\tilde{v}_z \\
& (\tilde{x} = x/L,\ \tilde{t} = t\alpha/L)
\end{aligned} \qquad [23]$$

In these equations, the quantities with tilde represent dimensionless ones, and $\tilde{v}$ and $\tilde{\tau}$ are the velocity and the stress of the solid; $\tilde{U}$ and $\tilde{P}$ are the velocity and the pressure of the fluid, ρ and ρ_f are the density of the solid and the fluid, μ and κ_f are the rigidity of the solid and the bulk modulus of the fluid, respectively. As shown in [23], the behavior of fluid-filled crack critically depends on a nondimensional parameter $c = \kappa_f/\mu \cdot L/d$, which was first introduced by Aki *et al.* (1977) and named as 'crack stiffness' by Chouet (1986).

Synthetic seismograms calculated by the fluid-filled model show strong similarities to observed volcanic signals in terms of both peaked spectra and long-lasting oscillations, and demonstrated the importance of active participation of fluids in source dynamics generating volcanic seismic signals. These studies include the interpretations of volcanic signals observed at Redoubt, Alaska (Chouet *et al.*, 1994); Galeras, Colombia (Gil Cruz and Chouet, 1997, **Figure 14**); Kusatsu-Shirane, Japan (Nakano *et al.*, 1998; Kumagai *et al.*, 2002); Kilauea, Hawaii (Kumagai *et al.*, 2005); and other volcanoes.

Kumagai and Chouet (1999, 2000) and Morrissey and Chouet (2001) further studied the dependencies of frequencies and attenuations of crack-resonant oscillations on the properties of fluid in more detail using various models of fluid–gas mixtures and fluid–particle mixtures, and demonstrated that the fluid properties and compositions can be estimated from frequencies and attenuations of observed signals. These results are used to assess the change of magmatic and hydrothermal system beneath the volcano at Kusatsu-Shirane, Japan (Kumagai *et al.*, 2002; **Figure 15**) and Tungurahua, Ecuador (Molina *et al.*, 2004).

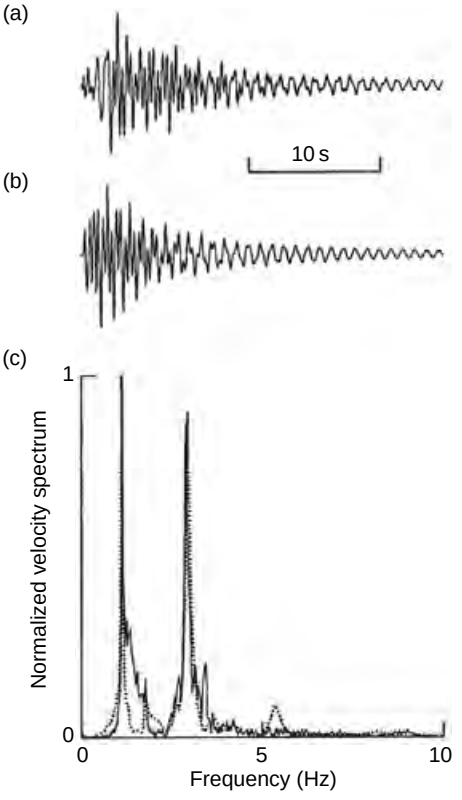


Figure 14 Example of the application of the fluid-filled crack model: (a) Vertical velocity seismogram observed at Galeras Volcano, Colombia. (b) Synthetic waveform obtained from the fluid-filled crack model. (c) Spectrum of observed data (thin line) and synthetic data (dotted line). Adapted from Chouet B (1996a) Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature* 380: 309–316.

4.13.4.3 Other Resonating Sources

Although the presence of regularly spaced spectral peaks may not straightforwardly warrant the presence of a resonator (Julian, 1994), it seems indeed natural to think of such presence like a crack as a dike, a cylinder as a conduit, and a sphere as a magma chamber, as they are all fundamental constituents of a volcano.

Resonators as models of the origins of volcanic seismic signals besides a crack have been suggested for a cylinder (Ferrick *et al.*, 1982; Chouet, 1985), and a sphere (Sassa, 1935; Kubotera, 1974). Crosson and Bame (1985) obtained an impulsive response of a spherical magma chamber with a spherical cavity inside. Fujita *et al.* (1995) solved eigenoscillation of a fluid sphere embedded in an infinite elastic medium. Fujita and Ida (2003) conducted a systematic survey of the geometrical effects on the eigenoscillations, and showed that the ratios of higher modes to the fundamental mode frequencies can be indicators of the geometry of the resonator. They also found a class of modes with a low attenuation named 'low-attenuation mode' (LAM), which share the similar out-of-phase characteristics of the slow waves discussed above, and suggested that LAM may be the origin of long-lasting low-frequency oscillation of volcanic seismic signals.

The complicated nature of vibrations of cylindrical conduits has been numerically studied (e.g., Neuberg *et al.*, 2000; Jousset *et al.*, 2003; Nishimura and Chouet, 2003). Neuberg *et al.* (2000) introduced a depth-dependent seismic velocity model to account for the varying gas content in the magma, and showed that a pressure change in the conduit was the most likely candidate for the physical process causing the peak-spectral shift often observed in many of volcanoes (**Figure 4**). Jousset *et al.* (2003) also showed that the end points of the conduit can act as secondary sources of volcanic seismic signals.

The excitation mechanism of these resonators is, on the other hand, not well understood. The proposed models include a jerky extension of the crack tip (Aki *et al.*, 1977), acoustic emissions from collapsing bubbles (Chouet, 1992), rapid discharge of fluids

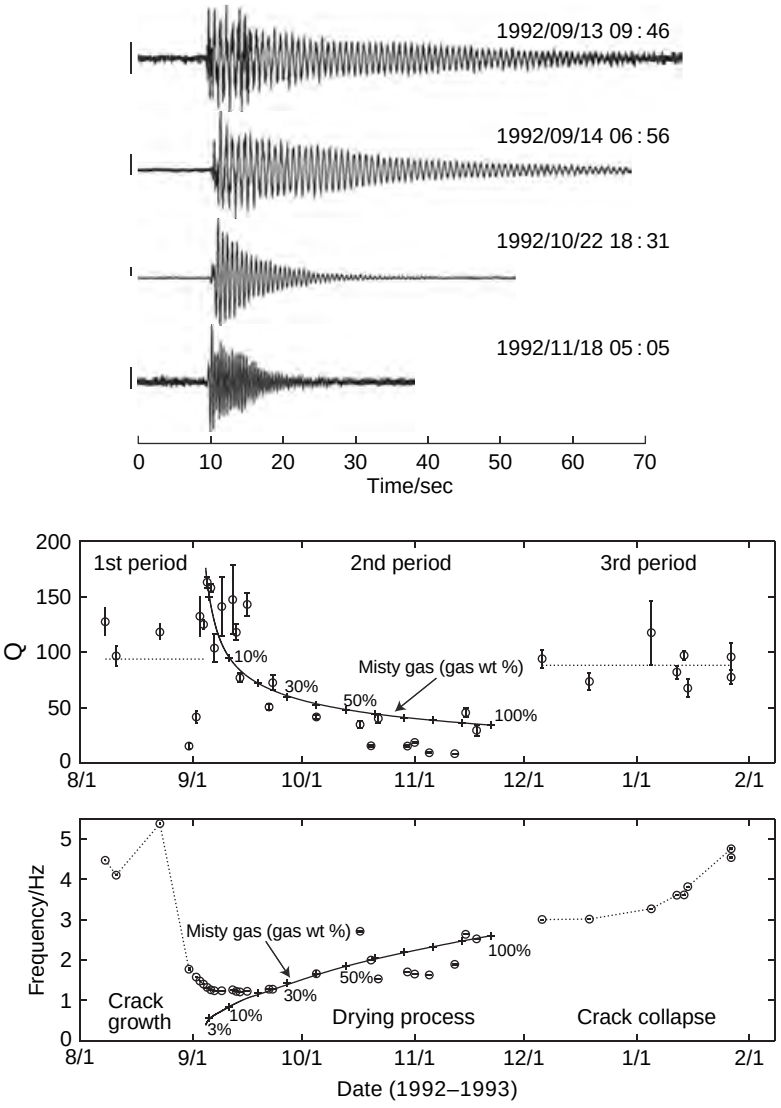


Figure 15 Temporal change of waveforms observed at Kusatsu-Shirane Volcano and the fluid properties estimated from the frequency and the attenuation of signals. From Kumagai H, Chouet B, and Nakano M (2002) Temporal evolution of a hydrothermal system in Kusatsu-Shirane Volcano, Japan, inferred from the complex frequencies of long-period events. *Journal of Geophysical Research* 107: ESE 9–1.

(Ferrick *et al.*, 1982), and a sudden pressure drop caused by unsteady choked flow (Morrissey and Chouet, 1997). Nakano *et al.* (1998) developed a method to extract the effective excitation function from seismic records based on an inhomogeneous autoregressive (AR) model of a linear dynamic system. Nakano *et al.* (2003) further extended the method to deal with the nonorthogonal nature of eigenfunctions of a resonator embedded in a rock matrix by using the biorthogonal eigenfunction expansion method of Yamamura and Kawakatsu (1998); the resolved effective excitation functions of the low-frequency events in Kusatsu-Shirane Volcano, Japan, are then modeled by the waveform inversion method of Ohminato *et al.* (1998) to obtain the moment tensor and single-force source time function, which are interpreted as a result of repeated activation of a subhorizontal crack located beneath the summit crater lake.

4.13.4.4 Flow-Induced Oscillation

Julian (1994) proposed nonlinear flow-induced oscillations in channels transporting volcanic fluid as a

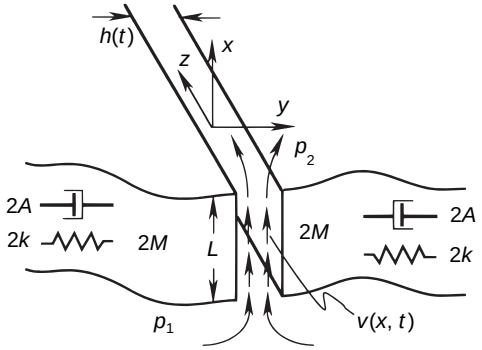


Figure 16 Lumped-parameter model of the generation of volcanic tremor. Viscous, incompressible fluid flows in x-direction from upstream (bottom) to downstream (top) reservoir through a channel with imperfectly elastic walls. Adapted from Julian BR (1994) Volcanic tremor: Nonlinear excitation by fluid flow. *Journal of Geophysical Research* 99: 11859–11878.

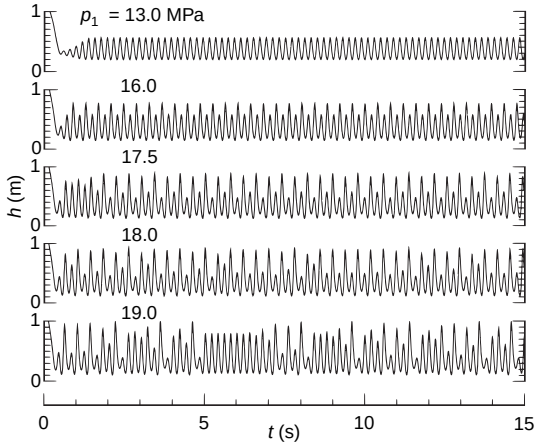


Figure 17 Synthetic tremor time series (channel thickness h versus time) for different values of the upstream pressure p_1. Behavior changes from a simple limit cycle (top) to chaos (bottom), with some period doubling in between. Adapted from Julian BR (1994) Volcanic tremor: Nonlinear excitation by fluid flow. *Journal of Geophysical Research* 99: 11859–11878.

possible generating mechanism for volcanic seismic signals, particularly for sustained volcanic tremors. The mechanism may have analogs such as the vibration of vocal chords, musical wind instruments, etc., and the periodic behavior, often characteristic of volcanic seismic signals, occurs without the presence of a resonator.

Figure 16 shows the simplified lumped-parameter model of the flow of a viscous fluid through a constricted channel with elastic walls. The channel walls are modeled as masses connected by a spring and dashpot to represent the effects of the elastic stiffness and radiation loss. The mechanism may be summarized as follows: an increase in a flow speed leads to a decrease in fluid pressure by the Bernoulli effect; as a result, the walls narrow the channel to constrict the flow, causing a pressure increase and forcing the channel open again; this again decreases the pressure and increases the flow speed. This positive feedback mechanism sustains the oscillation.

Julian (1994) derived a third-order nonlinear system of ordinary differential equations that describe the complete behavior of the system and may be solved numerically. The system without radiation damping exhibits two different types of behavior: oscillations that grow and approach a stable tremor-like limit cycle, and transient oscillations that decay, leading to a steady flow. Inclusion of the damping further complicates the behavior, as shown in **Figure 17**. By increasing the upstream pressure, the system experiences so-called period doubling, that is, an appearance of a subharmonic with a frequency half that of the fundamental frequency; further increase with successive period doubling eventually leads to chaotic behaviors. Julian suggests that these features resemble behavior of some of the volcanic tremors and the deep low-frequency event observed in Hawaii (Aki and Koyanagi, 1981).

Balmforth *et al.* (2005) extended Julian's treatment to include the dynamic behavior of the fluid and the elasticity of the surrounding walls, and considered the excitation of propagating waves. Their results confirm that Julian's mechanism is plausible, but the physical condition required for the instability is different. They conclude that magma itself is unlikely to generate flow-induced oscillations, but that the rapid flow through fractured rock of low-viscosity fluids exsolved from magma may. A similar flow-induced oscillation is also studied by Ida (1996) to explain the sawtooth nature of geodetic signals.

In fluid/gas two-phase system. The flow-induced oscillation described above assumes flows of a single-phase fluid. Volcanic fluids such as magma often contain ample gases, especially at a shallow depth, and it may be more reasonable to consider a flow of two-phase fluids consisting of liquid and gas phases rather than single-phase fluids as a source of volcanic seismic signals. Such a flow of liquid/gas two-phase fluid is also seen in, for example, the evaporation pipes in boiler systems, nuclear reactor cooling systems,

chemical industry equipments, and many studies have been conducted extensively in the field of engineering. These studies revealed the existence of flow-induced oscillatory phenomena, which is not observed in a flow of single-phase fluid, caused by complex interactions between two phases.

Iwamura and Kaneshima (2005) considered one of these flow-instability phenomena called 'density wave oscillation', which was originally studied in the nuclear reactor engineering, as a model of volcanic seismic signals, and evaluated the properties of generated seismic waves through a series of numerical experiments. Another kind of flow-induced oscillation in fluid/gas two-phase flows called 'pressure drop oscillation' was also invoked to interpret the cyclic tilt change observed at Miyake-jima Volcano by Fujita *et al.* (2004).

4.13.4.5 Bubble Dynamics

Gasses dissolved in magma are one of the main driving forces to cause various volcanic phenomena. The presence of gas bubbles in magma decreases the density of magma, and drives the magma ascent from the deep part of volcanoes (e.g., Wilson and Head, 1981). At shallower depths, the relatively high compressibility of bubbles contributes to the generation of oscillatory phenomena specific to volcanoes (e.g., Kieffer, 1977; Chouet, 1996a), and the transformation and the fragmentation of bubbles affect the behavior of volcanic eruptions (e.g., Alidibirov and Dingwell, 1996; Ichihara *et al.*, 2002).

In relation to volcano seismology, there are some aspects where bubbles play important roles. As described above, volcanic seismic signals are often characterized by long-lasting low-frequency oscillations with sharp spectral peaks (**Figure 1**), which are reminiscent of oscillations of some resonator filled with material with a slow acoustic velocity. Since the existence of bubbles in magma drastically changes the compressibility of magma, as pointed out by Kieffer (1977), these characteristics of volcanic seismic signals have been partly attributed to the presence of bubbles in magma. The rapid temporal changes of the observed spectral peaks (e.g., Benoit and McNutt, 1997; Neuberg *et al.*, 2000) have also been considered to reflect the change in the state of gas bubbles in volcanic conduits.

Bubbles in magma also act as an active source of seismic and acoustic signals: the bursting and the collapsing of bubbles at the lava surface preceding and/or accompanying volcanic eruptions generate acoustic pressure signals which are sometimes accompanied with seismic waves (e.g., Vergniolle *et al.*, 1996; Ripepe *et al.*, 1996); the collapse of small bubbles in hydrothermal fluids generates seismic tremors (e.g., Leet, 1988); the movement of large-scale bubbles (slugs) in magma conduit causes an exchange of their momentum with the surrounding rocks and generates seismic signals that show dominant single force components (e.g., Ohminato *et al.*, 1998; Ripepe and Gordeev, 1999; Chouet *et al.*, 2003, 2005); the forced oscillations of bubbles by strain waves can rapidly pump dissolved volatiles into bubbles by a mechanism called 'rectified diffusion' and the resultant pressure increase may trigger volcanic activities (e.g., Brodsky *et al.*, 1998; Ichihara and Brodsky, 2006).

In spite of the difference in their generation mechanisms and resultant phenomena, all of these phenomena are the manifestations of passive and active processes in which the thermochemical energy and/or the gravitational energy of the volcanic fluids are transduced into seismic/acoustic energy. The quantitative understanding of the behavior of bubbles, thus, is a critical key to elucidate the dynamics of volcanic phenomena.

4.13.5 Observation and Analysis Aspects

In addition to the conventional monitoring of active volcanoes using short-period seismometers and geodetic instruments, the digital era has brought new powerful components to seismometry of active volcanoes; broadband and array seismometries may be two of the most notable developments.

4.13.5.1 Broadband Seismometry

The first broadband seismic observation at an active volcano was conducted by Sassa (1935), who installed Wiechert horizontal-component seismographs (pendulum period $T_0 = 10.0$ s) and vertical ones ($T_0 = 4.6$ s), Galitzin seismographs ($T_0 = 8.0$ s), and short-period seismographs ($T_0 = 0.55$ s) at Aso Volcano. Summarizing the subsequent observations in a period of several years, Sassa classified volcanic seismic signals of Aso into 'eruption earthquakes' and four different kinds of tremors: periods of 0.2, 0.4–0.6, 0.8–1.5, and 3.5–8.0 s. The observation of these tremors in the wide frequency band ranging from 0.2 to 8.0 s clearly demonstrates the efficacy of broadband

seismometry at active volcanoes. Since Sassa's pioneering work, until 1990s (when it has become much easier to install portable broadband instruments), there have not been many published reports on attempts to observe long-period volcanic seismic signals except for a few cases. For example, Seidl *et al.* (1981) installed broadband seismometers at Etna, Italy, and observed 4–5 s period signals, which they suggested as one of the fundamental peaks of the Etna tremors.

Examples of deployments of the current-generation broadband seismometers are now plentiful, for example, Sakurajima (Kawakatsu *et al.*, 1992, 1994) Unzen (Yamasato *et al.*, 1993; Uhira *et al.*, 1994b) Aso (Kaneshima *et al.*, 1996; Yamamoto *et al.*, 1999; Kawakatsu *et al.*, 2000; Legrand *et al.*, 2000) Satsuma-Iwojima (Ohminato and Ereditato, 1997; Ohminato, 2006) Iwate (Nishimura *et al.*, 2000); Miyake-jima (Kumagai *et al.*, 2001; Fujita and Ida, 2003), Usu (Yamamoto *et al.*, 2002) Bandai (Nishimura *et al.*, 2003) Hachijo-jima (Kumagai, 2006) Asama (Ohminato *et al.*, 2006), Japan; Stromboli (Dreier *et al.*, 1994; Falsaperla *et al.*, 1994; Neuberg *et al.*, 1994; Chouet *et al.*, 2003), Italy; Semeru (Hellweg *et al.*, 1994) Merapi (Hidayat *et al.*, 2000, 2002), Indonesia; Arenal (Hagerty *et al.*, 2000), Costa Rica; Kilauea (Dawson *et al.*, 1998; Ohminato *et al.*, 1998; Almendros *et al.*, 2002a), Long Valley (Hill *et al.*, 2002; Hill and Prejean, 2005), Mt. St. Helens (Waite *et al.*, 2005), USA; Erebus (Rowe *et al.*, 1998, 2000; Aster *et al.*, 2003), Antarctica; Popocatépetl (Arciniega-Ceballos *et al.*, 1999, 2003; Chouet *et al.*, 2005), Mexico; Karymsky, Russia; and Sangay (Johnson and Lees, 2000), Ecuador. A variety of long-period volcanic seismic signals have been observed at different volcanoes.

Besides the advantages already mentioned, the long-period seismic wavefield observed with a broadband seismic network may be used to directly monitor the activity of a volcano. Dawson *et al.* (2004) implemented the radial semblance method of Kawakatsu *et al.* (2000; originally called 'waveform semblance') for the near real-time monitoring of the activities at the shallow magmatic conduit in Kilauea Volcano, Hawaii (**Figure 18**). Similarly, the moment tensor and single-force source real-time monitoring should be possible using a grid-based method as suggested by Kawakatsu (1998) and implemented at the Earthquake Research Institute of the University of Tokyo for the real-time seismicity monitoring, and as recently realized at Stromboli Volcano (Auger *et al.*, 2006).

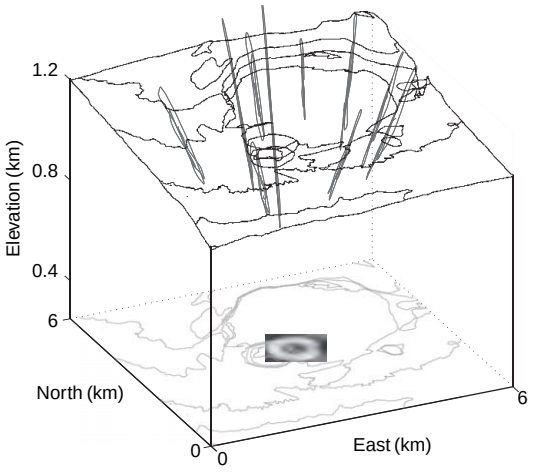


Figure 18 Particle motions recorded by the network during a 1 min long window of long-period signal. The area with bright colors with high semblance values indicates the source region. Adapted from Dawson P, Whilldin D, and Chouet B (2004) Application of near real-time radial semblance to locate the shallow magmatic conduit at Kilauea Volcano, Hawaii. *Geophysical Research Letters* 31: L21606.

4.13.5.2 Array Analysis

To locate the source of volcanic signals, various kinds of methods have been applied to the analyses of observed data. Classical hypocenter determination methods using phase arrivals have been applied to signals observed around active volcanoes and used to predict the eruptions (e.g., Scarpa and Tilling, 1996). More advanced methods such as relative hypocenter determination using the cross-spectral and the double difference have also been applied to volcanic earthquakes and have revealed vivid images of volcanic activities like the transportation of magma beneath active volcanoes (e.g., Hayashi and Morita, 2003).

Nevertheless, there still exist some specific issues to be explored in the analyses of volcanic signals due to complex source processes and complicated wave propagations in inhomogeneous structures. Lack of clear phases and relatively low dominant frequencies of, for example, low-frequency events and volcanic tremors make the direct application of the conventional hypocenter determination means rather difficult.

In the past decade, these difficulties have been partly overcome by the use of seismic arrays. In array observations, seismometers are placed at short distance intervals (a few tens to hundreds of meters), and thus

the wave properties of the incident waves can be examined by extracting coherent portion of the signals using the array data. Recent advance in portable instrumentation enables us to observe volcanic signals close to their sources with dense arrays and sheds new light on the understanding of the processes through the precise determination of location and distribution of volcano seismic sources. For example, the MUSIC (MUltiple SIgnal Classification) technique (Schmidt, 1986; Goldstein and Archuleta, 1991) has been applied to the analyses of the array data observed at Kilauea, Hawaii (e.g., Goldstein and Chouet, 1994; Almendros *et al.*, 2002a, 2002b; **Figure 19**), and at Stromboli, Italy (e.g., Chouet *et al.*, 1997), and revealed magmatic and hydrothermal activities beneath the volcanoes. Other array analysis techniques have been also applied to constrain the source locations and properties of volcanic signals (e.g., Gordeev *et al.*, 1990; Ibanez *et al.*, 2000).

Seismic arrays have also been used to constrain the shallow structure of volcanoes that is indispensable for the quantitative volcano seismology. Here, the wavefield observed by a seismic array is treated as a stationary stochastic one in time and space, and the statistical correlation method originally proposed by Aki (1957) is applied to the data to determine the wave properties such as wave types and phase velocities (e.g., Ferrazzini *et al.*, 1991; Metaxian and Lasage, 1997; De Luca *et al.*, 1997; Chouet *et al.*, 1998; Saccorotti *et al.*, 2001a).

4.13.6 Models for Volcanic Seismic Signals

Although the basic governing physics, some of which have been discussed above, are common, and the methods of data acquisition and analyses employed

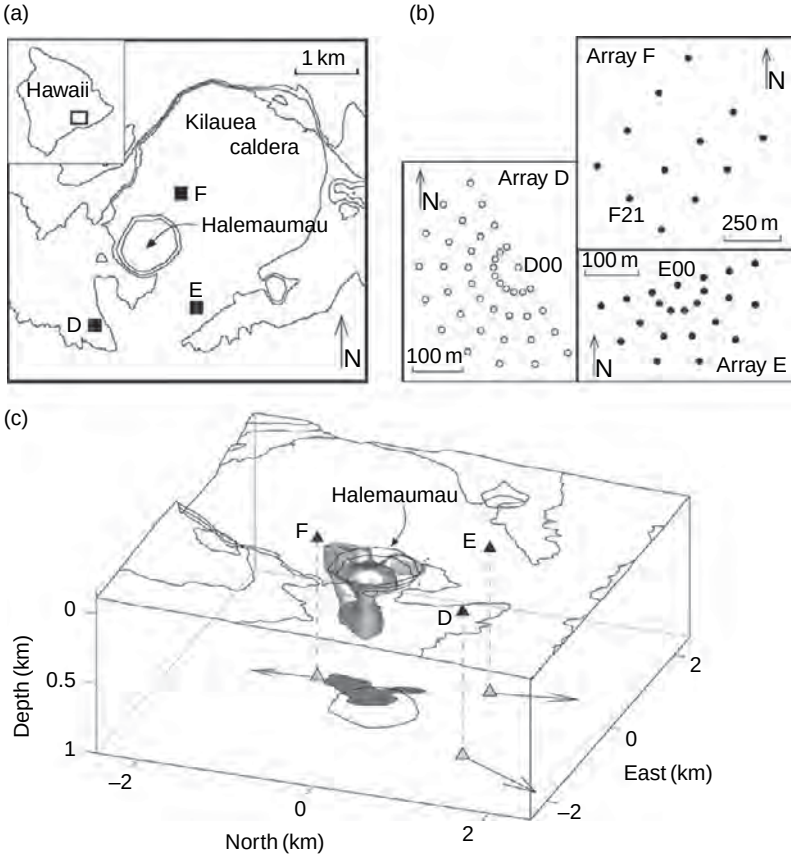


Figure 19 Example of array observation. (a) and (b) Configuration of the array observation at Kilauea, Hawaii. (c) Source location of the seismic signals obtained by array analyses. The arrows represent the apparent slowness vectors determined by frequency-slowness analyses, and the gray region is the estimated source region. Adapted from Almendros J, Chouet B, Dawson P, and Huber C (2002b) Mapping the sources of the seismic wavefield at Kilauea Volcano, Hawaii, using data recorded on multiple seismic antennas. *Bulletin of the Seismological Society of America* 92: 2333–2351.

may be the same, actual phenomena that we observed at volcanoes are different from one volcano to another. To appreciate the diversity, some of the well-studied volcanoes are reviewed. Since the names of volcanic seismic events sometimes differ from one volcano to another due to the difference in conventions as described in 'Terminology', in this section, we keep common terminology locally used at each volcano, and not necessarily follow the conventional classification of periods in earthquake seismology.

4.13.6.1 Aso

Aso Volcano is one of the most active volcanoes in Japan. It has erupted in Strombolian style repeatedly with intervals of 5–10 years, and recent activities take place at the youngest central cone which is composed of seven craters aligned northwest–southeast direction with a length of 1 km. The current eruptive activities have occurred at the first crater located at the northernmost of the chain of the craters.

Among the volcanic tremors observed by Sassa (1935), the most remarkable one is the volcanic microtremor of the second kind for its extraordinary long-period (7.5 s) and its highly repetitive occurrence. Observations using modern broadband seismometers revealed that the fundamental period of the long period tremor (hereafter LPT) is 15 s (Kaneshima *et al.*, 1996; Kawakatsu *et al.*, 2000), and that the second-kind tremor of Sassa is the higher mode of LPTs. The characteristics of LPTs are summarized as follows: (1) continually emitted from the volcano regardless of surface activity; (2) spectra show several common spectral peaks which align with almost equal spacing (15, 7.5, 5, 4 s); (3) decay fairly fast and the duration is only a few cycles; and (4) often accompanied with short-period tremors.

Kaneshima *et al.* (1996) and Kawakatsu *et al.* (2000) deployed a broadband seismic network in 1994 and analyzed the waveforms of the LPTs and those associated with phreatic eruptions which ejected composite of mud and water (**Figure 6**). They concluded that the sources of both phenomena are located a few hundred meters southwest of the first crater and at a depth of 1–1.5 km below ground surface. They further performed a source mechanism inversion of LPTs, and obtained a volumetric source mechanism which can be decomposed into an isotropic and a vertical tensile crack components (Legrand *et al.*, 2000). The detail of the tensile crack component was further studied by Yamamoto *et al.* (1999) using the data obtained by a dense broadband observation. Analyzing the spatial distribution of the signal amplitudes of LPTs (**Figure 20**), they revealed the existence of a crack-like conduit whose strike and width are almost the same as those of the chain of craters; the crack extends nearly vertically from a depth of 300–400 m below the surface to a depth of about 2.5 km, as illustrated in **Figure 21**. This observation revealed that the chain of craters is simply a surface expression of a buried crack-like conduit.

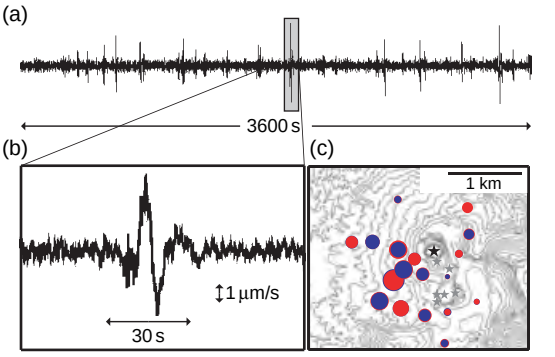


Figure 20 (a) A vertical-component broadband velocity seismogram observed at a distance of 1.5 km. Spike-like signals are LPTs. (b) Close-up view of an LPT. (c) Amplitude variation of LPTs.

The physical properties of the crack-like conduit were studied by Yamamoto (2005), who applied a boundary integral method to the oscillation of a fluid-filled crack with slits at the edges, which allow the fluid to escape, as an extension of Chouet's closed crack model (Chouet, 1986). He demonstrated that the spectral characteristics of LPTs (mode frequencies, spacing, and attenuation) can be explained by the oscillation of a 25 m thick crack-like conduit filled with gas–ash mixture. The existence of a crack-like conduit is also supported by other studies such as a reflection study by Tsutsui and Sudo (2004), and the crack-like conduit is considered as a subsurface path connecting a postulated magma chamber at a depth of around 5 km (Sudo and Kong, 2001) and the surface craters. The nature of other short-period seismic signals with a period of about 0.5 s and a period of around 0.4–0.1 s has been also studied with modern digital data obtained by dense observations using short-period seismometers (Yamamoto, 2005; Takagi *et al.*, 2006).

Figure 21 schematically summarizes the system beneath Aso Volcano that has been revealed by seismological analyses. Such a line of volcanic conduit

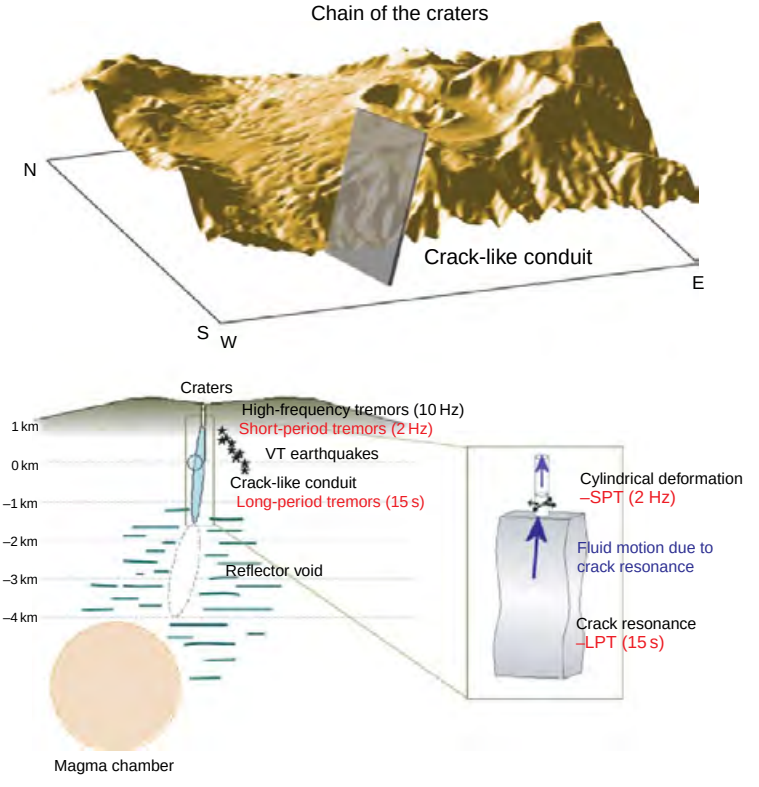


Figure 21 Image of crack-like conduit model. The crack is almost parallel to the chain of craters and the extension of the crack meets the active fumarole at the surface.

systems connecting the magma chamber and the ground surface has not been detected at any other active volcano in the world. Considering that the various volcanic signals are manifestations of dynamic interactions between volcanic fluids and volcanic edifice in the conduit system, Aso Volcano appears to be one of the best fields to study the dynamic behavior of volcanic fluid system beneath active volcanoes.

4.13.6.2 Kilauea

The Hawaiian volcanoes have been produced by the hot spot which is presently beneath the Hawaii island. The volcanoes primarily erupt basaltic magma, and the relatively low viscosity of the magma yields gentle slope of the edifices. Kilauea is the youngest and southeasternmost volcano on the Hawaii island, and it is located at the splitting point between the South-West Rift Zone and the East Rift Zone. Owing to the high level of activities and easy access to it, Kilauea may be the best-studied volcano in the world. A magma-rising system from mantle through narrow pipe-like conduit located beneath the volcano, for instance, was detected by the analyses of spatial distribution of VT earthquakes (e.g., Klein and Koyanagi, 1989), and gives a rough image of the deep magma transport and storage system beneath the volcano (**Figure 22**).

The magma transport system at shallower depths has been also studied using broadband seismic networks and short-period seismic arrays. Performing the moment tensor inversions of broadband data, Ohminato *et al.* (1998) revealed that the source of the very-long-period (VLP; 5–10 s) signals associated with a magmatic activity in 1996 is a crack-like magma pathway at a depth of 1 km, which acts like a buffer for surging magma from beneath summit of Kilauea to the East Rift Zone. Their analyses suggest that ~1000–4000 m^3 of magma is injected into the crack during a slow (1–3 min) accumulation phase, and it is ejected from the crack toward the East Rift Zone during a rapid (5–10 s) deflation phase, which together produce a sawtooth-like displacement signal shown in **Figure 23**. Dawson *et al.* (1998) also analyzed VLP signals observed in 1997, and concluded that the source location is almost

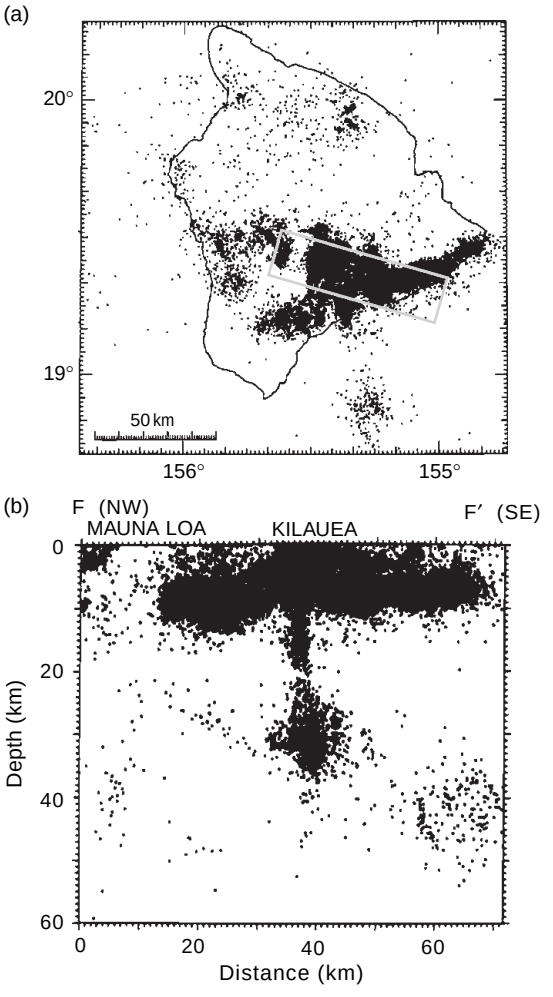


Figure 22 Seismicity of Hawaii. (a) Hypocenters of shallow earthquakes during the period 1970–84. (b) Cross section through Kilauea and Mauna Loa for earthquakes during 1970–83. The area of the cross section is that delimited by the box in the top panel. Adapted from Klein FW and Koyanagi RY (1989) The seismicity and tectonics of Hawaii. In: *The Geology of North America*, Vol. N, The Eastern Pacific Ocean and Hawaii, pp. 238–252. Geological Society of America.

the same as that determined by Ohminato *et al.* (1998), although the source mechanism is different.

The source of long-period (LP; 1–2 s) signals, on the other hand, is determined using seismic array data (e.g., Saccorotti *et al.*, 2001b; Almendros *et al.*, 2001), and the hypocenter of LP signals is located directly above the source of VLP signals. Kumagai *et al.* (2005) analyzed the source mechanism of LP signals, and imaged an expansion and contraction of a nearly horizontal crack through the moment tensor inversions. Based on the fluid-filled crack model developed by Chouet (1986), they further suggest that LP signals represent a resonance of a hydrothermal crack containing bubbly water and/or steam, and that the heat from the underlying magma conduit (VLP source) may cause the pressurization of hydrothermal fluid in the crack and trigger these LP signals. Other geophysical observations, such as a positive correlation between summit SO_2 emissions and the shallow seismicity which was pointed out by Sutton *et al.* (2001), also suggest the interaction between magmatic and hydrothermal systems at the shallow part of the volcano.

These results indicate that the detailed quantitative analyses of volcanic seismic signals are critically important for a better understanding of volcanic processes, and such analyses may be crucial steps toward eruption prediction and the assessment of volcanic hazards.

4.13.6.3 Miyake

Miyake-jima is a basaltic stratovolcanic island in the Izu-Bonin arc, Japan, and has erupted quasiperiodically with an approximate interval of 20 years; the recent activities took place in 1940, 1962, and 1983. The eruption in 2000 was expected in this sense; however, what actually happened was totally unexpected: a new caldera was formed for the first time in the past 2500 years (**Figure 24**). This caldera-forming process was recorded with contemporary geophysical equipments.

Figure 25 shows the seismicity observed during the 2000 Miyake-jima activity. The earthquake swarm started right beneath the volcano, and migrated to the southwest ((Uhira *et al.*, 2005), not well-resolved in **Figure 25**), which may correspond to the subsurface migration of magma beneath the volcano. The swarm activity next shifted to northwest of the island, indicating a major subsurface magma migration from the volcano to the outside dike system. After a small summit eruption, the caldera formation took place, and a number of low-frequency earthquakes are observed beneath the summit area. It appears that sucking of magma from the magma storage system beneath the volcano to the northwest dike system, and the resulting magma vacancy in the volcano, are the direct causes of the caldera formation (Furuya *et al.*, 2003).

During roughly 40 days of the caldera formation, a number of peculiar bell-shaped long-period (50 s) seismic (velocity) signals, as well as steps in tiltmeters, are observed once or twice a day (**Figure 26**). Based on

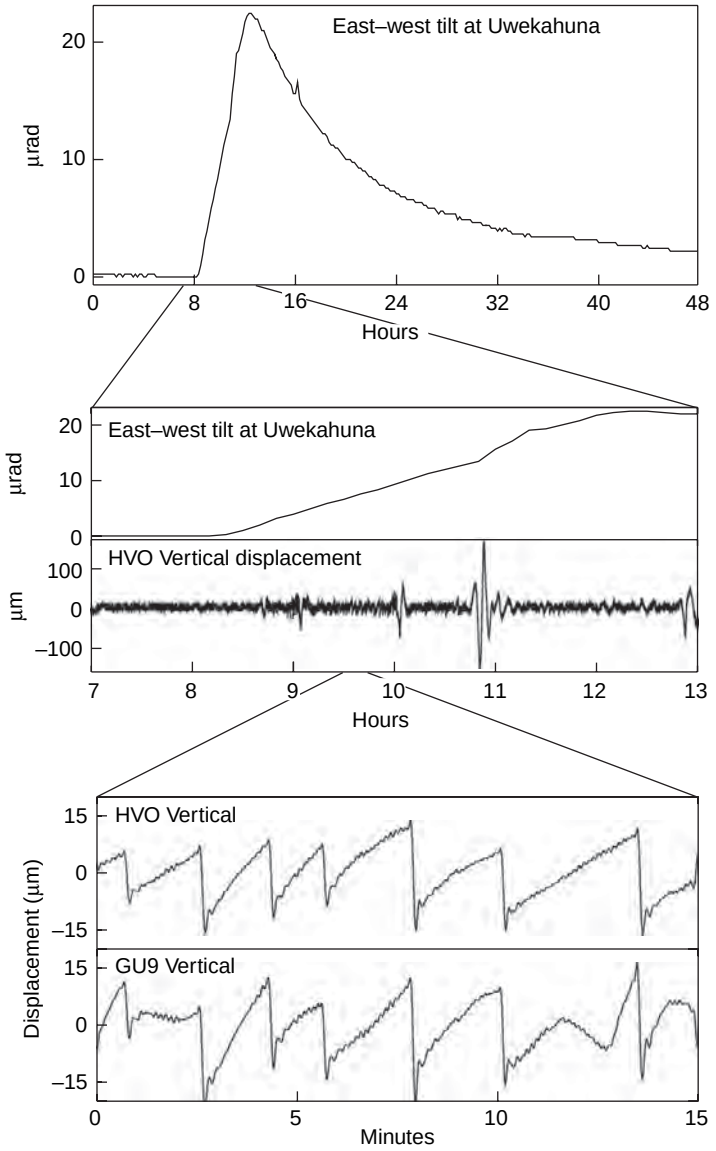


Figure 23 Tilt and broadband record observed around the summit caldera of Kilauea. Adapted from Ohminato T, Chouet B, Dawson P, and Kedar S (1998) Waveform inversion of very long period impulsive signals associated with magmatic injection beneath Kilauea Volcano, Hawaii. *Journal of Geophysical Research* 103: 23839–23862.

the moment tensor inversion of the near-field and far-field broadband seismic records, Kumagai *et al.*, (2001) associated these long-period velocity pulses with inflations of a nearly vertical crack-like magma chamber due to its pressure increase as response to the repetitive stick-slip-type falling of a piston-like mass within the conduit, that results in the caldera formation at the summit. Analyzing the signals of the same sequence of events recorded by the tiltmeters and broadband seismometers deployed in the island, Fujita *et al.* (2004) arrived at a different conclusion. They suggest that these events are caused by a cyclic expansion of a subsurface sill-like magma plumbing system, and introduce a model based on a two-phase flow instability, called a pressure drop oscillation as the mechanism for the repeated activity. Either proposed model seems to explain a part of the observations, but not all. Although the available data may not be able to eventually resolve the difference, nevertheless the 2000 Miyake-jima activity has provided us a rare opportunity to observe a caldera-forming event with contemporary geophysical instruments.

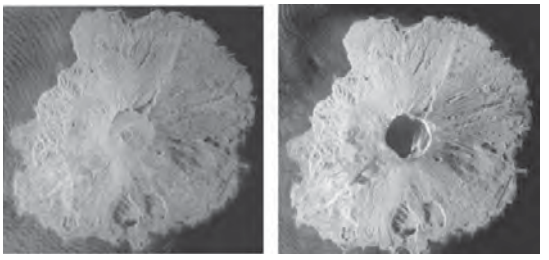


Figure 24 Airborne SAR images of Miyake-jima Volcano before and after the 2000 caldera formation. Adapted from Kumagai H, Ohminato T, and Nakano M, *et al*. (2001) Very-long-period seismic signals and caldera formation at Miyake island, Japan. *Science* 293: 687–690.

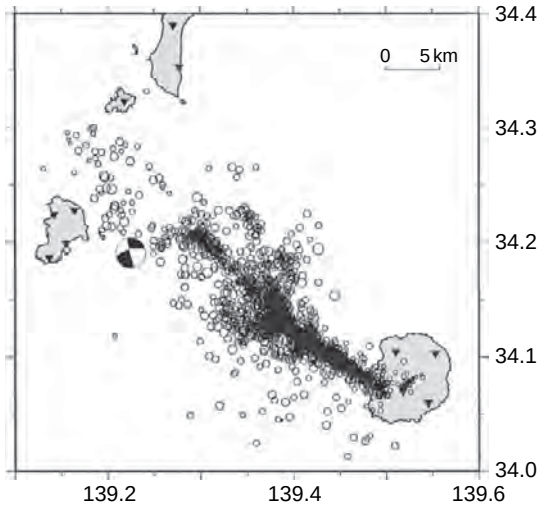


Figure 25 The seismicity of the 2000 Miyake-jima activity. Adapted from Sakai S, Yamada T, and Ide S, *et al*. (2001) Magma migration from the point of view of seismic activity in the volcanism of Miyake-jima island in 2000. *Journal of Geography* 110: 145–155.

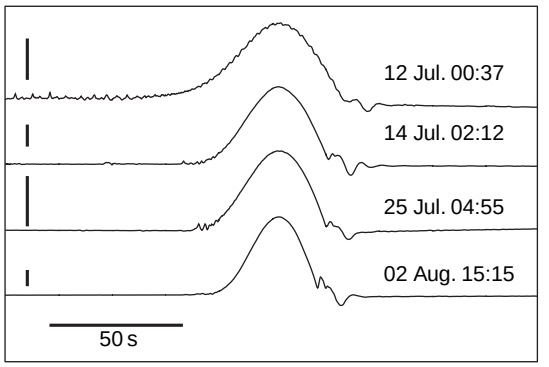


Figure 26 Vertical-component near-field velocity waveforms observed in the Miyake-jima island. Adapted from Kumagai H, Ohminato T, and Nakano M, *et al*. (2001) Very-long-period seismic signals and caldera formation at Miyake island, Japan. *Science* 293: 687–690.

4.13.6.4 Stromboli

Stromboli is one of the Aeolian islands of Italy, and has been almost continuously in eruption for more than at least 2000 years. In contrast to the widely used term 'Strombolian eruption' that is undiscerningly used to describe a variety of volcanic eruptions, the activity of Stromboli is characterized by short-term explosive bursts of lava ejected into the air. Since they eject relatively viscous basaltic lava, building up of the gas pressure is required to fragment the magma, and results in intermittent episodic explosions.

From the seismic point of view, such activities generate a considerable number of explosion quakes, persistent volcanic tremors, and long-period signals. One of the most characteristic features of seismic activities at Stromboli is the extremely wide frequency contents of these signals (**Figure 27**). Spindle-shaped signals superimposed on background continuous tremor are generated by summit eruptions, and the nature of these short-period signals has been extensively studied by many researchers (e.g., Del Pezzo *et al*., 1974; Del Pezzo, 1992; Ripepe and Gordeev, 1999; Saccorotti and Del Pezzo, 2000). These studies shed new light on the seismic activity of the volcano, which is mainly characterized by a very shallow seismicity. Such very shallow origin of the seismic signals is also confirmed by the analyses of seismic array observations (e.g., Chouet *et al*., 1997; Saccorotti and Del Pezzo, 2000), and the sources of explosion quakes and tremors are located at depths shallower than 200 m beneath the summit crater.

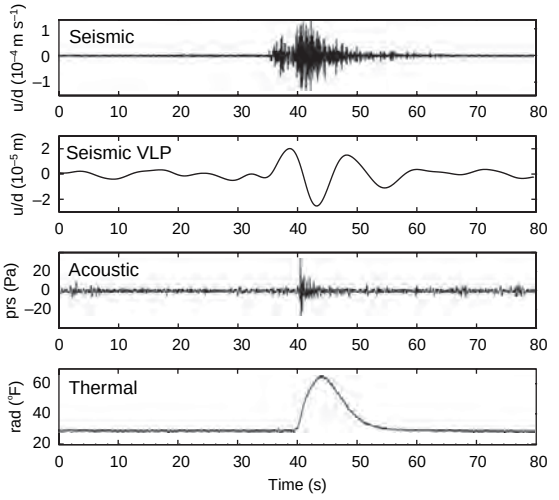


Figure 27 Example of the synchronous occurrence of seismic, acoustic, and thermal signals observed in Stromboli. From Ripepe *et al*., unpublished data.

Recent observations using broadband seismometers reveal the existence of VLP signals with a period of about 10 s which are accompanied with explosions, and give additional constraints on the understanding of Strombolian activities (e.g., Neuberg *et al.*, 1994; Chouet *et al.*, 2003). Analyses of broadband seismograms using the independent component analysis method indicate that further longer-period signals (30–40 s) are generated at almost the same region as the VLP signals (De Martino *et al.*, 2005).

Based on various field observations and laboratory experiments, Ripepe and Gordeev (1999) and Ripepe *et al.* (2001) arrived at a model of Strombolian explosions that is a reminiscent of the analog experimental model by Jaupart and Vergniolle (1989), in which dissolved gas bubbles are trapped at the roof of a magma reservoir as a foam layer and the periodical collapse of the layer causes ascent of gas slugs through a vertical conduit connecting the reservoir and the surface vent. The growth, flow, and burst of the coalescent gas bubbles observed in their laboratory experiments well capture the characteristics of sequential occurrence of very-long-period/short-period seismic signals, acoustic signals, and thermal/light emissions observed at Stromboli. **Figure 28** shows such a direct link between gas flux, magma volume flux, and seismicity, where the balance between gas/magma flux and gas overpressurization appears to determine the type of activities (i.e., effusive or explosive). The common trend in the VLP event rate and the SO_2 emission rate during the effusive phase suggests that the rate of gas flux controls the frequency of foam coalescence and the development of gas slugs which generate VLP signals; the decrease in SO_2 and VLP rates coincides with the increase in the explosion rate at the summit measured by the count of thermal transients and in the amplitude of acoustic pressure; this may be understood in terms of the reduction of gas/magma supplies that results in sealing of eruptive fractures and the upward migration of magma in the conduit.

Performing detailed analyses of the broadband network data, Chouet *et al.* (2003) concluded that the source mechanism of VLP events associated with Strombolian explosions consists of both moment tensor and single force components, which respectively correspond to the opening/closing of an inclined crack just beneath the active vents and the magma flow in the conduit caused by the piston-like rise of a gas slug. To investigate the behavior of coalescence and ascent of gas slugs in such an inclined conduit, James *et al.* (2004) carried out laboratory experiments of the two-phase flow in vertical and inclined pipes, and demonstrated that the inclination of the conduit may play an important role in controlling the flow types and size/velocity distributions of gas slugs in the fluid-filled conduit.

In spite of the difference between models in details, these results suggest that understanding of the bubble dynamics may be essential to elucidate the source process of the seismic signals and the dynamics of Strombolian activities.

4.13.7 Other Volcano-Specific Issues

4.13.7.1 Explosion Quakes

Explosion quakes are observed when explosive eruptions occur (e.g., **Figures 3** and **4**). An explosive volcanic eruption (vulcanian type) can be macroscopically understood as the deflation process of a reservoir system beneath a volcano which may be observed by geodetic (e.g., Ishihara, 1990) or broadband seismic (e.g., **Figure 6**) means. The source mechanism of the accompanied explosion quake may be represented by a combination of a single force and an implosive moment tensor (**Figure 11**). Since the pioneering work of Kanamori *et al.* (1984), there are now a number of observations to quantify the force mechanism and size of volcanic explosions (e.g., Nishimura *et al.*, 1995; Chouet *et al.*, 2005; Ohminato *et al.*, 2006).

Explosion quakes are extensively studied at Sakurajima Volcano, Japan; comparing seismic and video records, Ishihara (1985) demonstrated that explosion quakes occur at a depth of 1–2 km beneath the active crater about a few seconds prior to the explosive eruptions at the summit which generate air shock waves. Uhira and Takeo (1994) showed that an explosive and implosive cylindrical moment tensor component at the source depth is the dominating source mechanism for the explosion quakes. By carefully analyzing borehole seismic network records, Tameguri *et al.* (2002) were able to decompose explosion quakes into two processes: the first one is an isotropic expansion followed by a contraction of a cylinder at a depth of 2 km, and the second is an isotropic expansion and subsequent horizontal contraction at depths of 0.25–0.5 km beneath the crater bottom, which occur about 1 s after the onset of the first and coincide with the generation of the shock wave. The plausibility of this second process was numerically demonstrated by Nishimura and Chouet (2003), who

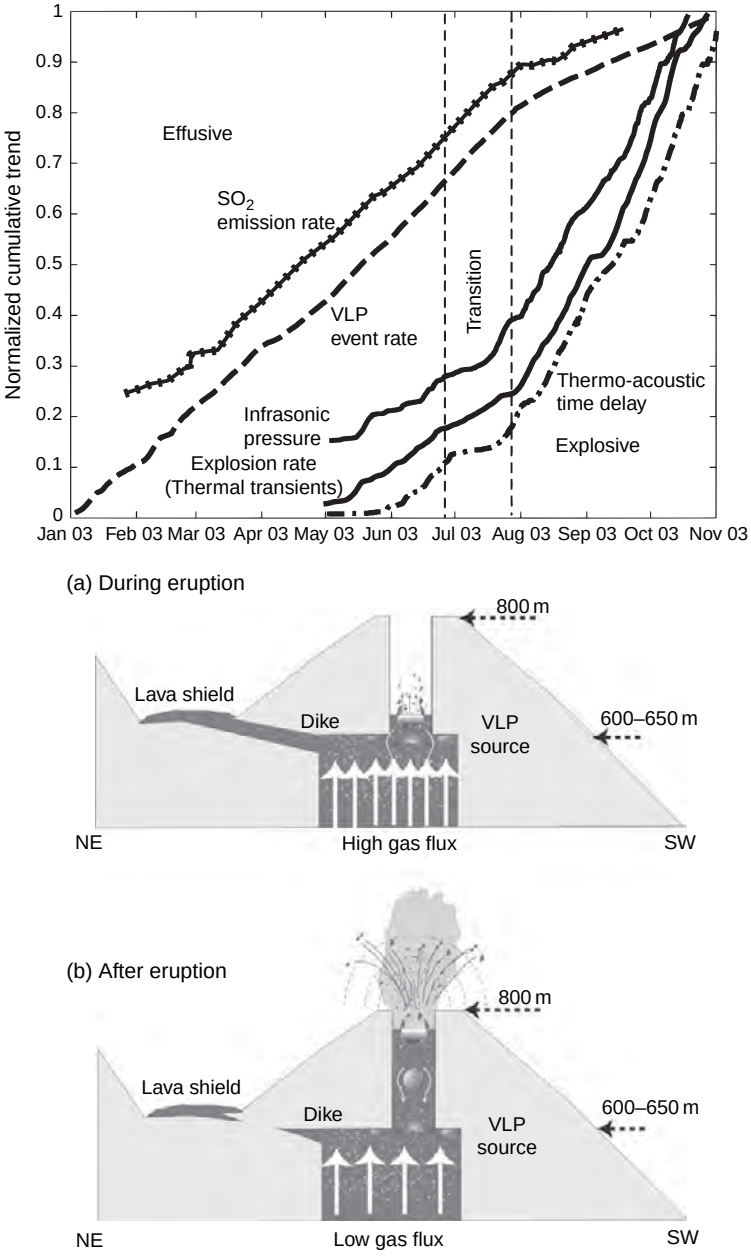


Figure 28 Example of the relation between gas flux rate and seismic/acoustic activities. During the lava effusion phase (model A), few explosions at the summit were observed while the VLP seismicity well correlated with the SO_2 emission rate. On the other hand, in the explosive phase (model B), magma rose up in the conduit due to the sealing of eruptive fractures, resulting in the gas overpressurization and the explosive activity at the summit craters. Adapted from Ripepe M, Marchetti E, and Ulivieri G, *et al.* (2005) Effusive to explosive transition during the 2003 eruption of Stromboli Volcano. *Geology* 33: 341–344.

considered the behavior of a reservoir–conduit system embedded in a homogeneous crust which is filled with a compressive fluid; it was shown that an abrupt pressure increase occurs just beneath the vent which can generate pulse-like Rayleigh waves as observed by Tameguri *et al.* (2002), and that the strength of the lid played an important role in defining eruption types.

Scaling law of volcanic explosions. Nishimura (1998) showed that the far-field displacement due to an explosion is proportional to the area (square of radius) of the crater vent, and obtained a scaling law of volcanic explosion quakes, as similar to that of earthquakes. **Figure 29** compares the vent radii and estimated seismic magnitudes of the largest event for

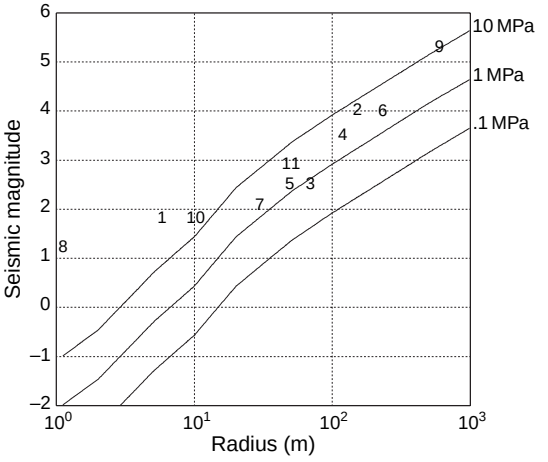


Figure 29 Relation between the seismic magnitude of eruption quakes and the observed vent radius. Numbers from 1 to 11 indicate representative points of different volcanos. Lines show theoretical relation for assumed initial excess pressures. Adapted from Nishimura T (1998) Source mechanisms of volcanic explosion earthquakes: Single force and implosive sources. *Journal of Volcanology and Geothermal Research* 86: 97–106.

studied volcanoes by various investigators; it shows that the magnitude is essentially proportional to the cross-sectional area of the vent for vent radii ranging from 10 to 600 m, as predicted by the theory, and that the initial excess pressure in the reservoir is of the order of a few magapascals, a similar value as the stress drops of earthquakes. This suggests that the size of explosion quakes may be estimated by knowing the vent radius, although it does not necessarily imply that it is possible to predict the size of future eruptions, as it is not guaranteed that the size of present vent is the same for the future ones.

4.13.7.2 Triggered Seismicity

It has been well known among local observers that volcanoes are 'sensitive', as seismicity there often shows changes apparently associated with external disturbances (tides or large earthquakes). For example, Sassa (1936) documented diurnal and semi-diurnal periodicities in the occurrence of volcanic seismicity of Aso Volcano. McNutt and Beavan (1981) also suggested a tidal correlation of seismicity at Pavlof Volcano (cf. Emter, 1997).

There are also convincing lines of evidence that the passage of large amplitude seismic waves dynamically trigger seismicity, especially at volcanoes. Hill *et al.* (1993) documented the seismicity increase in Long Valley caldera, USA, following the $M_w = 7.3$ 1992 Landers earthquake, California (also Gomberg *et al.*, 2001; Prejean *et al.*, 2004). Brodsky and Prejean (2005) made a systematic survey of remotely triggered seismicity at Long Valley; showing the evidence for a frequency-dependent triggering threshold, they concluded that long-period (>30 s) waves are more effective at generating local seismicity than short-period waves of comparable magnitude; among the several proposed mechanisms for such remote dynamic triggering, they suggested that the mechanism that invokes fluid flow through a porous medium can produce such a frequency dependence of triggering threshold (~5 kPa). This mechanism also seems to explain why volcanoes are sensitive to external disturbances, as unclogging fractures to drive fluid flow (Brodsky *et al.*, 2003) may be easily initiated in volcanic/geothermal areas where the constant precipitation from volcanic fluids generate fragile blockages in fractures and faults. West *et al.* (2005) observed periodically (20–30 s) triggered seismicity at Mt. Wrangell, Alaska, when long-period Rayleigh waves generated by the $M_w = 9.0$ 2004 Sumatra earthquake swept across Alaska, and concluded that the extensional stresses of ~25 kPa due to the Rayleigh waves triggered shear failure on normal faults, which may be also influenced by the aforementioned fluid pumping mechanism. These observations suggest that the presence of ample fluid in volcanic systems makes them ideal experimental field for elucidating the earthquake-triggering mechanism.

4.13.8 Concluding Remarks

One of the ultimate goals of volcano seismology is to extract quantitative information about the volcanic fluid transport systems beneath active volcanoes and to utilize such information for forecasting short- and long-term eruptive behaviors. As described in this chapter, macroscopic behaviors of dynamic interactions between volcanic fluids and volcanic edifices have been extensively studied in volcano seismology, and significant progress has been made; the observational/numerical/experimental studies have established firm foundations for explaining the macroscopic processes responsible for generating observed volcanic seismic signals and other related volcanic activities. Considering the rapid progress in technologies for the field observations and numerical simulations, the future of these lines of studies seems to be promising; the accumulation of high-quality digital data at different

volcanoes around the world will further bring significant advances in volcano seismology, as well as volcanology in general; the increase in computer power remarkably extends our capability to simulate complicated dynamic interactions between phases, and enables us to identify and separate distinct signatures of volcanic seismic signals; the recent advance in high-sampling, high-precision GPS and InSAR observations may also help our understanding from the observational aspect.

Nevertheless, our understanding of the elementary physics of multiphase systems, such as the dynamics of bubbles in magmatic fluids and their flows, is still immature, although such understanding fused with the established studies are indispensable for the elucidation of the whole volcanic fluid systems. For instance, recently James *et al.* (2006) experimentally demonstrated that the geometry of the conduit is the key parameter governing the flow pattern of ascending slugs in viscous magma which act as a trigger/source of volcanic seismic signals. Such a parameter of the conduit system may be retrieved by acoustic measurements as demonstrated by Garcés *et al.* (2000), and the integration of these studies will provide a great feedback to volcano seismology. Such studies on the modeling of physical processes, in cooperation with multiparameter observations and numerical simulations, may bring a new boost to volcano seismology in the near future. Volcano seismology, thus, remains to be one of the most challenging fields in Earth science as an integrated science of multidisciplinary researches on multiscale/multiphase systems.

References

Aki K (1957) Space and time spectra of stationary stochastic waves with special reference to microtremors. *Bulletin of the Earthquake Research Institute, University of Tokyo* 35: 415–456.

Aki K (1984) Evidence for magma intrusion during the Mommoth Lakes earthquakes of May 1980 and implications of the absence of volcanic (harmonic) tremor. *Journal of Geophysical Research* 89: 7689–7696.

Aki K, Fehler M, and Das S (1977) Source mechanism of volcanic tremor: Fluid-driven crack models and their application to the 1963 Kilauea eruption. *Journal of Volcanology and Geothermal Research* 2: 259–287.

Aki K and Koyanagi R (1981) Deep volcanic tremor and magma ascent mechanism under Kilauea, Hawaii. *Journal of Geophysical Research* 86: 7095–7109.

Aki K and Lee WHK (2003) Appendix 1 – Glossary of Interest to Earthquake and Engineering Seismologists. In: Lee WHK, Kanamori H, Jennings P, and Kisslinger C (eds.) *International Handbook of Earthquake & Engineering Seismology*, Part B, pp. 1793–1856. New York: Academic Press.

Aki K and Richards PG (2002) *Quantitative Seismology,* 2nd edn. Sausalito, CA: University Science Books.

Alidibirov M and Dingwell DB (1996) Magma fragmentation by rapid decompression. *Nature* 380: 146–149.

Almendros J, Chouet B, and Dawson P (2001) Spatial extent of a hydrothermal system at Kilauea Volcano, Hawaii, determined from array analyses of shallow long-period seismicity. 2: Result. *Journal of Geophysical Research* 106: 13581–13598.

Almendros J, Chouet B, Dawson P, and Bond T (2002a) Identifying elements of the plumbing system beneath Kilauea Volcano, Hawaii, from the source locations of very-long-period signals. *Geophysical Journal International* 148: 303–312.

Almendros J, Chouet B, Dawson P, and Huber C (2002b) Mapping the sources of the seismic wavefield at Kilauea volcano, Hawaii, using data recorded on multiple seismic antennas. *Bulletin of the Seismological Society of America* 92: 2333–2351.

Aoyama H and Takeo M (2001) Wave properties and focal mechanisms of N-type earthquakes at Asama volcano. *Journal of Volcanology and Geothermal Research* 105: 163–182.

Arciniega-Ceballos A, Chouet B, and Dawson P (1999) Very long-period signals associated with vulcanian explosions at Popocatepetl volcano. *Geophysical Research Letters* 26: 3013–3016.

Arciniega-Ceballos A, Chouet B, and Dawson P (2003) Long-period events and tremor at Popocatepetl volcano (1994–2000) and their broadband characteristics. *Bulletin of Volcanology* 65: 124–135.

Aster R, Mah S, Kyle P, *et al.* (2003) Very long period oscillations of Mount Erebus Volcano. *Journal of Geophysical Research* 108: 2522.

Auger E, D'Auria L, Martini M, Chouet B, and Dawson P (2006) Real-time monitoring and massive inversion of source parameters of very long period seismic signals: An application to Stromboli volcano. *Geophysical Research Letters* 33: L04301.

Backus G and Mulcahy M (1976) Moment tensors and other phenomenological descriptions of seismic sources. I: Continuous displacements. *Geophysical Journal of the Royal Astronomical Society* 46: 341–361.

Balmforth EJ, Craster RV, and Rust AC (2005) Instability in flow through elastic conduits and volcanic tremor. *Journal of Fluid Mechanics* 527: 353–377.

Ben-Menahem A and Singh SJ (1981) *Seismic Waves and Sources*. Berlin: Springer.

Benoit JP and McNutt SR (1997) New constraints on source processes of volcanic tremor at Arenal Volcano, Costa Rica, using broadband seismic data. *Geophysical Research Letters* 24: 449–452.

Biot MA (1952) Propagation of elastic waves in a cylindrical bore containing a fluid. *Journal of Applied Physics* 23: 997–1005.

Biot MA (1956) The theory of propagation of elastic waves in a fluid-saturated porous solid. I: Low-frequency range. *Journal of the Acoustical Society of America* 28: 168–178.

Brodsky EE and Prejean SG (2005) New constraints on mechanisms of remotely triggered seismicity at Long Valley Caldera. *Journal of Geophysical Research* 110: B04302.

Brodsky EE, Roeloffs E, Woodcock D, Gall I, and Manga M (2003) A mechanism for sustained groundwater pressure changes induced by distant earthquakes. *Journal of Geophysical Research* 108: 2390.

Brodsky EE, Sturtevant B, and Kanamori H (1998) Earthquakes, volcanoes, and rectified diffusion. *Journal of Geophysical Research* 103: 23827–23838.

Chouet B (1985) Excitation of a buried magmatic pipe: A seismic source model for volcanic tremor. *Journal of Geophysical Research* 90: 1881–1893.

Chouet B (1986) Dynamics of a fluid-driven crack in three dimensions by the finite difference method. *Journal of Geophysical Research* 91: 13967–13992.

Chouet B (1992) A seismic model for the source of long-period events and harmonic tremor. In: *Volcanic Seismology*, pp. 133–156. Berlin: Springer.

Chouet B (1996a) Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature* 380: 309–316.

Chouet B (1996b) New methods and future trends in seismological volcano monitoring. In: *Monitoring and Mitigation of Volcano Hazards*, pp. 23–97. New York: Springer.

Chouet B (2003) Volcano seismology. *Pure and Applied Geophysics* 160: 739–788.

Chouet B, Dawson P, and Arciniega-Ceballos A (2005) Source mechanism of vulcanian degassing at Popocatepetl volcano, Mexico, determined from waveform inversions of very long period signals. *Journal of Geophysical Research* 110: B07301.

Chouet B, Dawson P, Ohminato T, *et al.* (2003) Source mechanisms of explosions at Stromboli Volcano, Italy, determined from moment-tensor inversions of very-long-period data. *Journal of Geophysical Research* 108: 2019.

Chouet B and Julian BR (1985) Dynamics of an expanding fluid-filled crack. *Journal of Geophysical Research* 90: 11187–11198.

Chouet B, Luca GD, Milana G, Dawson PB, Martini M, and Scarpa R (1998) Shallow velocity structure of Stromboli Volcano, Italy, derived from small-aperture array measurements of Strombolian tremor. *Bulletin of the Seismological Society of America* 88: 653–666.

Chouet B, Page RA, Stephens CD, Lahr JC, and Power JA (1994) Precursory swarms of long-period events at Redoubt Volcano (1989–1990), Alaska: Their origin and use as a forecasting tool. *Journal of Volcanology and Geothermal Research* 62: 95–135.

Chouet B, Saccorotti G, Martini M, *et al.* (1997) Source and path effects in the wave fields of tremor and explosions at Stromboli Volcano, Italy. *Journal of Geophysical Research* 102: 15129–15150.

Crosson RS and Bame DA (1985) A spherical source model for low frequency volcanic earthquake. *Journal of Geophysical Research* 90: 10237–10247.

Dahm T (1996) Relative moment tensor inversion based on ray theory: Theory and synthetic tests. *Geophysical Journal International* 124: 245–257.

Dawson P, Dietel C, Chouet B, Honma K, Ohminato T, and Okubo P (1998) A digitally telemetered broadband seismic network at Kilauea Volcano, Hawaii. *US Geological Survey, Open File Report* 98–108: 1–121.

Dawson P, Whilldin D, and Chouet B (2004) Application of near real-time radial semblance to locate the shallow magmatic conduit at Kilauea Volcano, Hawaii. *Geophysical Research Letters* 31: L21606.

De Luca G, Scarpa R, Del Pezzo E, and Simini M (1997) Shallow structure of Mt. Vesuvius volcano, Italy, from seismic array analysis. *Geophysical Research Letters* 24: 481–484.

De Martino S, Falanga M, Scarpa R, and Godano C (2005) Very-long-period volcanic tremor at Stromboli, Italy. *Bulletin of the Seismological Society of America* 95: 1186–1192.

Del Pezzo E (1992) Wave polarization and location of the source of the explosion quakes at Stromboli volcano. In: Gasperini P, Scarpa R, and Aki K (eds.) *Volcanic Seismology*, pp. 279–296. Heidelberg: Springer.

Del Pezzo E, Guerra I, Bascio AL, Luongo G, Nappi G, and Scarpa R (1974) Microtremors ad volcanic explosions at Stromboli - Part 2. *Bulletin of Volcanology* 38: 1023–1036.

Dreger DS, Tkali H, and Johnston M (2000) Dilational processes accompanying earthquakes in the Long Valley Caldera. *Science* 288: 122–125.

Dreier R, Widmer R, Schick R, and Zurn W (1994) Stacking of broad-band seismograms of shocks at Stromboli. *Acta Vulcanologica* 5: 165–172.

Dziewonski AM, Chou TA, and Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research* 86: 2825–2852.

Eissler HK and Kanamori H (1987) A single-force model for the 1975 Kalapana, Hawaii, earthquake. *Journal of Geophysical Research* 92: 4827–4836.

Ekström G, Nettles M, and Abers GA (2003) Glacial earthquakes. *Science* 302: 622–624.

Emter D (1997) Tidal triggering of earthquakes and volcanic events. In: Wilhelm H, Zürn W, and Wenzel HG (eds.) *Tidal Phenomena, Lecture Notes in Earth Sciences*, pp. 293–309. Berlin: Springer.

Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proceedings of the Royal Society of London A* 241: 376–396.

Falsaperla S, Langer H, Martinelli B, and Schick R (1994) Seismic measurements on Stromboli volcano in a wide frequency range. *Acta Vulcanologica* 5: 173–178.

Ferrazzini V and Aki K (1987) Slow waves trapped in a fluid-filled infinite crack: Implication for volcanic tremor. *Journal of Geophysical Research* 92: 9215–9223.

Ferrazzini V, Aki K, and Chouet B (1991) Characteristics of seismic waves composing Hawaiian volcanic tremor and gas-piston events observed by a near-source array. *Journal of Geophysical Research* 96: 6199–6209.

Ferrick M, Qamar A, and Lawrence WFS (1982) Source mechanism of volcanic tremor. *Journal of Geophysical Research* 87: 8675–8683.

Foulger GR, Julian BR, Hill DP, Pitt AM, Malin PE, and Shalev E (2004) Non-double-couple microearthquakes at Long Valley caldera, California, provide evidence for hydraulic fracturing. *Journal of Volcanology and Geothermal Research* 132: 45–71.

Fujita E and Ida Y (2003) Geometrical effects and low-attenuation resonance of volcanic fluid inclusions for the source mechanism of long-period earthquakes. *Journal of Geophysical Research* 108: 2118.

Fujita E, Ida Y, and Oikawa J (1995) Eigen oscillation of a fluid sphere and source mechanism of harmonic volcanic tremor. *Journal of Volcanology and Geothermal Research* 69: 365–378.

Fujita E, Ukawa M, and Yamamoto E (2004) Subsurface cyclic magma sill expansions in the 2000 Miyakejima volcano eruption: Possibility of two-phase flow oscillation. *Journal of Geophysical Research* 109: B04205.

Furuya M, Okubo S, Sun W, *et al.* (2003) Spatiotemporal gravity changes at Miyakejima Volcano, Japan: Caldera collapse, explosive eruptions and magma movement. *Journal of Geophysical Research* 108: 2219.

Garcés M, McNutt SR, Hansen RA, and Eichelberger JC (2000) Application of wave-theoretical seismoacoustic models to the interpretation of explosion and eruption tremor signals radiated by Pavlof volcano, Alaska. *Journal of Geophysical Research* 105: 3039–3058.

Gil Cruz F and Chouet B (1997) Long-period events, the most characteristic seismicity accompanying the emplacement and extrusion of a lava dome in Galeras Volcano, Colombia, in 1991. *Journal of Volcanology and Geothermal Research* 77: 121–158.

Gilbert F (1970) Excitation of the normal-modes of the Earth by earthquake sources. *Geophysical Journal of the Royal Astronomical Society* 23: 119–123.

Goldstein P and Archuleta RJ (1991) Deterministic frequency-wavenumber methods and direct measurements of rupture propagation during earthquakes using a dense array: Theory

and methods. *Journal of Geophysical Research* 96: 6173–6185.
Goldstein P and Chouet B (1994) Array measurements and modeling of sources of shallow volcanic tremor at Kilauea Volcano, Hawaii. *Journal of Geophysical Research* 99: 2637–2652.
Gomberg J, Reasenberg PA, Bodin P, and Harris RA (2001) Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature* 411: 462–466.
Gordeev EI, Saltykov VA, Sinitsin VI, and Chebrov VN (1990) Temporal and spatial characteristics of volcanic tremor wave fields. *Journal of Volcanology and Geothermal Research* 40: 89–101.
Hagerty MT, Schwartz SY, Garcés MA, and Protti M (2000) Analysis of seismic and acoustic observations at Arenal Volcano, Costa Rica, 1995–1997. *Journal of Volcanology and Geothermal Research* 101: 27–65.
Hasegawa A and Yamamoto A (1994) Deep, low-frequency microearthquakes in or around seismic low-velocity zones beneath active volcanoes in northeastern Japan. *Tectonophysics* 233: 233–252.
Hasegawa HS and Kanamori H (1987) Source mechanism of the magnitude 7.2 Grand Banks earthquake of November 1929; double couple or submarine landslide? *Bulletin of the Seismological Society of America* 77: 1984–2004.
Hayashi Y and Morita Y (2003) An image of a magma intrusion process inferred from precise hypocentral migrations of the earthquake swarm east of the Izu Peninsula. *Geophysical Journal International* 153: 159–174.
Hellweg P, Seidl D, Brotopuspito KS, and Brustle W (1994) Team investigates activity at Mt. Semeru, Java, volcano. *EOS, Transactions of AGU* 75: 313–317.
Hidayat D, Chouet B, Voight B, Dawson P, and Ratdomopurbo A (2002) Source mechanism of very-long-period signals accompanying dome growth activity at Merapi volcano, Indonesia. *Geophysical Research Letters* 29: 2118.
Hidayat D, Voight B, Langston C, Ratdomopurbo A, and Ebeling C (2000) Broadband seismic experiment at Merapi Volcano, Java, Indonesia: Very-long-period pulses embedded in multiphase earthquakes. *Journal of Volcanology and Geothermal Research* 100: 215–231.
Hill DP, Dawson P, Johnston MJS, Pitt AM, Biasi G, and Smith K (2002) Very-long-period volcanic earthquakes beneath Mammoth Mountain, California. *Geophysical Research Letters* 29: 1370.
Hill DP and Prejean S (2005) Magmatic unrest beneath Mammoth Mountain, California. *Journal of Volcanology and Geothermal Research* 146: 257–283.
Hill DP, Reasenberg PA, Michael A, *et al.* (1993) Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* 260: 1617–1623.
Hudson JA, Pearce RG, and Rogers RM (1989) Source type plot for inversion of the moment tensor. *Journal of Geophysical Research* 94: 765–774.
Ibanez JM, Del Pezzo E, Almendros J, *et al.* (2000) Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling. *Journal of Geophysical Research* 105: 13905–13932.
Ichihara M and Brodsky EE (2006) A limit on the effect of rectified diffusion in volcanic systems. *Geophysical Research Letters*. 33: L02316.
Ichihara M, Rittel D, and Sturtevant B (2002) Fragmentation of a porous viscoelastic material: Implications to magma fragmentation. *Journal of Geophysical Research* 107: 2229.
Ida Y (1996) Cyclic fluid effusion accompanied by pressure change: Implication for volcanic eruptions and tremor. *Geophysical Research Letters* 23: 1457–1460.
Ishihara K (1985) Dynamical analysis of volcanic explosion. *Journal of Geodynamics* 3: 327–349.
Ishihara K (1990) Pressure sources and induced ground deformation associated with explosive eruptions of an andesitic volcano, Sakurajima volcano, Japan. In: Ryan MP (ed.) *Magma Transport and Storage*, pp. 335–356. New York: Wiley.
Iwamura K and Kaneshima S (2005) Numerical simulation of the steam-water flow instability as a mechanism of long-period ground vibrations at geothermal areas. *Geophysical Journal International* 163: 833–851.
James MR, Lane SJ, Chouet B, and Gilbert JS (2004) Pressure changes associated with the ascent and bursting of gas slugs in liquid-filled vertical and inclined conduits. *Journal of Volcanology and Geothermal Research* 129: 61–82.
James MR, Lane SJ, and Chouet BA (2006) Gas slug ascent through changes in conduit diameter: Laboratory insights into a volcano-seismic source process in low-viscosity magmas. *Journal of Geophysical Research* 111: B05201.
Jaupart C and Vergniolle S (1989) The generation and collapse of a foam layer at the roof of a basaltic magma chamber. *Journal of Fluid Mechanics* 203: 347–380.
Johnson JB and Lees JM (2000) Plugs and chugs–seismic and acoustic observations of degassing explosions at Karymsky, Russia and Sangay, Ecuador. *Journal of Volcanology and Geothermal Research* 101: 67–82.
Jousset P, Neuberg J, and Sturton S (2003) Modelling the time-dependent frequency content of low-frequency volcanic earthquakes. *Journal of Volcanology and Geothermal Research* 128: 201–223.
Julian BR (1983) Evidence for dyke intrusion earthquake mechanisms near Long Valley Caldera, California. *Nature* 303: 323–325.
Julian BR (1994) Volcanic tremor: Nonlinear excitation by fluid flow. *Journal of Geophysical Research* 99: 11859–11878.
Julian BR and Foulger GR (1996) Earthquake mechanisms from linear-programming inversion of seismic-wave amplitude ratios. *Bulletin of the Seismological Society of America* 86: 972–980.
Julian BR, Miller AD, and Foulger GR (1997) Non-double-couple earthquake mechanisms at the Hengill-Grensdalur volcanic complex, southwest Iceland. *Geophysical Research Letters* 24: 743–746.
Julian BR, Miller AD, and Foulger GR (1998) Non-double-couple earthquakes. 1: Theory. *Review of Geophysics* 36: 525–549.
Kamo K, Furuzawa T, and Akamatsu J (1977) Some natures of volcanic micro-tremors at the Sakurajima volcano. *Bulletin of the Volcanological Society of Japan* 22: 41–58.
Kanamori H, Ekström G, Dziewonski A, Barker JS, and Sipkin SA (1993) Seismic radiation by magma injection: An anomalous seismic event near Tori Shima, Japan. *Journal of Geophysical Research* 98: 6511–6522.
Kanamori H and Given JW (1982) Analysis of long-period seismic waves excited by the May 18, 1980, eruption of Mount St. Helens - A terrestrial monopole. *Journal of Geophysical Research* 87: 5422–5432.
Kanamori H, Given JW, and Lay T (1984) Analysis of seismic body waves excited by the Mount St. Helens eruption of May 18, 1980. *Journal of Geophysical Research* 89: 1856–1866.
Kanamori H and Mori J (1992) Harmonic excitation of mantle Rayleigh waves by the 1991 eruption of Mount Pinatubo Philippines. *Geophysical Research Letters* 19: 721–724.
Kaneshima S, Kawakatsu H, Matsubayashi H, *et al.* (1996) Mechanism of phreatic eruptions at Aso Volcano Inferred from near-field broadband seismic observations. *Science* 273: 643–645.
Kawakatsu H (1989) Centroid single force inversion of seismic waves generated by landslides. *Journal of Geophysical Research* 94: 12363–12374.

Kawakatsu H (1998) On the realtime monitoring of the long-period seismic wavefield. *Bulletin of the Earthquake Research Institute, University of Tokyo* 73: 267–274 (also in Methods and applications of signal processing in seismic network operations (2002), Springer, pp.251–257).

Kawakatsu H, Kaneshima S, Matsubayashi H, *et al.* (2000) Aso94: Aso seismic observation with broadband instruments. *Journal of Volcanology and Geothermal Research* 101: 129–154.

Kawakatsu H, Ohminato T, and Ito H (1994) 10s-period volcanic tremors observed over a wide area in southwestern Japan. *Geophysical Research Letters* 21: 1963–1966.

Kawakatsu H, Ohminato T, Ito H, *et al.* (1992) Broadband seismic observation at the Sakurajima volcano, Japan. *Geophysical Research Letters* 19: 1959–1962.

Kieffer SW (1977) Sound speed in liquid-gas mixtures: Water–air and water–steam. *Journal of Geophysical Research* 82: 2895–2904.

Kikuchi M and Kanamori H (1982) Inversion of complex body waves. *Bulletin of the Seismological Society of America* 72: 491–506.

Klein FW and Koyanagi RY (1989) The seismicity and tectonics of Hawaii. In: *The Geology of North America*, Vol. N, The Eastern Pacific Ocean and Hawaii, pp. 238–252. Geological Society of America.

Knopoff L and Randall M (1970) The compensated linear-vector dipole: A possible mechanism for deep earthquakes. *Journal of Geophysical Research* 75: 4957–4963.

Kubotera A (1974) Volcanic tremors at Aso volcano. In: Civetta L, Gasparini, Luongo G, and Rapolla A (eds.) *Physical Volcanology*, pp. 29–47. Amsterdam: Elsevier.

Kumagai H (2006) Temporal evolution of a magmatic dike system inferred from the complex frequencies of very long period seismic signals. *Journal of Geophysical Research* 111: B06201.

Kumagai H and Chouet B (1999) The complex frequencies of long-period seismic events as probes of fluid composition beneath volcanoes. *Geophysical Journal International* 138: F7–F12.

Kumagai H and Chouet B (2000) Acoustic properties of a crack containing magmatic or hydrothermal fluids. *Journal of Geophysical Research* 105: 25493–25512.

Kumagai H, Chouet B, and Dawson PB (2005) Source process of a long-period event at Kilauea volcano, Hawaii. *Geophysical Journal International* 161: 243–254.

Kumagai H, Chouet B, and Nakano M (2002) Temporal evolution of a hydrothermal system in Kusatsu-Shirane Volcano, Japan, inferred from the complex frequencies of long-period events. *Journal of Geophysical Research* 107: ESE 9–1.

Kumagai H, Ohminato T, and Nakano M, *et al.* (2001) Very-long-period seismic signals and caldera formation at Miyake island, Japan. *Science* 293: 687–690.

Leet RC (1988) Saturated and subcooled hydrothermal boiling in groundwater flow channels as a source of harmonic tremor. *Journal of Geophysical Research* 93: 4835–4849.

Legrand D, Kaneshima S, and Kawakatsu H (2000) Moment tensor analysis of near-field broadband waveforms observed at Aso Volcano, Japan. *Journal of Volcanology and Geothermal Research* 101: 155–169.

McNutt SR (1996) Seismic monitoring and eruption forecasting of volcanoes: A review of the state-of-the-art and case histories. In: Scarpa R and Tilling RI (eds.) *Monitoring and Mitigation of Volcano Hazards*, pp. 99–146. New York: Springer.

McNutt SR (2002) Volcano seismology. In: Lee WHK, Kanamori H, Jennings P, Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology*, Part A, Ch. 25, pp. 383–406. New York: Academic Press.

McNutt SR (2005) Volcanic seismology. *Annual Review of Earth and Planetary Sciences* 33: 461–491.

McNutt SR and Beavan RJ (1981) Volcanic earthquakes at Pavlof Volcano correlated with the solid earth tide. *Nature* 294: 615–518.

Metaxian J-P and Lasage P (1997) Permanent tremor of Masaya Volcano, Nicaragua: Wave field analysis and source location. *Journal of Geophysical Research* 102: 22529–22545.

Miller AD, Foulger GR, and Julian BR (1998) Non-double-couple earthquakes. 2: Observations. *Review of Geophysics* 36: 551–568.

Minakami T (1974) Seismology of volcanoes in Japan. In: Civetta L, Gasparini P, Luongo G, and Rapolla A (eds.) *Physical Volcanology*, pp. 1–27. Amsterdam: Elsevier.

Mogi K (1958) Relation between the eruptions of various volcanoes and the deformations of the ground surfaces around them. *Bulletin of the Earthquake Research Institute* 36: 99–134.

Molina I, Kumagai H, and Yepes H (2004) Resonances of a volcanic conduit triggered by repetitive injections of an ash-laden gas. *Geophysical Research Letters* 31: L03603.

Mori J, Patia H, McKee C, *et al.* (1989) Seismicity associated with eruptive activity at Langila volcano, Papua New Guinea. *Journal of Volcanology and Geothermal Research* 38: 243–255.

Morrissey MM and Chouet B (1997) Burst conditions of explosive volcanic eruptions recorded on microbarographs. *Science* 275: 1290–1293.

Morrissey MM and Chouet B (2001) Trends in long-period seismicity related to magmatic fluid compositions. *Journal of Volcanology and Geothermal Research* 108: 265–281.

Nakamichi H, Hamaguchi H, Tanaka S, Ueki S, Nishimura T, and Hasegawa A (2003) Source mechanisms of deep and intermediate-depth low-frequency earthquakes beneath Iwate volcano, northeastern Japan. *Geophysical Journal International* 154: 811–828.

Nakano M, Kumagai H, and Chouet B (2003) Source mechanism of long-period events at Kusatsu-Shirane Volcano, Japan, inferred from waveform inversion of the effective excitation functions. *Journal of Volcanology and Geothermal Research* 122: 149–164.

Nakano M, Kumagai H, Kumazawa M, Yamaoka K, and Chouet B (1998) The excitation and characteristic frequency of the long-period volcanic event: An approach based on an inhomogeneous autoregressive model of a linear dynamic system. *Journal of Geophysical Research* 103: 10031–10046.

Nettles M and Ekström G (1998) Faulting mechanism of anomalous earthquakes near Bardarbunga volcano, Iceland. *Journal of Geophysical Research* 103: 17973–17984.

Neuberg J, Luckett R, Baptie B, and Olsen K (2000) Models of tremor and low-frequency earthquake swarms on Montserrat. *Journal of Volcanology and Geothermal Research* 101: 83–104.

Neuberg J, Luckett R, Ripepe M, and Braun T (1994) Highlights from a seismic broad-band array on Stromboli volcano. *Geophysical Research Letters* 21: 749–752.

Nishida K, Kobayashi N, and Fukao Y (2000) Resonant oscillations between the solid earth and the atmosphere. *Science* 287: 2244–2246.

Nishimura T (1998) Source mechanisms of volcanic explosion earthquakes: Single force and implosive sources. *Journal of Volcanology and Geothermal Research* 86: 97–106.

Nishimura T and Chouet B (2003) A numerical simulation of magma motion, crustal deformation, and seismic radiation associated with volcanic eruptions. *Geophysical Journal International* 153: 699–718.

Nishimura T, Hamaguchi H, and Ueki S (1995) Source mechanisms of volcanic tremor and low-frequency earthquakes associated with the 1988–89 eruptive activity of Mt. Tokachi, Hokkaido, Japan. *Geophysical Journal International* 121: 444–458.

Nishimura T, Nakamichi H, Tanaka S, *et al.* (2000) Source process of very long period seismic events associated with the 1998 activity of Iwate Volcano, northeastern Japan. *Journal of Geophysical Research* 105: 19135–19147.

Nishimura T, Ueki S, Yamawaki T, *et al.* (2003) Broadband seismic signals associated with the 2000 volcanic unrest of Mount Bandai, northeastern Japan. *Journal of Volcanology and Geothermal Research* 119: 51–59.

Ohminato T (2006) Characteristics and source modeling of broadband seismic signals associated with the hydrothermal system at Satsuma-iwojima volcano, Japan. *Journal of Volcanology and Geothermal Research* 158: 467–490.

Ohminato T and Chouet B (1997) A free-surface boundary condition for including 3D topography in the finite-difference method. *Bulletin of the Seismological Society of America* 87: 494–515.

Ohminato T, Chouet B, Dawson P, and Kedar S (1998) Waveform inversion of very long period impulsive signals associated with magmatic injection beneath Kilauea Volcano, Hawaii. *Journal of Geophysical Research* 103: 23839–23862.

Ohminato T and Ereditato D (1997) Broadband seismic observations at Satsuma-Iwojima volcano, Japan. *Geophysical Research Letters* 24: 2845–2848.

Ohminato T, Takeo M, Kumagai H, *et al.* (2006) Vulcanian eruptions with dominant single force components observed during the Asama 2004 volcanic activity in Japan. *Earth Planets Space* 58: 583–593.

Pitt AM and Hill DP (1994) Long-period earthquakes in the Long Valley caldera region, eastern California. *Geophysical Research Letters* 21: 1679–1682.

Prejean SG, Hill DP, Brodsky EE, *et al.* (2004) Remotely triggered seismicity on the United States west coast following the Mw 7.9 Denali fault earthquake. *Bulletin of the Seismological Society of America* 94: S348–S359.

Ripepe M, Ciliberto S, and Schiava MD (2001) Time constraints for modeling source dynamics of volcanic explosions at Stromboli. *Journal of Geophysical Research* 106: 8713–8727.

Ripepe M and Gordeev E (1999) Gas bubble dynamics model for shallow volcanic tremor at Stromboli. *Journal of Geophysical Research* 104: 10639–10654.

Ripepe M, Marchetti E, Ulivieri G, *et al.* (2005) Effusive to explosive transition during the 2003 eruption of Stromboli Volcano. *Geology* 33: 341–344.

Ripepe M, Poggi P, Braun T, and Gordeev E (1996) Infrasonic waves and volcanic tremor at Stromboli. *Geophysical Research Letters* 23: 181–184.

Ross A, Foulger GR, and Julian BR (1996) Non-double-couple earthquake mechanisms at The Geysers geothermal area, California. *Geophysical Research Letters* 23: 877–880.

Rowe CA, Aster RC, Kyle PR, Dibble RR, and Schlue JW (2000) Seismic and acoustic observations at Mount Erebus Volcano, Ross Island, Antarctica, 1994–1998. *Journal of Volcanology and Geothermal Research* 101: 105–128.

Rowe CA, Aster RC, Kyle PR, Schlue JW, and Dibble RR (1998) Broadband recording of Strombolian explosions and associated very-long-period seismic signals on Mount Erebus volcano, Ross Island, Antarctica. *Geophysical Research Letters* 25: 2297–2300.

Saccorotti G, Almendros J, Carmona E, Ibanez JM, and Del Pezzo E (2001a) Slowness Anomalies from Two Dense Seismic Arrays at Deception Island Volcano, Antarctica. *Bulletin of the Seismological Society of America* 91: 561–571.

Saccorotti G, Chouet B, and Dawson P (2001b) Wavefield properties of a shallow long-period event and tremor at Kilauea Volcano, Hawaii. *Journal of Volcanology and Geothermal Research* 109: 163–189.

Saccorotti G and Del Pezzo E (2000) A probabilistic approach to the inversion of data from a seismic array and its application to volcanic signals. *Geophysical Journal International* 143: 249–261.

Sakai S, Yamada T, Ide S, *et al.* (2001) Magma migration from the point of view of seismic activity in the volcanism of Miyake-jima island in 2000. *Journal of Geography* 110: 145–155.

Sassa K (1935) Volcanic micro-tremors and eruption-earthquakes (Part 1 of the geophysical studies on the volcano Aso). *Memories of the College of Science, Kyoto Imperial University* 18: 255–293.

Sassa K (1936) Micro-seismometric study on eruption of the Volcano Aso. *Memories of the College of Science, Kyoto Imperial University* 19: 11–56.

Scarpa R and Tilling RI (1996) *Monitoring and Mitigation of Volcanic Hazards*. New York: Springer.

Schmidt RO (1986) Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation* 34: 276–280.

Schoenberg M and Sen PN (1983) Properties of a periodically stratified acoustic half-space and its relation to a Biot fluid. *Journal of the Acoustical Society of America* 73: 61–67.

Seidl D, Schick R, and Riuscetti M (1981) Volcanic tremors at Etna: A model for hydraulic origin. *Bulletin of Volcanology* 141: 43–56.

Sudo Y and Kong LSL (2001) Three-dimensional seismic velocity structure beneath Aso Volcano, Kyushu, Japan. *Bulletin of Volcanology* 63: 326–344.

Sutton AJ, Elias T, Gerlach TM, and Stokes JB (2001) Implications for eruptive processes as indicated by sulfur dioxide emissions from Kilauea Volcano, Hawaii, 1979–1997. *Journal of Volcanology and Geothermal Research* 108: 283–302.

Takagi N, Kaneshima S, Kawakatsu H, *et al.* (2006) Apparent migration of tremor source synchronized with the change in the tremor amplitude observed at Aso volcano, Japan. *Journal of Volcanology and Geothermal Research* 154: 181–200.

Takei Y and Kumazawa M (1994) Why have the single force and torque been excluded from seismic source models? *Geophysical Journal International* 118: 20–30.

Takeo M (1992) The rupture process of the 1989 offshore Ito earthquakes preceding a submarine volcanic eruption. *Journal of Geophysical Research* 97: 6613–6627.

Takeo M, Hamada N, Kashiwabara S, and Uhira K (1984) Analysis of long-period seismic waves excited by the explosive eruption of Mt. Asama on April 8. *Bulletin of the Volcanological Society of Japan* 29: 31–44.

Takeo M, Yamasato H, Furuya I, and Seino M (1990) Analysis of long-period seismic waves excited by the November 1987 eruption of Izu-Oshima volcano. *Journal of Geophysical Research* 95: 19377–19393.

Tameguri T, Iguchi M, and Ishihara K (2002) Mechanism of explosive eruptions from moment tensor analyses of explosion earthquakes at Sakurajima volcano, Japan. *Bulletin of the Volcanological Society of Japan* 47: 197–216.

Tsutsui T and Sudo Y (2004) Seismic reflectors beneath the central cones of Aso Volcano, Kyushu, Japan. *Journal of Volcanology and Geothermal Research* 131: 33–58.

Uhira K, Baba T, Mori H, Katayama H, and Hamada N (2005) Earthquake swarms preceding the 2000 eruption of Miyakejima volcano, Japan. *Bulletin of Volcanology* 67: 219–230.

Uhira K, Ikeda S, and Takeo M (1994a) Source process of explosion earthquakes deduced from short-period records at Sakurajima volcano. *Bulletin of the Volcanological Society of Japan* 40: 295–310.

Uhira K and Takeo M (1994) The source of explosive eruptions of Sakurajima volcano, japan. *Journal of Geophysical Research* 99: 17775–17789.

Uhira K, Yamasato H, and Takeo M (1994b) Source mechanism of seismic waves excited by proclastic flows observed at Unzen volcano, Japan. *Journal of Geophysical Research* 99: 17757–17773.

Ukawa M and Ohtake M (1987) A monochromatic earthquake suggesting deep-seated magmatic activity beneath the Izu-Ooshima Volcano, Japan. *Journal of Geophysical Research* 92: 12649–12663.

Vergniolle S, Brandeis G, and Mareschal J-C (1996) Strombolian explosions. 2: Eruption dymanics determined form acoustic measurements. *Journal of Geophysical Research* 101: 20449.

Waite GP, Chouet B, Dawson P, and Moran SC (2005) Very-long-period seismic source at Mount St. Helens. *EOS, Transactions of AGU* 86 (Fall Meet. Suppl.,): Abstract V53D–1601.

Wallace TC (1985) A reexamination of the moment tensor solutions of the 1980 Mammoth Lakes earthquakes. *Journal of Geophysical Research* 90: 11171–11176.

Watada S (1995) *Part I: Near-source Acoustic Coupling Between the Atmosphere and the Soild Earth During Volcanic Eruptions*. PhD Thesis, California Institute of Technology.

West M, Sanchez JJ, and McNutt SR (2005) Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake. *Science* 308: 1144–1146.

White R (1996) Precursory deep long-period earthquakes at Mount Pinatubo: Spatiotemporal link to a basalt trigger. In: Newhall CG and Punongbayan RS (eds.) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, pp. 307–327. Seattle: University of Washington Press.

Widmer R and Zürn W (1992) Bichromatic excitation of long-period Rayleigh and air waves by the Mount Pinatubo and E1 Chichon volcanic eruptions. *Geophysical Research Letters* 19: 765–768.

Wilson L and Head JW (1981) Ascent and eruption of basaltic magma on the earth and moon. *Journal of Geophysical Research* 86: 2971–3001.

Yamamoto M (2005) *Volcanic Fluid System Inferred from Broadband Seismic Signals*. PhD Thesis, University of Tokyo.

Yamamoto M, Kawakatsu H, Kaneshima S, *et al.* (1999) Detection of a crack-like conduit beneath the active crater at Aso volcano, Japan. *Geophysical Research Letters* 26: 3677–3680.

Yamamoto M, Kawakatsu H, Yomogida K, and Koyama J (2002) Long-period (12 sec) volcanic tremor observed at Usu 2000 eruption: Seismological detection of a deep magma plumbing system. *Geophysical Research Letters* 29: 1329.

Yamamura K (1997) *Slow Waves in Solid-Liquid Two Phase System* (in Japanese). Master's Thesis, University of Tokyo.

Yamamura K and Kwakatsu H (1998) Normal-mode solutions for radiation boundary conditions with an impedance contrast. *Geophysical Journal International* 134: 849–855.

Yamasato H, Fukui K, Uhira K, Hashimoto T, and Hori H (1993) Analysis of seismic and acoustic signals excited by pyroclastic flows at Unzen volcano. *Bulletin of the Volcanological Society of Japan* 38: 79–90.

Relevant Website

http://wwweic.eri.u-tokyo.ac.jp – Grid-Based Real-Time Determination of Moment Tensors (GRiD MT), Earthquake Information Center.

4.14 Interaction of Solid Earth, Atmosphere, and Ionosphere

T. Tanimoto, University of California, Santa Barbara, CA, USA

J. Artru-Lambin, Centre National d'Etudes Spatiales, Toulouse, France

Nomenclature

e	electron charge
f	frequency in Hertz (Hz)
f_1	primary carrier frequency of GPS
f_2	secondary carrier frequency of GPS
g	gravitational acceleration
k_x, k_y, k_z	three components of wave number
m_e	electron density
$\mathbf{v}$	velocity (three-component vector)
z	vertical coordinate ($z = 0$ is surface and positive upward)
C_s	sound speed
G	gravitational constant (6.67×10^{-11}) (MKS unit)
N_e	electron density in the upper atmosphere
γ	specific heat ratio
ϵ_0	permittivity of vacuum
ζ	ocean-surface displacement
$\overline{\zeta}$	averaged ocean-surface displacement over its wavelength
ρ	density
ω	$2\pi f$
ω_a	acoustic cut-off frequency
ω_g	Brunt–Väisälä frequency
Δg	perturbation to gravitational acceleration
ΔP	surface-pressure perturbation
$\Delta\rho$	density perturbation

Glossary

acoustic cut-off frequency The frequency of radio waves below which acoustic waves are reflected back to the Earth.

barometer, barograph Instrument to measure atmospheric pressure.

Brunt–Väisälä frequency When the medium is stably stratified, displaced particle in the medium starts to oscillate at this frequency. In an unstable medium, this frequency becomes complex.

GPS (Global Positioning System) Satellite-based positioning system. Two carrier frequencies, L1 (1.57542 GHz) and L2 (1.22760 GHz), are used.

Hum Free oscillation peaks observed as background noise continuously. Frequency range is between 0.002 and 0.007 Hz, approximately.

infrasound Low-frequency sound waves. Definition of frequency range is vague.

ionosonde, ionogram Ionosonde is a land-based system to send radio waves upward and record reflection from the upper atmosphere. Ionogram is its record. Ionosonde has been used to monitor the state of upper atmosphere.

microbarom Atmospheric low-frequency waves that have close frequencies to microseisms. Believed to have the same origin with microseisms.

microseism Largest seismic noise in seismograms observed continuously. Frequency range is about 0.1–0.4 Hz (period 2–10 s). Ocean waves are known to generate this noise. The largest peak in seismic noise is found at about twice the frequency of ocean waves. Small-amplitude noise at the frequency of ocean waves is called the primary-frequency microseisms, while 'microseisms' usually refer to the larger-amplitude double-frequency microseisms. Doubling (and sometimes tripling) of frequency at the maximum amplitudes are believed to be caused by nonlinear mechanism in the Navier–Stokes equation.

NLNM New Low Noise Model proposed by Peterson (1993).

plasma frequency Low-frequency cut-off for radio-wave propagation.

PREM (Preliminary Reference Earth Model) One-dimensional reference model published by Dziewonski and Anderson (1981).

superconducting gravimeter Sensitive gravimeter which uses superconductivity, typically Ni (niobium) ball floated in the magnetic field is used as its sensor.

TEC (total electron density) An integral of electron density along a path.

4.14.1 Introduction

The atmosphere has been traditionally ignored by seismologists in their analysis of seismograms. Seismologists typically regard the surface of the solid Earth as the upper boundary of the Earth and this boundary is treated as a traction-free boundary. The atmospheric pressure is not zero, however, and the correct boundary condition must reflect the existence of the atmosphere with its diminishing density with height. Strictly speaking, the

actual physical upper-boundary condition should be in the atmosphere as the radiation-boundary conditions.

Despite this approximate treatment of the surface boundary, seismology has been successful mainly because of two factors. The first is that the impedance contrast at the atmosphere/solid-earth boundary is quite large. Density contrast across this layer is larger than 2000 (1.2 vs 2700 in $\mathrm{kg\,m^{-3}}$) and P-wave velocity contrast is about 20 (300 vs 6000 in $\mathrm{m\,s^{-1}}$). If we take the case of P waves impinging on the solid surface from below, there is obviously some transmitted P-wave energy but the majority of the energy is reflected back into the solid media due to this high impedance ratio. Resulting seismograms hardly indicate the effects of the leakage of transmitted energy.

The second important factor is the low mass of the atmosphere. The mass of the solid Earth is 6.0×10^{24} kg, while the mass of the atmosphere is only 5.1×10^{18} kg. Therefore, phenomena in the solid Earth are affected very little even if we include atmospheric effects in the analysis. It is thus justifiable to ignore the atmosphere and use the free-surface boundary conditions for many phenomena.

But does this mean that there are no benefits in including the atmospheric layer in the analysis? In a nutshell, the purpose of this chapter is to point out that there are many benefits in doing so. We can immediately think of two potential advantages: the first is the possibility of obtaining information on solid-Earth processes from analyses of atmospheric waves. It has been pointed out that the eruption of Pinatubo volcano in 1991 generated air-coupled waves that provided information on the nature of the eruption process. Second, some earthquakes have been known to generate Rayleigh-wave-coupled air waves which provide quantitative information on excitation-source processes. These examples are analogous to the use of tsunami for the study of earthquake-source process, because though tsunamis themselves are mostly confined to the liquid layer (ocean), their level and of excitation often provides information about the slow rupturing process on an earthquake fault. Analyses of atmospheric waves have not been as useful as analyses of tsunami so far, but more careful studies of atmospheric waves may lead to useful applications to solid-Earth processes.

Another benefit comes from the analysis of seismic noise. Most seismic noise in seismograms is generated by interactions between the solid Earth and the atmosphere or the ocean. In the past 5–6 years, such noise was shown to be practically useful because we can recover Green's functions between any pair of stations using a cross-correlation technique. Its application to Earth structure study is under way: seismic noise can provide surface-wave dispersion curves between two stations and promise to provide results that were not available from earthquake data. The cross-correlation technique also provides information on the flow (direction) of propagating seismic energy, thereby providing potential information about the source location of noise, which was not possible to determine before.

This chapter consists of two main parts: in the first part, the nature of seismic noise and the Green's function cross-correlation technique are described. A review of the study of seismic noise and how noise illuminates our understanding of atmosphere–solid-Earth interactions is presented here. Since the atmosphere and oceans are strongly coupled and often not separable, the oceans are involved in many of these processes.

In the second part, events such as large atmospheric explosions or earthquakes, the effects of transient sources in the solid Earth, and the resulting effects in the atmosphere are discussed. These events radiate energy in various types of waves, such as seismic waves in the solid Earth, tsunamis in the oceans, and internal waves in the atmosphere. Dynamical coupling allows the transmission of a part of the wave energy across the sea or Earth's surface, hence the generation of atmospheric, ocean, or tsunami waves from sources external to their respective domains. Sources of such coupled waves include (1) atmospheric explosions: natural (meteors, volcanoes) or man made, (2) earthquakes, and (3) sonic booms from supersonic jets (Concorde, Space Shuttle). Some notable effects observed after the great Sumatra–Andaman earthquake will also be discussed.

The first and second parts take two different views on modeling the source–medium relationship. The first part follows the view that seismic waves generated by atmospheric effects are generated by surface atmospheric-pressure variations. The Earth is basically separated into two distinct media in contact. In this view, generation of acoustic/gravity waves in the atmosphere by sources in the solid Earth is mainly by vertical surface displacement. On the other hand, the second follows the so-called full-coupling approach. There is no artificial boundary between the atmosphere and the solid Earth and the whole Earth is treated as a single medium. Normal modes are the normal modes of the whole system and propagating waves cross the atmosphere–solid-Earth boundary just like other boundaries inside the Earth. In this view, sources are stress gluts or the Reynolds stresses. The approach in the first part is clearly an approximation, but it seems to holds well.

4.14.2 Seismic Noise

Peterson (1993) summarized seismic-noise characteristics from global distribution of 75 stations. One of the most frequently quoted results from this study is the parametrized model called the New Low Noise Model (NLNM). **Figure 1** shows a plot of this model in the form of acceleration power spectral density. This model has often been used as a reference in the discussion of seismic-noise studies.

From the perspective of a seismologist, one of the main features in **Figure 1** is the existence of a low-noise frequency band from 2 to 20 mHz. Existence of this low-noise band explains why analyses of long-period body and surface waves have been successful in the past few decades as these waves exist in this low-noise band (1 mHz–10 Hz). Detection of small-amplitude waves has been possible because of this low-noise characteristics which is not as easily done outside this frequency band.

From the perspective of a scientist interested in the atmosphere–solid Earth interactions, **Figure 1** shows three prominent peaks, from the low-frequency end: (1) the low-frequency peak below 0.1 mHz, whose peak is outside the range of this plot but the decreasing trend with frequency up to about 2 mHz is clearly recognized; (2) a small peak or a bump at about 7–10 mHz (denoted as Hum in **Figure 1**); and (3) large-amplitude peak(s) at 0.1–0.2 Hz (microseisms). In this section, each of the three peaks from the low-frequency end is discussed, particularly focusing on the question which components in the Earth system interact to create these peaks.

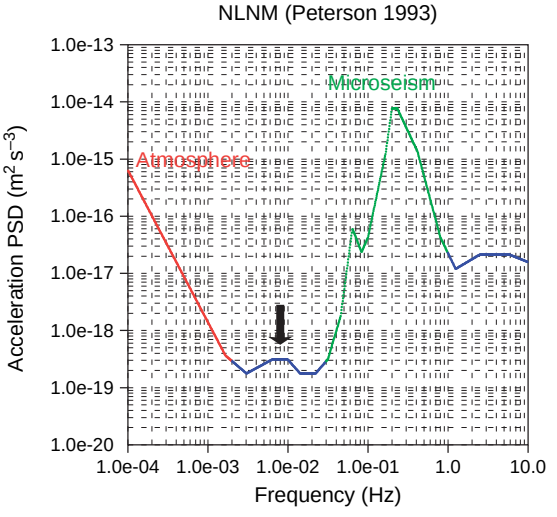


Figure 1 New Low Noise Model by Peterson (1993). Power spectral density (PSD) in unit for acceleration is used for this plot. Frequency range is from 1/10 000 to 10 Hz.

The model NLNM was derived from vertical-component seismograms. The noise in horizontal components is often an order of magnitude larger than the noise in vertical component because horizontal components are affected by tilts that are generated by temperature variations and atmospheric pressure variations. They are often dominant sources of noise in horizontal-component seismographs but are confined to shallow depths and local regions. In this chapter, these noise in horizontal components will not be discussed.

4.14.2.1 Low-Frequency Peak: Atmospheric Effects

Seismic noise below 2 mHz (millihertz) increases toward lower frequencies (**Figure 1**). This is a global feature that has been confirmed repeatedly from long-period seismographs and gravimeters.

4.14.2.1.1 Cause

The main cause of this noise is density changes in the atmosphere. For example, as weather patterns change, density changes occur in the atmosphere. These changes affect the gravitational field and generate small perturbations in gravitational acceleration at the surface of the Earth.

This mechanism was pointed out by Warburton and Goodkind (1977) by analyzing two co-located instruments, a barometer and a superconducting gravimeter. Barometric (surface pressure) data represent an integration of density anomalies along a vertical column above a station:

$$\Delta P = \int_0^\infty \Delta\rho g \, dz \qquad [1]$$

where ΔP is the surface pressure perturbation, $\Delta\rho$ is density perturbation in this formula, and the integration is from the surface of the Earth ($z = 0$) to infinity. Effects of atmospheric density perturbations above an observing station can be approximated by a formula similar to the Bouguer gravity formula:

$$\Delta g = 2\pi G \int_0^\infty \Delta\rho \, dz \qquad [2]$$

where Δg is the perturbation to gravitational acceleration and G is the gravitational constant. Since g does not vary very much near the surface, barometric data and gravity data are related, to a good approximation, by

$$\Delta g = \frac{2\pi G}{g} \Delta P \qquad [3]$$

Warburton and Goodkind showed that co-located gravimeter data and barometric data correlate with coefficients close to a value predicted by $2\pi G/g$.

There is, however, an assumption in using the Bouguer gravity formula; in applying this formula, density perturbation in the atmosphere is assumed to be shaped like a disk-like pattern, and density variations laterally away from the station location (in latitude and longitude) do not affect the gravitational acceleration very much. More careful evaluation of such effects, using a meteorological simulation model, showed that deviation from this formula is small (Boy *et al.*, 1998; Hinderer and Crossley, 2000).

Changes in the atmospheric pressure also cause deformation of the elastic Earth as a surface load. Correction due to this effect is usually added to the above formula in the analyses (Warburton and Goodkind, 1977; Spratt, 1982; Van Dam and Wahr, 1987; Merriam, 1992; Crossley *et al.*, 1995). This effect is not negligible and in fact was claimed to be larger than the above effect in earlier period of seismic-noise study (Sorrells, 1971; Sorrells *et al.*, 1971). But it is smaller than the right-hand side of the above formula.

A clear demonstration of atmospheric pressure effects at the time of a cold front passage was documented by Müller and Zürn (1983) and some careful quantitative computational analysis was reported by Rabbel and Zschau (1985).

Figure 2 shows the evidence of correlation between gravity data and surface-pressure data after the removal of tidal signals from gravimeter data. The original gravity record (top) is dominated by tidal signals but after their removal, the middle trace is obtained. This time series matches the barometric data at bottom closely, indicating peak-to-peak correlations between the middle trace and the bottom trace.

4.14.2.1.2 Turbulence

While it seems that the correlation between gravity (ground motions) and surface-pressure data is good as shown in **Figure 2**, the correlation coefficient hovers around 0.5 and never becomes close to 1. Uncorrelated part of the signal may be due to additional forces, such as wind stress (Reynolds stress) in atmospheric turbulence, which dominates the frequency range higher than above 1 mHz (Tanimoto, 1999). Frequency dependence of surface pressure in this frequency band shows the $1/f$ behavior as observed spectra in **Figure 3** show (e.g., Tanimoto and Um, 1999). This $1/f$ behavior is consistent with the prediction by the theory of turbulence with the Kolgomorov-type scaling relation (Landau and Lifshitz, 1987) and directly supports the importance of turbulence in this frequency band.

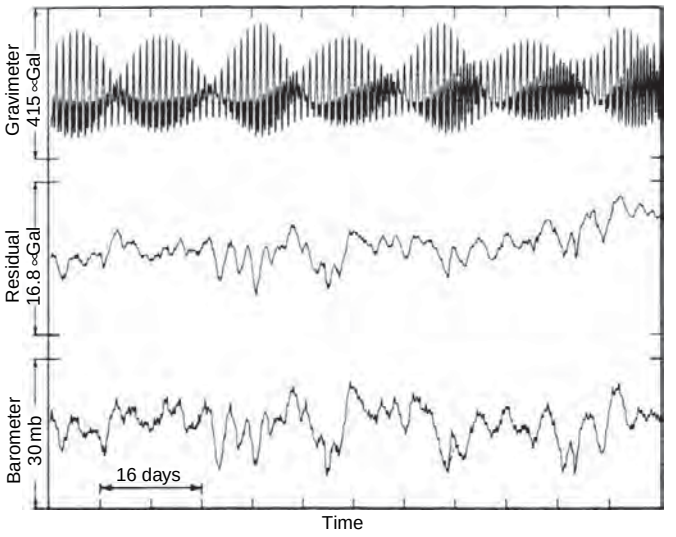


Figure 2 *Top*: Original gravity signal from Warburton and Goodkind (1977). Data were recorded by superconducting gravimeter. *Middle*: Gravity signal after tidal signal removal. *Bottom*: Barometer signal, which correlates well with the trace in the middle.

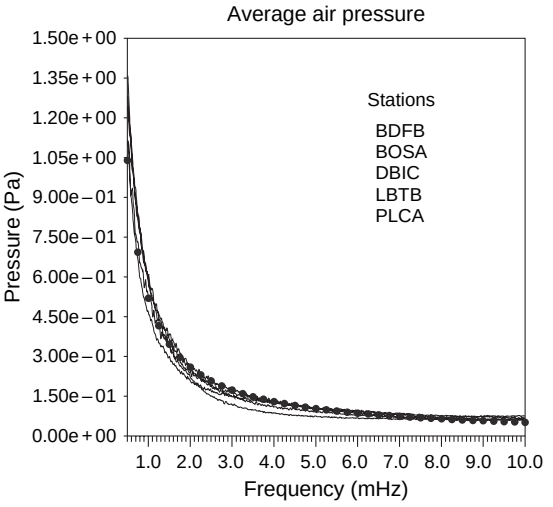


Figure 3 Barometric data (pressure) from five stations in North and South America. Analysis was performed for 20-day long signals and spectral amplitudes were averaged over 0.2 mHz range. Solid circles denote 1/*f* trend.

4.14.2.1.3 Reduction of seismic noise

Existence of the above mechanism suggests that, if a seismic/gravity instrument is co-located with a barometer, one can reduce the level of noise considerably by taking advantage of this correlation. For analyses of tidal and lower-frequency data, this has become a common practice since the study of Warburton and Goodkind (1977). For seismic data, especially for normal-mode studies (about 1 mHz and above), Zürn and Widmer (1995), Beauduin *et al.* (1996), and Roult and Crawford (2000) showed significant noise reduction. These studies showed that noise level below 1 mHz can be made lower than the model NLNM by this correlation technique.

4.14.2.2 Hum

The second peak denoted by Hum in **Figure 1** was identified as a broad peak by Peterson (1993). It was shown later that this peak is associated with multiple modal peaks, primarily on the left-hand side of its maximum. **Figure 4** shows an example from data in this frequency range, obtained from global network data by averaging 11 stations located at various parts of the world. The bottom panel is an enlarged figure within the small box in the top panel and shows that modal peaks exist for the frequency range between 2 and 7 mHz. They are shown to match the eigenfrequencies of fundamental spheroidal modes almost exactly, as the eigenfrequencies for the Preliminary

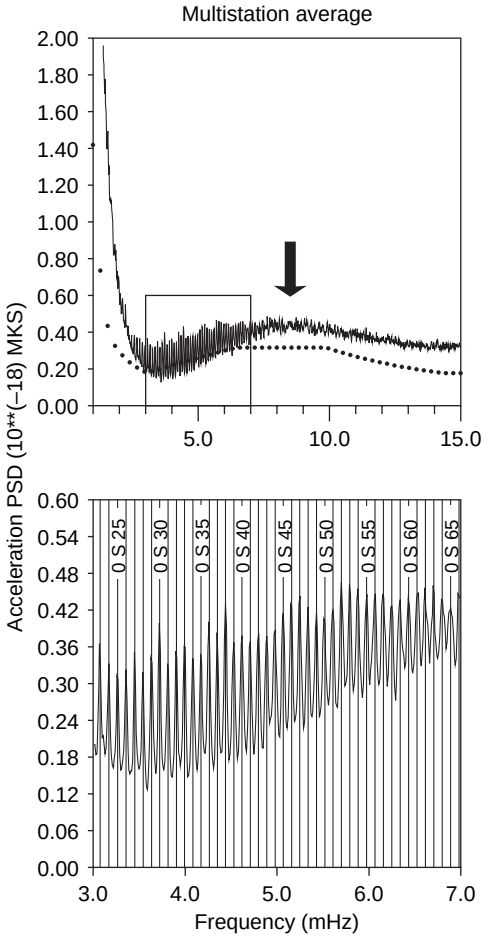


Figure 4 Averaged acceleration PSD from 11 stations, distributed from various parts of the world, is shown at top. Circles in the top panel are NLNM. Spectra in the small box (top) are enlarged in the bottom panel. Each modal peak is shown to match the eigenfrequency of fundamental spheroidal mode (vertical lines in the bottom panel).

Reference Earth Model (PREM; Dziewonski and Anderson, 1981) are drawn in the bottom panel by vertical lines. These continuously excited modes are all fundamental spheroidal modes and do not seem to contain any overtone modes. They are now commonly referred to as the Earth's Hum.

4.14.2.2.1 Discovery

Discovery of continuously excited modal peaks was made in 1997. A broad peak in Peterson's model has been noted since 1993, or perhaps even earlier, but the hum was discovered as an independent feature from it. The initial reports were by Suda *et al.* (1998), Kobayashi and Nishida (1998), and Tanimoto *et al.* (1998). These and subsequent studies showed that all

sites where the noise level is as low as 10^{-18} ($m^2 s^{-3}$) in acceleration spectral density (for the frequency band 2–15 mHz) show signals of the Hum. Since the global minimum of horizontal noise is an order of magnitude higher than this value, all observations were from vertical-component seismograms or gravimeters. The types of seismic intruments that led to these discoveries included broadband seismometers STS-1 and gravimeters with spring sensor (LaCoste–Romberg gravimeters) or superconducting sensors. It was recently shown that seismic data from broadband seismometers STS-2 (with lower pendulum period than STS-1, about 100 s) at the Black Forest Observatory (BFO) in Germany show the Hum (Widmer-Schnidrig, 2003). This may not be surprising, however, since the noise level is as low as 10^{-18} ($m^2 s^{-3}$) for STS-2 at this low-noise site. The most important point seems to be that the signal of the Hum is observed if the noise level is about 10^{-18} ($m^2 s^{-3}$) or less, in unit for acceleration spectral density.

Nawa *et al.* (1998) reported the existence of the Hum in their superconducting gravimeter data at Showa station in Antarctica. Their report was made before the three papers cited above, but the reported results contain peaks from grave spheroidal modes (angular degree less than 10) that have not been confirmed independently. In addition, their later study (Nawa *et al.*, 2000) showed that the noise level at this station seems to be much higher than 10^{-18} ($m^2 s^{-3}$) for the frequency band 1–5 mHz.

4.14.2.2.2 Seasonal variations

These modal amplitudes display seasonal variations; claims of the predominant 6 months periodicity were made by Tanimoto and Um (1999) from a frequency-domain analysis and by Ekström (2001) from an entirely independent time-domain analysis. A claim for an annual seasonality was made by Nishida *et al.* (2000), although they claimed that annual signal was dominant and was thus mildly different from the above two studies.

Detection of seasonal variations had major implications for the source of excitation of the Hum because it basically removed causes in the solid Earth, because phenomena in the solid Earth do not usually have clear seasonal signatures. Up until these discoveries on seasonality were made, slow earthquakes (e.g., Beroza and Jordan, 1990) were considered to be one of the major candidates for the cause of the Hum.

4.14.2.2.3 Excitation mechanism

The cause of the Hum may be in the atmosphere or in the oceans. The atmospheric excitation was advocated by Kobayashi and Nishida (1998), Tanimoto and Um (1999), and Fukao *et al.* (2002). In these papers, pressure fluctuations in the turbulent atmosphere were postulated as the cause. Modal amplitudes of individual peaks were shown to be explained by the atmospheric excitation model. But there was an uncertain part in this scenario, particularly on the correlation length in atmospheric turbulence within the frequency band (2–7 mHz range). This was critical because the excitation of modes by turbulent atmosphere is proportional to this parameter (Tanimoto, 1999; Goldreich and Keeley, 1977). The atmospheric excitation mechanism was advocated by assuming that this correlation length was about 1 km or larger for the frequency band 2–7 mHz. This correlation length has not been confirmed by observation although the existence of turbulent boundary layer of thickness 1 km seems to imply that it may be a viable candidate.

In addition to this uncertainty in source strength, the atmospheric excitation hypothesis does not provide a good reason for the existence of the broad noise peak between 3 and 15 mHz. Both Tanimoto and Um (1999) and Fukao *et al.* (2002) assumed this broad spectral peak to be caused by an unknown background noise and only the modal peaks above this background noise were modeled by the atmospheric-pressure variations at the surface.

The oceanic excitation hypothesis was advanced by Rhie and Romanowitz (2004) and Tanimoto (2005). Rhie and Romanowitz (2004) used seismic-array data and located the sources of Rayleigh waves (noise) using two arrays in Japan and California. They found the excitation sources to be in the oceans, especially in the mid-latitude bands (30–60°) in the Northern and Southern Hemispheres. They also claimed that source locations switched rather abruptly between the Northern and Southern Hemispheres. This feature seems to be compatible with general patterns of storm behaviors, although further independent confirmation is desirable.

Tanimoto (2005) showed that the overall spectral shape of the hum, the modal peaks, and the broad spectral peak depicted in **Figure 1** (3–15 mHz) can be explained by a single mechanism if the oceanic infragravity waves were the cause. A rather *ad hoc* feature in the atmospheric excitation hypothesis, which has to find separate causes for modal peaks

and for the broad spectral peak, can then be avoided. Nishida *et al.* (2005) argue that it may still be possible to create this broad peak by atmospheric effects but a detailed mechanism is still missing in the atmospheric excitation hypothesis.

One of the uncertainties in the oceanic excitation hypothesis lies in our lack of knowledge on the oceanic infragravity waves. Limited number of observations (e.g., Watada and Masters 2001) are now supplemented by new observations. Also, some new understanding as to the generation of infragravity waves from oceanic swells is emerging from observation (Dolenc *et al.*, 2005). Although there have been some studies on the oceanic infragravity waves and many results (Webb *et al.*, 1991; Okihiro *et al.*, 1992; Dolenc *et al.*, 2005) point to near-coastal generation of waves, a comprehensible, total picture of the generation mechanism seems to be missing. This is still an active research area.

Satellite ocean-wave data provide semihemispheric switching of activities, as shown in **Figure 5**. Ocean waves occasionally reach 10 m or more in high-activity regions which generates pressure perturbations higher than surface atmospheric pressure. This behavior seems to explain the 6-month periodicity naturally as well as source locations of Rayleigh waves observed by Rhie and Romanowitz (2004). This satellite evidence does not necessarily prove the oceanic excitation mechanism for the Hum because strong atmospheric winds are associated with these ocean-wave behavior and the atmosphere also contains 6-month periodicity. There is clearly an inherent difficulty in the argument of atmospehric versus oceanic excitation because the atmosphere and the oceans are coupled in almost all scales.

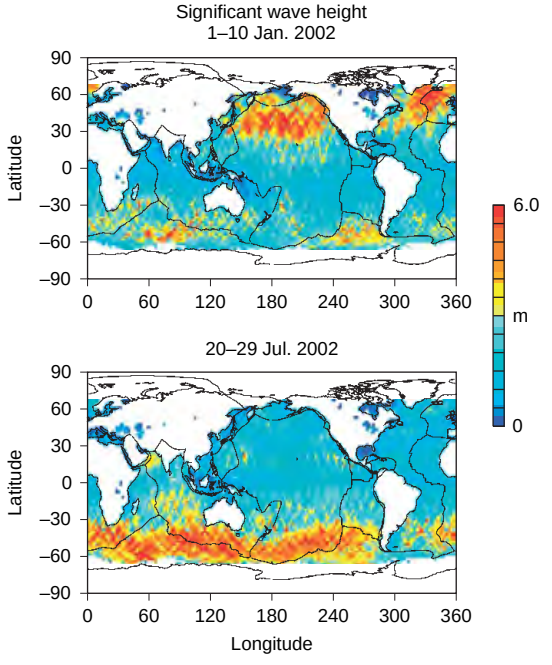


Figure 5 Significant wave-height data from TOPEX/POSEIDON data. Data from two selected periods, one in January (top) and the other in July (bottom), are shown. In January, ocean waves are energetic in the mid-latitude region in the Northern Hemisphere, while ocean waves in the Southern Hemisphere become energetic in July.

4.14.2.2.4 Ubiquitous Rayleigh waves

One of the most important notions developed in these studies is that the energy associated with the Hum, for the entire frequency range 3–15 mHz, consists of Rayleigh-wave energy. From observation, Nishida *et al.* (2002) showed that signals between 2 and 20 mHz have similar phase velocities to Rayleigh-wave phase velocities predicted by the PREM (Dziewonski and Anderson, 1981). Tanimoto (2005) showed that the whole spectra in this frequency band can be systhesized by normal-mode summation of spheroidal modes, potentially excited by the oceanic infragravity waves, thereby indicating that they are Rayleigh waves.

Along with the microseisms in the next section, a majority of seismic noise seems to be dominated by Rayleigh waves, at least those signals (noise) in vertical-component seismograms.

4.14.2.3 Microseisms

It has been noted for a long time, since the early twentieth century, that the most obvious and perhaps annoying noise in seismograms are the microseisms with the peak frequency at about 0.1–0.4 Hz (**Figure 1**). Amplitudes of this noise was so overwhelming that, before the development of high-dynamic-range digital seismic instruments, seismologists recorded seismic waves separately for high-frequency range (above 0.5 Hz) and for low-frequency range (below 0.1 Hz) in order to avoid this microseismic noise. World-Wide Standard Seismograph Network, which played the central role for the development of global seismology from 1960s to 1980s, had such separate (short-period and long-period) instruments.

Modern instruments with high dynamic range and digital recordings removed such a cumbersome recording procedure (Wielandt and Steim, 1986).

They record seismic signals from about 1 mHz to 10 Hz using the same sensor. A benefit of such recording procedure is that microseisms are now recorded continuously and this has raised interests among some seismologists since we now have access to raw microseism data from modern high-dynamic-range digital data.

4.14.2.3.1 Nature of waves

Microseisms mostly consist of Rayleigh waves. This was shown by Lee (1935) and Haubrich *et al.* (1963) by particle motion analysis which indicated a characteristic retrograde elliptical particle motion. Later, array analysis (Lacoss *et al.*, 1969; Capon, 1972) showed that some higher-mode energy and Love-wave energy were mixed in the signals. Earlier, Gutenberg (1958) also discussed two types of microseisms, apparently referring to predominant Rayleigh waves and occasional S waves. From modern three-component data, it is easy to confirm the dominance of Rayleigh waves in microseisms from phase velocity measurements from array observations (Lacoss *et al.*, 1969; Capon, 1972) or from phase-shift observations between horizontal and vertical components which indicate mostly 90° phase shifts.

4.14.2.3.2 Excitation mechanisms

One of the important characteristics in microseisms is the fact that larger amplitudes occur for the double-frequency microseisms (0.1–0.4 Hz) than the primary-frequency microseisms that have the same predominant frequencies with ocean swells (0.05–0.07 Hz). It is still not clear why the double-frequency microseisms have such large amplitudes but the most widely accepted (basic) mechanism for the double-frequency microseisms is that of Longuet-Higgins (1950). Longuet-Higgins (1950) showed that interactions of two ocean waves in opposite direction can create the double-frequency microseisms through the nonlinear (advection) term in the Navier–Stokes equaiton. The crux of his theory is that, for surface displacement ζ, pressure perturbation at sea bottom is given by

$$p(t) = \rho \frac{\partial^2}{\partial t^2} \left\{ \frac{1}{2} \bar{\zeta}^2 \right\} \qquad [4]$$

even when the depth extent of colliding waves, that make surface displacement ζ, do not reach the sea bottom. In fact, this pressure arises as a constant in the Bernoulli equation (Longuet-Higgins, 1950), implying that this pressure occurs at all depths. The bar denotes an averaging procedure over a wavelength. For example, if two ocean waves with the same frequency are propagating in opposite directions

$$\zeta(x, t) = a_1 \cos(\omega t - kx) + a_2 \cos(\omega t + kx) \qquad [5]$$

we get

$$p(t) = -\rho a_1 a_2 \omega^2 \cos(2\omega t) \qquad [6]$$

from eqn [4]. This formula contains two important features; first, it shows the occurrence of double-frequency pressure variations. These double-frequency pressure variations at sea bottom generate the double-frequency microseisms. Second, this formula shows that if there exists a unidirectional propagating wave, say $+x$ meaning $a_2 = 0$, this pressure term goes to zero because it is proportional to $a_1 a_2$. The double-frequency pressure variation requires existence of colliding waves. This mechanism was experimentally confirmed by Cooper and Longuet-Higgins (1950). Phillips (1977) showed a derivation of the same formula from the Navier–Stokes equation by carefully analyzing the vertical momentum balance.

Hasselmann (1963) casts the problem in a more general context, expressing the ocean wavefield as a wave-number integral, and then showing that the predominant term is equivalent to the Longuet-Higgins formula. Normal-mode excitation analysis of seismic wavefields by nonlinear interactions of ocean waves also showed recently that the Longuet-Higgins pressure formula naturally arises from the analysis (Tanimoto, 2007).

4.14.2.3.3 Source location

While it is clear that nonlinear interaction of ocean waves is necessary to generate microseisms, there is some confusion as to exact locations of excitation sources. This is partly due to the fact that the Longuet-Higgins' mechanism is effective at any depth; regardless of ocean depths, the pressure in the above formula is transmitted to the ocean bottom. What we know for sure is the fact that a source must be a place where ocean-wave collisions occur. This leads us to two potential locations. One is a coast where we can expect some reflected ocean waves from a coast can collide with incoming ocean waves. The other is an eye of low-pressure systems.

Close relationship between microseisms and ocean swells near the shore was shown by Haubrich *et al.* (1963) and more recently by Bromirski *et al.* (1999). The latter involved a correlation study

between microseisms from seismograms and ocean waves from buoy data and demonstrated a high correlation and thus a causal relationship.

On the other hand, it seems also true that storms provide a condition that ocean waves collide from opposite directions near their eyes. Tracking of a storm was done by Santo (1960), Sutton and Barstow (1996), and Bromirski (2001) in order to understand the source process of the double-frequency microseisms. The work by Bromirski (2001) provided the most quantitative analysis and reported that the dominant source area for the double-frequency microseisms was not in the open ocean, where the highest waves occurred, but was near the coast. It seems therefore that standing ocean waves that occur near the coasts, even at the time of large low-pressure system, are the source of double-frequency microseisms. This does not mean that the eye of the low-pressure system does not send out seismic waves; it simply seems to be dwarfed by near-coastal processes.

Friedrich *et al.* (1998) performed a careful array analysis of seismic data and reported that the primary microseisms contained more Love waves than the double-frequency microseisms, suggesting that the exciting mechanisms may be quite different between them.

It has also been noted that source area of primary microseims is often spread out along the coasts whereas source areas of the double-frequency microseisms appear to be specific locations. This was observed in Europe (Friedrich *et al.*, 1998), on the Atlantic coast (Bromirski 2001), and the Pacific coast of the United States (Bromirski *et al.*, 1999; Schulte-Pelkum *et al.*, 2004). Also Schulte-Pelkum *et al.* (2004) showed a temporary effect of microseismic signals in California caused by ocean waves on the Atlantic coasts, although occurrence of such phenomena does not seem to be common.

In summary, there is still confusion as to the exact locations of excitation but the importance of near coastal region has become clear, at least for double-frequency microseisms.

4.14.2.3.4 Microbaroms

Microbaroms are atmospheric low-frequency waves, having frequencies close to those of double-frequency microseisms. Similarity of power spectra to those of microseisms suggest that they are of the same origin (e.g., Donn and Naini, 1973; Nishida *et al.*, 2005). Posmentier (1967) presented a theory analogous to the mechanism proposed by Longuet-Higgins (1950). Arendt and Fritts (2000) published a more complete theory, carefully analyzing various types of waves that arise from interactions of ocean waves. The basic mechanism in both studies is the same with the Longuett-Higgins mechanism for microseism generation and assumes interactions of two surface waves propagating in opposite directions. Standing ocean waves near the coasts appear to be generating both microseisms and microbaroms.

4.14.2.3.5 Implication to past climate

Changes in the climate are likely to be related to ocean-wave behaviors (Bromirski *et al.*, 1999; Bromirski and Duennebier, 2002), which may threaten coastal areas if they become too energetic. However, recovery of past ocean-wave heights from meteorological data is generally difficult. Grevemeyer *et al.* (2000) proposed to use activities of microseisms in historical seismograms as a proxy for obtaining information for historical ocean-wave behavior. They reported that the number of high microseismic days increased from 7 to 14 days in the last 50 years and suggested that this may be related to global warming indirectly. While this result should be regarded as preliminary, it is reasonable to expect some correlation and microseism records may turn out to be a useful source of information for ocean-wave behavior and thus for climate changes.

4.14.2.4 From Noise to Structure

One of the motivations to study seismic noise is its potential use for Earth structure study. Many techniques have been developed, each focusing on different aspects of data; examples include, in historical order, spatial variation of correlation amplitudes, array analysis for Rayleigh-wave phase velocity measurements, use of horizontal–vertical amplitude ratios for Rayleigh-wave signals, and the two-station correlation technique to recover the Green's functions between a pair of stations. Below we first discuss the most recent development, the correlation technique to recover Green's functions in diffuse seismic wavefields, and then discuss other techniques under classification as the traditional methods.

4.14.2.4.1 The correlation technique in diffuse wavefield

The first application of the correlation technique to recover structural information was demonstrated in helioseismology (Duvall *et al.*, 1993). The essence of the technique was to recover traveltime versus

distance curve for multiply reflected waves through cross-correlation of data at various distances. In the case of the Sun, seismic sources are stochastic in time and space, and thus there are no specific noise sources; it can be anywhere and any time. Therefore, such a technique was most needed and worked effectively to get the results.

Applications to other fields, such as geophysical exploration (Rickett and Claerbout, 1999) and ultrasonics (Weaver and Lobkis, 2000), followed before seismology adopted it recently (Campillo and Paul, 2003). Theoretical basis to recover the Green's functions from noise was discussed in Weaver and Lobkis (2001), Lobkis and Weaver (2001), van Tiggelen (2003), Snieder (2004), and Roux *et al.* (2005a).

Applications to seismic data started with Campillo and Paul (2003). Recovered signals were dominated near the microseismic frequency bands (about period 5–10 s) in earlier demonstrations. The technique was soon applied to local to regional-scale problems such as Southern California (Shapiro *et al.*, 2005; Sabra *et al.*, 2005a, 2005b). It is now applied to much wider frequency bands (up to 100 s in period) and is also applied to continental-scale structure.

Typical seismic applications first recover the Green's functions by cross-correlation of seismograms at two stations. Since data are mainly from vertical components and are dominated by Rayleigh waves, traditional dispersion measurements for group velocity are often used to retrieve Earth structure. Dominance of fundamental-mode surface waves in noise is perhaps unavoidable because the excitation sources are in the atmosphere and the oceans, although there are now some studies which report successful recovery of body-wave signals (e.g., Roux *et al.*, 2005b). This is currently a rapidly expanding field and the landscape of the research field is expected to change quickly in a few years.

4.14.2.4.2 Traditional techniques

Aki (1957, 1965) proposed a method, often referred as the spatial autocorrelation method (SPAC), to determine local seismic structure. This technique also uses the cross-correlation among stations but focuses on amplitude variations of the correlated signals in space; in a flat layered media, the correlation function becomes proportional to Bessel functions (Aki, 1965) from which one can obtain phase velocity. Therefore, it is quite different from the above correlation technique that focusses on phase information.

Lacoss *et al.* (1969) showed that an array-based phase velocity measurement is a powerful approach. This is basically a beam-forming technique, using array data, and is becoming increasingly popular because of availability of dense seismograph networks.

Both SPAC and the beam-forming approach are still used at present and an enormous body of literature exists in this area, especially in geotechnical engineering. In geotechnical engineering, the frequency range extends up to 10–20 Hz, exceeding the microseismic frequency band (0.1–0.4 Hz). These studies not only use microseismic signals but higher-frequency signals generated by other sources and can only be called a cultural noise. A good recent summary of this field was given by Okada (2003).

These traditional techniques have been around 40–50 years but their practical use may expand as quality and density of seismic instruments have improved recently. We expect to see more use of these methods because there are urgent needs, especially in urban area, to understand near-surface structure as shallow seismic structure plays a critical role for ground motion amplification at the time of major earthquakes.

4.14.3 Localized Sources of Interactions

In this section we will focus on the interaction between the solid Earth and the atmosphere after localized and transient events such as earthquakes and volcanic eruptions.

4.14.3.1 Historical Context

An early example of observations can be traced back to more than a century ago, the eruption of the Krakatoa volcano (Indonesia) on 26 August 1883. After the eruption, coupled air-sea waves were observed in barographs and tide gauges worldwide (Harkrider and Press, 1967). The meteor or comet explosion in Tunguska (Central Siberia) on 30 June 1908 also induced both atmospheric and seismic waves. The latter was similar in amplitude to a $M = 5$ earthquake (Whipple, 1930), although the true source of those waves was an explosion in the air, at an estimated altitude of 8 km (Ben-Menahem, 1975).

The interest for the study of such interactions rose significantly later, during Cold War periods, because atmospheric gravity waves were emitted by nuclear

explosions and detection and characterization of such waves became an active research field (Hines, 1972). This led to three major advances during the 1960s. First, a better and efficient theoretical description of internal waves in the atmosphere (Hines, 1960), from the ground up to ionospheric heights, was developed. Second, ionospheric sounding networks capable of monitoring the ionospheric response to those waves (Davies, 1962) were deployed. And third, the occurrence of several major earthquakes (Chile, 22 May 1960, $M = 9.5$; Alaska, 28 March 1964, $M = 9.2$) revealed the generation of internal acoustic waves in the atmosphere by global Rayleigh-wave propagation (Donn and Posmentier, 1964).

More recently, significant advances were made toward a quantitative interpretation and prediction of such coupled phenomena. In particular, new types of atmospheric observations have been made possible by the development of Global Positioning System (GPS) ionosphere monitoring (Mannucci *et al.*, 1998; Calais and Minster, 1998) and by the deployment of the International Monitoring System (global seismological, hydroacoustic, radionuclide, and infrasound network aimed at ensuring compliance with the Comprehensive Nuclear Test Ban Treaty). Increased accessibility of numerical computations has also made it possible to refine theoretical modeling of coupled systems (LePichon *et al.*, 2003; Artru *et al.*, 2004).

4.14.3.2 Theoretical Preliminaries

4.14.3.2.1 Wave propagation in the atmosphere

After localized events, energy is transmitted in the form of atmospheric internal waves from the surface to the ionosphere. These waves arise from the interactions of compressional and gravitational forces and are divided into two classes: long-period gravity waves and short-period acoustic waves. The basic physics of acoustic–gravity waves was formulated by Hines (1960), and his formalism is now widely used. Let us review the main features of this theory.

We consider an isothermal atmosphere, initially in hydrostatic equilibrium, and include forces from inertia, gravity, and pressure gradients. We assume that disturbances can be regarded as adiabatic process because wave propagation is a sufficciently fast process. We do not include the effects from the rotation of the Earth, and therefore large-scale tidal and planetary waves are not considered here. We focus on much shorter-wavelength waves.

Let us use the notations that ρ is the density, p is the pressure, $\mathbf{v}$ is the neutral gas velocity, $\mathbf{g}$ is the gravitational acceleration and C_s is the constant sound speed. We then have the following three basic equations:

Conservation of mass:

$$\frac{\partial \rho}{\partial t} + \mathbf{v}\cdot\nabla\rho = -\rho\nabla\cdot\mathbf{v} \qquad [7a]$$

Conservation of momentum:

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} = \mathbf{g} - \frac{1}{\rho}\nabla p \qquad [7b]$$

Adiabaticity:

$$\frac{\partial p}{\partial t} + \mathbf{v}\cdot\nabla p = C_s{}^2\left(\frac{\partial \rho}{\partial t} + \mathbf{v}\cdot\nabla p\right) \qquad [7c]$$

In the equilibrium state, $\mathbf{v}_0 = 0$ and both ρ_0 and p_0 are proportional to $\exp(-z/H)$, where $H = C_s^2/\gamma|g|$ is the density scale height and γ is the specific heat ratio. Assuming ρ_1, ρ_1 and $\mathbf{v}$ are small perturbations with no dependency on the y-axis, we may solve the linearized equations as harmonic solutions by assuming that ρ_1, p_1, and $\mathbf{v}$ are proportional to $\exp\left[i\left(\omega t - k_x x - k_z z\right)\right]$. The full dispersion relation takes the form

$$\omega^4 - \omega^2 C_s^2\left(k_x^2 + k_z^2\right) + (\gamma - 1)g^2 k_x^2 + i\gamma g\omega^2 k_z = 0 \qquad [8]$$

This equation means that, in the presence of gravity, no solution exists in which both k_x and k_z are purely real and different from zero. Let us assume that k_x is real and seek a solution that propagates in the x-direction as a harmonic wave. There are now two possibilities, either k_z is purely imaginary or

$$k_z = k_z' + \mathrm{i}\frac{\gamma g}{2c^2} = k_z' + \mathrm{i}\frac{1}{2H} \qquad [9]$$

The first case (k_z pure imaginary) is appropriate for horizontally propagating surface waves but it permits no variation of phase with height. The second case is appropriate for traveling disturbances in the atmosphere under the influence of gravity. Thus, the second case is pursued here. The dispersion relation can be rewritten as

$$\omega^2 C_s^2 k_z'^2 = \omega^4 - \omega^2\left(C_s^2 k_x^2 + \frac{\gamma^2 g^2}{4C_s^2}\right) + (\gamma - 1)g^2 k_x^2 \qquad [10]$$

For a given horizontal wave number, a real solution for k_z' exists only when the right-hand side of eqn [7], which is a second-order polynomial in ω, is positive.

The two roots ω_1 and ω_2 define the following three different cases:

- If $\omega_1 < \omega < \omega_2$, k'_z is purely imaginary. In this case we can only have waves trapped at the surface propagating only horizontally (Lamb waves).
- If $\omega < \omega_1 < (\gamma-1)^{1/2} g/C_s$, k'_z is purely real; this correspond to the internal gravity-waves domain, governed primarily by buoyancy. $\omega_g = (\gamma-1)^{1/2} g/C_s$ is the Brunt–Väisälä frequency.
- If $\omega > \omega_2 > \gamma g/2C_s$, k'_z is purely real; this corresponds to the internal acoustic waves domain, governed primarily by compression. The high-frequency limit of those waves is usual sound wave (with exponential increase in amplitude). $\omega_a = \gamma g/2C_s$ is called the acoustic cut-off frequency. In the short-wavelength limit, however, the root ω_2 varies as $C_s k_x$.

Typically $\omega_a/2\pi = 3.3$ mHz and $\omega_g/2\pi = 2.9$ mHz in the lower atmosphere. This classification in acoustic, gravity, and Lamb wave is widely used in aeronomy. It can be extended to nonisothermal models when the scale height is large in comparison to the vertical wavelength. An adaptation of normal-mode theory to such a nonisothermal model, for the coupled solid Earth–ocean–atmosphere system, has been developed which allows simulation of wave propagation for the whole system (Lognonne *et al.*, 1998). The main difference with traditional seismic normal-mode simulation is in the treatment of the upper-boundary condition to reflect the absence of free surface.

Attenuation of waves arise from viscous and thermal dissipations with comparable magnitudes. But except for short-period acoustic waves, attenuation can be neglected up to 100 km of altitude. This is because dominant atmospheric waves that are coupled to Rayleigh waves have frequency content of about 10–20 mHz.

Stronger limitations of this model arise at high altitude because of two reasons; first, amplification with altitude implies that linearized equations will cease to be valid. Second, upward-propagating internal waves eventually reach the ionosphere, where the dynamics is strongly constrained by the influence of magnetic field on charged particles. The ionospheric response to gravity waves is still an object of an extensive literature (Yeh and Liu, 1972).

4.14.3.2.2 Frequency–wavelength domains of interest

Seismic, tsunami, and atmospheric waves result from the action of gravity and elastic forces. Significant effects from both types of forces exist for the relevant frequency range (a few tens of millihertzs). **Figure 6**

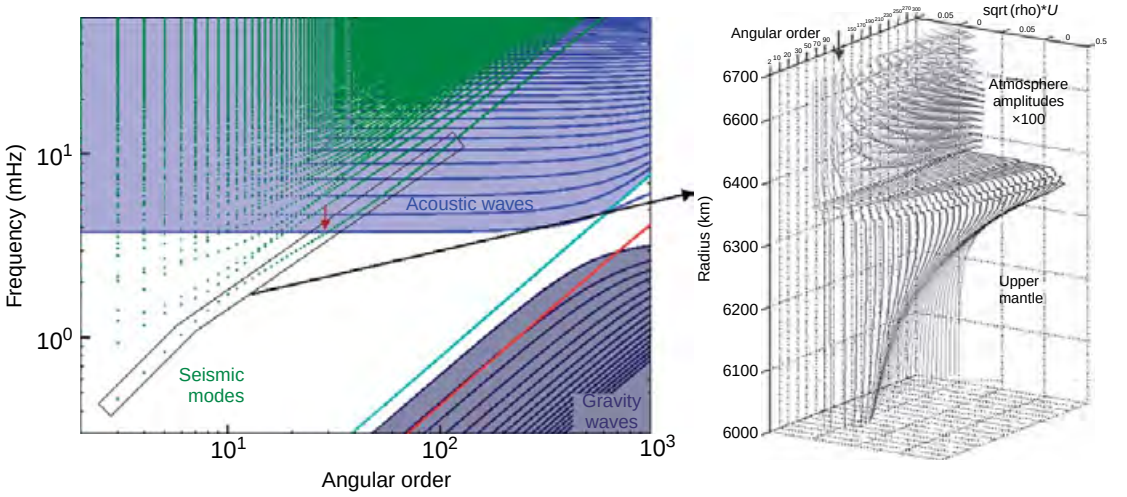


Figure 6 Domains of existence of waves in the solid Earth, ocean, and atmosphere. This figure represents the set of free oscillation modes obtained using standard minos software (Woodhouse, 1988) for a Earth model composed of PREM for solid Earth and US Standard Atmosphere 1976 up to 86 km of altitude. Green dots indicate seismic spheroidal modes; Red dots indicate 'tsunami' mode (corresponding to the 3 km ocean layer in PREM); Blue dots indicate acoustic and gravity mode. As minos software has a free boundary at the top of the model, some obvious errors exists: the atmospheric-mode spectrum is not in general discrete because there is no upper-boundary condition, and furthermore, the lowest acoustic branch shown on this plot would correspond to the nonphysical surface atmospheric mode at the top of the model.

represents a normal-mode representation of the whole Earth–ocean–atmosphere system (Lognonne *et al.*, 1998; Artru *et al.*, 2001). The range of existence of seismic, tsunami, acoustic, and gravity waves can be clearly identified, and the different areas where they coexist (seismic and acoustic, tsunami, and gravity) mark the potential for coupled waves. Outside those area, the dynamic coupling is limited to interface waves (e. g., Lamb wave in the atmosphere).

Study of atmospheric signals caused by sources in the solid Earth is mostly based on observations at the surface (seismograms, barograms, tides gauges, infrasound arrays), or at high altitudes through the ionospheric response to upward-propagating internal waves. The latter measurement can only concern the acoustic or gravity frequency ranges. In particular, seismic and tsunami waves are much more likely to produce strong atmospheric signals at high altitude than many other natural or artificial sources. This is because, despite the very small size of displacements at the surface, they present a unique combination of frequency and horizontal wavelength range necessary for an efficient coupling with internal waves in the atmosphere. On the other hand, major energy from ocean swell, located in the same frequency range, may induce some infrasonic signal trapped at the base of the atmosphere (Garces *et al.*, 2003), but will not in general induce internal (i.e., upward propagating) acoustic waves in the atmosphere, because the wavelength is much shorter than for Rayleigh waves (Arendt and Fritts, 2000).

4.14.3.3 Observation Techniques

4.14.3.3.1 Surface observation: seismometer, microbarograph, hydrophones

Both seismic and pressure sensors have been used to characterize the coupling at the Earth surface. In particular, recent worldwide deployment of infrasound arrays for the International Monitoring System has enabled us to detect atmospheric-pressure fluctuations related to solid Earth activity (**Figure 7**). The frequency band for such instruments is typically between 0.1 and 10 Hz (Mutschlecner and Whitaker, 2005). The noise level depends on the wind conditions, from 0.1 to 10 Pa, but it can be reduced by filtering. Those arrays can also be used to determine the azimuth and velocity of the observed signals, which provide information on their origin. Essentially two types of infrasound signals are observed after earthquakes (LePichon *et al.* 2002):

- Seismic-coupled air waves, essentially a local conversion of seismic-wave vertical motion into sound pressure (Donn and Posmentier, 1964).
- Infrasound waves remotely generated. Those can be generated at the epicenter (Bolt, 1964) or backscattered by mountain ranges (Young and Green, 1982).

In the latter case, infrasounds travel obliquely upward in the atmosphere but are reflected or refracted at various altitudes because of sound velocity variations in the atmosphere (Drob *et al.* 2003). Such a signal arrives later than the coupled air wave, and has

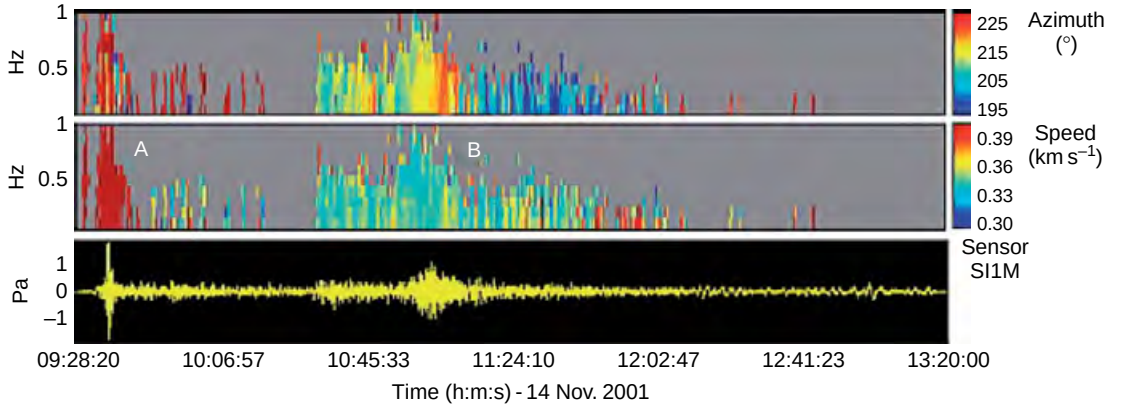


Figure 7 Typical mid-latitude electron density profile in the ionosphere. Profiles were obtained using the International Reference Ionosphere model (Bilitza, 2000). Main ions species (dashed lines are their contribution to ionization) and the plasma frequencies (eqn [11]) corresponding to different electron densities are indicated.

different azimuth and velocity. Mutschlecner and Whitaker (2005) showed that amplitude of the observed pressure perturbations depends not only on the epicentral distance and magnitude of the earthquake but also on the stratospheric wind speed, which has a major impact on the propagation of infrasounds.

4.14.3.3.2 High-altitude observation: ionosondes, Doppler sounding, transmission (GPS), in situ

Detection of acoustic and gravity waves at intermediate altitudes in the neutral atmosphere is not possible due to the absence of *in situ* measurement and because atmospheric remote sensing generally lacks the resolution that would be needed. However, at higher altitudes, in addition to the exponential amplification of the wave amplitude, the interaction with the local plasma leads to perturbations of the ionosphere that are detectable by using radio sounding techniques.

The ionosphere is the intermediate region between the neutral atmosphere and the magnetosphere, ranging approximately between 60 and 1500 km in altitude. It is a stratified medium, partially ionized by solar radiation (due to ultraviolet light and X-rays). The maximum in electron density is reached between 200 and 400 km of altitude, and takes values in the range 10^5–10^7 e^- m^{-3} (**Figure 8**).

Because it is a plasma, the ionosphere has a strong influence on electromagnetic-wave propagation. The

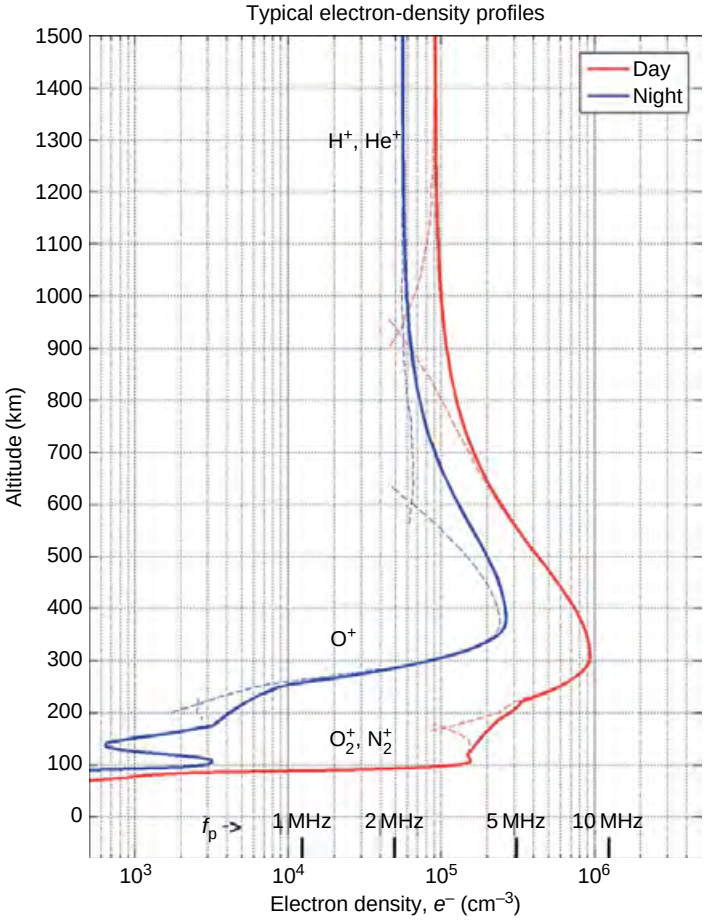


Figure 8 Infrasound record from the I34MN IMS infrasound array (Mongolia) after the 2001 Kunlun earthquake (Northern China, *M*s = 8.1 on 14 November 2001). *Bottom panel*: Atmospheric-pressure fluctuations recorded filtered between 0.05 and 4 Hz. *Top two panels*: The color scales indicate the values of the azimuth and the trace velocity measured at the infrasound station. *Y*-axis corresponds to frequencies from 0.05 to 1 Hz. In the first part of the signal, coherent wave trains referred as Group A are related to local seismically coupled air waves. Due to the coupling at the Earth–air interface, the horizontal trace velocity of the ground-coupled air waves and the seismic waves are identical (greater than 3 km s^{-1}). Group B corresponds to the propagation of infrasonic waves backscattered by the Kunlun mountain range.

plasma frequency is the low frequency cut-off for radio-wave propagation, and depends on the local electron density N_e as

$$\omega_p = \left(\frac{N_e e^2}{m_e \epsilon_0}\right)^{1/2} \quad [11]$$

where e and m_e are the electron charge and mass, and ϵ_0 is the permittivity of vacuum. Typical ionospheric plasma frequencies range from 1 to 20 MHz and their maximum occurs at the maximum of ionization. Ionospheric sounding techniques are based on measurement of either the reflection of radio waves below the maximum plasma frequency, or the refraction delay of higher-frequency signals transmitted across the ionosphere from satellites.

4.14.3.3.2.(i) Reflection For frequencies lower than 10 MHz, waves are reflected by the ionospheric layer whose plasma frequency (eqn [11]) exceeds the frequency of signal. One type of radio sounding of the ionosphere is vertical sounding (ionosonde), measuring the traveltime of a reflected signal as a function of frequency. This provides information on the electron density profile. Doppler sounding is based on the measurement of the Doppler shift of monochromatic signal, sent vertically upward and reflected back to the ground. The frequency shift between the emitted and the reflected wave is directly proportional to the vertical velocity of the reflecting layer. Using such ionospheric sounding network, Blanc (1985) gave a review of natural and artificial sources of signals recorded in those 'ionospheric seismometers'. Both types of sounding can be either ground based, probing the lower ionosphere up to the electron density maximum, or satellite based ('top-side sounding').

4.14.3.3.2.(ii) Transmission For higher frequencies (typically L-band, 1–2 GHz and higher), electromagnetic waves can propagate across the ionosphere. Ionospheric refraction induces a delay in the traveltime, which depends on the frequency of the signal and on the local plasma frequency (eqn [11]), hence on the local electron density along the ray:

$$\delta\rho_{iono} = \pm\frac{e^2}{4\pi^2 m_e \epsilon_0}\int_{ray}\frac{N_e}{f^2}\,dl \quad [12]$$

Using traveltime measurements at two different frequencies gives access to the total electron content (TEC), which is the integral of electron density along a path between a radio transmitter and a receiver. This measurement on TEC has become widely accessible from the development in satellite altimetry and positioning, since it is a key correction required to be applied to those systems. In particular, GPS relies on the accurate measurement of satellites – GPS receiver distances and uses a combination of traveltime measurements at two frequencies (L1 and L2, respectively, $f_1 = 1.575$ and $f_2 = 1.227$ GHz). Once other sources of error are taken into account (Lanyi and Roth, 1988; Sardon *et al.*, 1994), TEC along the ray can be obtained as a linear combination of the two estimated distances ρ_1 and ρ_2:

$$\text{TEC} = \int_{ray} N_e\,dl = \frac{1}{40.3}\frac{f_2^2 - f_1^1}{f_2^2 f_1^2}(\rho_1 - \rho_2) \quad [13]$$

4.14.3.4 Sources in the Solid Earth

Ionospheric perturbations that follow earthquakes have been observed both near the seismic source and at teleseismic distances (**Figure 9**). The first published observations were related to the Great Alaska Earthquake in 1964. Using ionospheric sounding networks, Bolt (1964) observed atmospheric perturbations propagating from the epicenter region and the ionospheric signatures of the Rayleigh-wave propagation at the sounder location.

4.14.3.4.1 Away from the source: surface waves and tsunamis

4.14.3.4.1.(i) Seismic waves Seismic surface waves have typical group velocity in the range 3–4 km s^{-1}, whereas the sound velocity in the atmosphere is much smaller (340 m s^{-1}). Therefore, the acoustic wave is sent almost vertically upward and reaches the ionosphere with a delay of 8–15 min after the passage of Rayleigh waves on the ground. Amplification due to density decrease with altitude can reach 10^4–10^5, and attenuation is only observed for short-period signal above 100 km of altitude. Large-scale vertical oscillations of the ionospheric layers can be easily monitored using the Doppler sounding networks described in Section 4.14.3.3.2 (Yuen *et al.*, 1969; Blanc, 1985). **Figure 10** presents a recent example of such measurements. Although monitoring networks are still sparse, such ionospheric oscillations are currently observed systematically for most $M \geq 6.5$ earthquakes worldwide.

Using a normal-mode approach, Artru *et al.* (2004) have modeled oscillations of ionospheric layers induced by surface Rayleigh waves generated by a

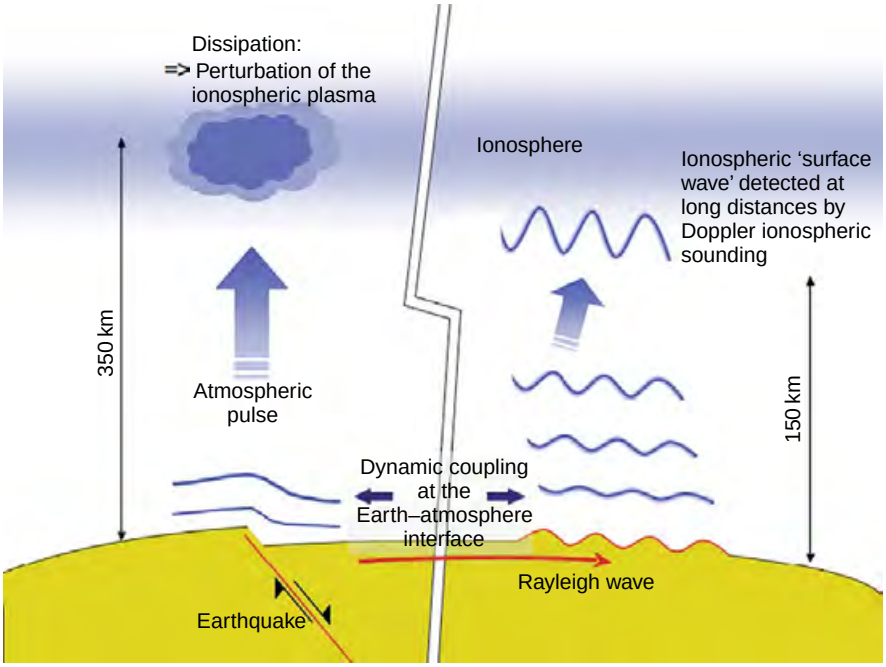


Figure 9 Schematic view of atmospheric waves generated after an earthquake.

major earthquake. Their analysis retrieved source parameters from atmospheric events and showed that atmospheric perturbations generated by surface Rayleigh waves may be amplified by a factor of 10 000–100 000 upon attaining heights between 150 and 250 km.

4.14.3.4.1.(ii) Tsunami waves Tsunami waves are expected to induce a similar type of coupling with the atmosphere; despite their small amplitude compared with ocean swells, they can generate atmospheric gravity waves because of their long wavelengths. A possibility of detecting tsunami by

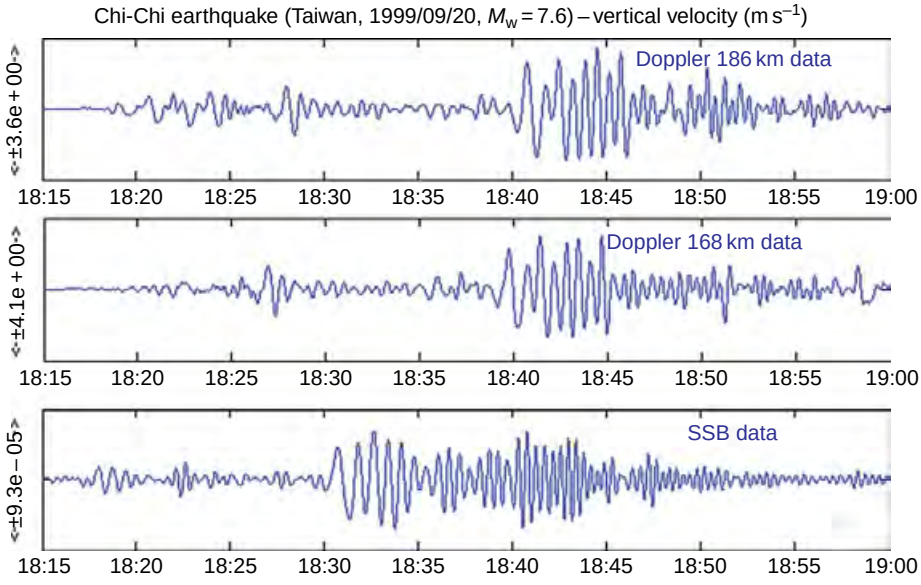


Figure 10 Seismogram and Doppler sounding record (two altitudes: 168 and 186 km) taken in France after Chi-Chi earthquake. Both traces are band-pass filtered between 1 and 50 mHz.

monitoring the ionospheric signature of the induced gravity waves was proposed by Peltier and Hines (1976). They discussed theoretical issues on the coupling and concluded that it is feasible.

Geometry of tsunami–gravity wave coupling is very different from the coupling between seismic and acoustic waves. Tsunami wave is nondispersive, and its velocity depends only on gravity g and water depth d, as the velocity is given by $\sqrt{gd}$. From the gravity-wave dispersion equation, one can estimate group velocity of the induced gravity wave. Horizontally, it appears to be very close to typical tsunami-wave speed, while vertical component is much slower than sound speed (about $50\,\mathrm{m\,s^{-1}}$). Depending on the period, it would therefore take from one to a few hours for the gravity wave to reach the ionosphere (in contrast to about 10 min for seismic–acoustic waves). The ionospheric perturbation should be behind the tsunami front, with a delay increasing with altitude.

Early papers that proposed using ionospheric measurements to detect tsunami-generating earthquakes (Najita *et al.*, 1973; Najita and Yuen, 1979), focused on perturbations induced by Rayleigh waves preceeding a potentially destructive tsunami. Artru *et al.* (2005) used ionospheric sounding from a very dense GPS network in Japan (GEONET) to detect the perturbations associated with the arrival of tsunami wave, generated by the 23 June 2001 Peru eathquake. The observed arrival time, wavelengths, and orientation were shown to fit well with theoretical predictions estimated by a simple simulation (**Figure 11**).

4.14.3.4.2 Direct atmospheric waves

Atmospheric perturbations, observed either at the ground level or in the ionosphere, are often generated away from the measurement location. Propagation of the signal is therefore essentially in the atmosphere. In most cases, the origin of the atmospheric disturbance is the ground motion near an earthquake source (or an underground nuclear explosion). Calais and Minster (1995) detected perturbations in the ionospheric TEC above Southern California after the Northridge earthquake ($M = 6.7$, 17 January 1994) using GPS measurements. Davies and Archambeau (1998) developed a direct modeling of these waves for a simple representation of shallow seismic sources, including high-frequency components of the wave-packet and nonlinear effects. Their results confirmed the seismic origin of the signal, observed after the Northridge earthquake.

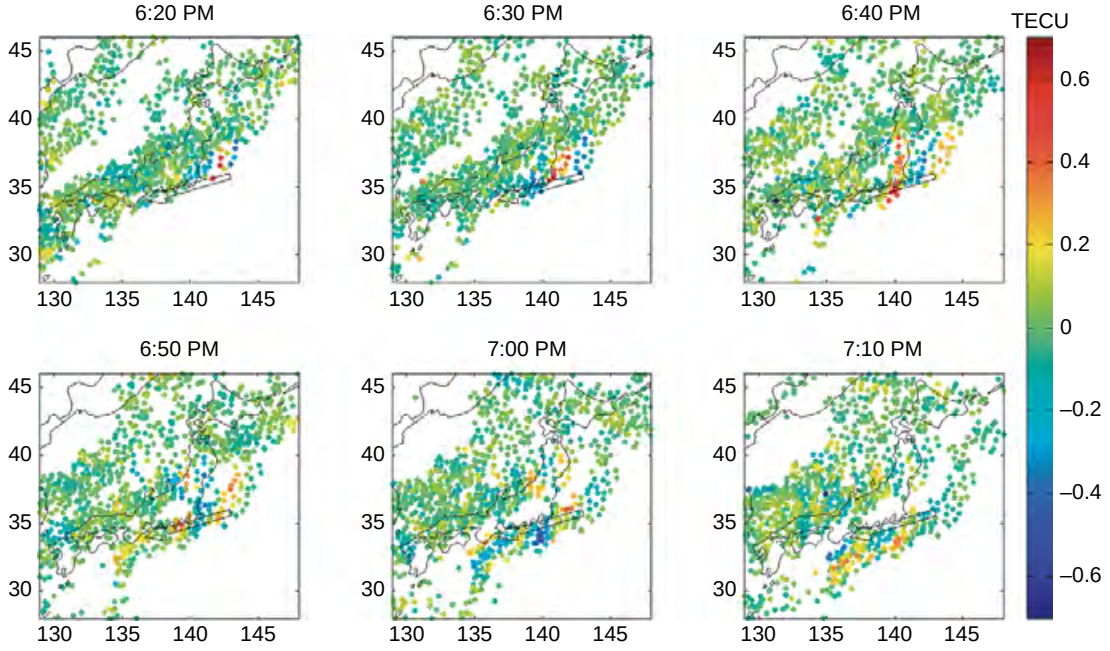


Figure 11 Observed signal for the 23 June 2001 tsunami (initiated offshore Peru): TEC variations plotted at the ionospheric piercing points. A wave-like disturbance is propagating towards the coast of Honshu. This perturbation presents the expected characteristics of a tsunami-induced gravity waves, and arrives approximately at the same time as the tsunami wave itself.

The displacement field generated at the Earth's surface produces a piston-like impulse on the atmosphere. Afraimvich *et al.* (2001) proposed a model for the atmospheric perturbation in the form of 'shock–acoustic waves'. Drob *et al.* (2003), Il-Young *et al.* (2002), and Virieux *et al.* (2004) focused on solving the acoustic-wave propagation in order to model data from CTBT verification network.

4.14.3.4.2.(i) Other sources of infrasounds

In addition to the coupled surface waves and the direct acoustic–shock waves, some atmospheric perturbations related to earthquakes can be observed as scattered waves away from the source. For example, a mountain range can become a source of diffracted infrasound waves at the time of passage of Rayleigh waves (LePichon *et al.*, 2002, 2003; Mutschlecner and Whitaker, 2005).

Other 'solid Earth' sources of atmospheric infrasounds include avalanches and rockfall (Bedard, 2000), volcanic activity (Garces, 1997), and chemical explosions, and mining blasts (Hagerty *et al.*, 2002).

4.14.3.5 The Case of the Great Sumatra–Andaman Earthquake

The great ($M_w = 9.1$) Sumatra–Andaman earthquake and tsunami of 26 December 2004 (Lay *et al.*, 2005) provided a unique opportunity to study solid Earth–ocean–atmosphere coupling. Indeed, by itself this particularly large thrust event would generate not only seismic wave and their associated atmospheric Rayleigh wave, but also direct atmospheric perturbation. In addition it triggered an exceptionally large tsunami that propagated across the Indian Ocean. Satellite altimetry observation could detect the open-ocean wave, with peak-to-peak amplitude of 40–50 cm (Song *et al.*, 2005). Tide gauge stations also provided tsunami data, allowing Tanioka *et al.* (2006) to estimate the rupture process somewhat independently from purely seismic data. The propagation of the tsunami wave across the Indian Ocean lasted for several hours, which was enough, according to the tsunami–gravity-wave coupling theory described earlier, to generate ionospheric signal. Let us review some recent observations published on the subject, although some other works might still be under progress and unpublished at this time.

Le Pichon *et al.* (2005) and Garces *et al.* (2005) analyzed infrasound array data. They observed distinct packets of signal arriving successively: first, the pressure perturbation generated at the sensor location by the seismic waves, with horizontal trace velocity greater than $3\,km\,s^{-1}$, and second, an infrasonic wave train with a mean trace velocity of $0.35\,m\,s^{-1}$ and a dominant period of 10 s, associated with infrasound radiated from the epicenter region. A third sequence consisted of large coherent infrasonic wave, similar in velocity as the previous train, with a dominant period of 30 s, for with back azimuth reconstruction indicate a source area extending from the northern tip of Sumatra to the northern margin of the Bay of Bengal: this suggests that those infrasonic waves were generated by the tsunami wave, either through the interaction of the tsunami with the shoreline, or as the tsunami reaches shallow water and presents shorter wavelength (a few tens of kilometers, comparable to 30 s infrasound waves).

At higher altitudes, ionospheric perturbations related to the Sumatra earthquake and tsunami have been reported as well. Liu *et al.* (2006a) used Doppler sounding network in Taiwan, monitoring the vertical motion of a specific ionospheric layer, and detected two distinct disturbances interpreted as the Rayleigh acoustic wave, then to the direct acoustic–gravity wave emitted by the crustal motion around the earthquake. Other works published were based on continuous GPS data (i.e., the detection of perturbation in the integrated electron content of the ionosphere), although the distribution of permanent receivers in the Indian Ocean region was quite scarce at the time. Heki *et al.* (2006) studied the signals related to the direct acoustic–gravity wave from the source region and used such observation to retrieve information on the rupture process, in particular, the rupture propagation speed. Using sensibly the same GPS stations in Sumatra and Thailand, Otsuka *et al.* (2006) focused on the variations in the TEC perturbation observed between different stations, interpreted as the consequence of the directivity in the ionospheric response with respect to the neutral atmosphere perturbation. DasGupta *et al.* (2006) reported smooth variations in TEC detected by GPS stations located on the East coast of India without giving yet any specific interpretation for their origin. Iyemori *et al.* (2005) reported a rather different kind of observation, using ground-based fluxgate magnetometer. They observe localized, long-period geomagnetic pulsation in Thailand shortly after the origin time of the earthquake and speculated that it was due to the resonant interaction of magnetic field lines with the upward propagating magnetosonic wave emitted from the earthquake area.

Searching for ionospheric disturbances related to the tsunami-wave propagation, two works reported some successful observation: Liu *et al.* (2006b) used GPS data from five permanent receivers in the southern Indian Ocean and detected TIDs with period in the range 10–20 minutes, with horizontal propagation speed consistent with the theory of tsunami–gravity-wave coupling. Occhipinti *et al.* (2006) took advantage of the simultaneous sea surface height and TEC measurements provided by altimetry satellites Jason-1 and TOPEX/Poseidon, hence giving data in the open ocean, away from possible coastal perturbation. In addition, they were able to perform a direct three-dimensional (3D) modeling of the tsunami-generated gravity wave including its interaction with the ionospheric plasma. The simulated and observed perturbations agreed remarkably well, considering the large uncertainties inherent to the combination of data, models, and theory from very different fields of Earth sciences.

4.14.3.6 Sources in the Atmosphere

Atmospheric events that generate infrasounds or gravity waves radiate, in some cases, enough energy to induce seismic signals through a dynamic coupling and the inverse energy flow from the atmosphere to the solid Earth can occur. Two main categories are discussed below: volcanic eruptions and sonic booms. The latter include data from the shuttle, supersonic jets, or meteors that enter into the atmosphere.

4.14.3.6.1 Eruptions

Atmospheric waves produced by the Krakatoa eruption in 1883 were observed worldwide by barometric measurements. The propagation of such atmospheric disturbances is similar to those produced by nuclear explosions, and investigations were made in the 1960s to characterize these waves and their impact on the ionosphere (Row, 1967; Harkrider and Press, 1967). Following the eruption of Mount St. Helens in 1980, Roberts *et al.* (1982) detected a long-lived, large-scale traveling ionospheric disturbance in TEC measurements. Bolt and Tanimoto (1980) reported air waves in barograph records that circled more than once around the globe. The eruption of Mount Pinatubo (Philippines) on 15 June 1991 provided a remarkable example of the interaction between the solid and gaseous envelopes of the Earth system. The energy released in several explosions is estimated to be more than 100 MT (TNT), generating significant atmospheric-pressure waves (see **Figure 6**). Signals related to these waves were observed worldwide in barographs, ionograms, Doppler soundings, TEC measurements, and seismic data, resulting in multitude of analyses for different data. Igarashi *et al.* (1994) used the Japanese ionospheric observation network to determine the characteristics of the gravity wave and the associated traveling ionospheric disturbances. Kanamori *et al.* (1994) investigated the source mechanism of atmospheric oscillations from both barographic and seismographic records. Johnson (2003) gave a review of infrasound observations emitted from volcanic eruptions and pointed out that it is to distinguish regular seismic signals related to subsurface seismicity from the seismicity associated with gas release.

4.14.3.6.2 Sonic boom

Seismic and underwater perturbations induced by aircraft sonic booms have been studied since the mid-1960s (Cook *et al.*, 1972). Frequency range is much higher than those for previously discussed phenomena. It should be noted that these sonic booms are one of the rare examples of controlled sources. The main effects of sonic booms arise in the waveforms of the strain with the shape of the N-wave overpressure, and also in air-coupled Rayleigh wave trains following each N-wave transient. Seismic waves generated by the coupling can then propagate, faster than the original shock wave. More recently, the Concorde (LePichon *et al.*, 2003). and the space shuttle provided further opportunities for studying those effects (Sorrells *et al.*, 2002).Using seismic records of the sonic boom by the space shuttle Columbia, returning to the California Edwards Air Force base, Kanamori *et al.* (1991) observed P-wave pulse across the Los Angeles basin, probably excited through the motion of high-rise buildings in response to the sonic boom. Sonic boom was also shown to induce ionospheric perturbations through the atmosphere–ionosphere coupling using data from GPS measurements (Calais and Minster, 1998).

4.14.4 Conclusion

It has been a traditional practice for seismologists to ignore the atmosphere, the outermost layer of the Earth. To a large extent, this has been justified because of its relatively small effect on seismograms.

On the other hand, because of the existence of the atmospheric layer, we receive some benefits. In addition to understanding the causes of noise, which shed

some light on the mechanism in the atmosphere–ocean–solid Earth interactions, one of the recent benefits has been the development of the correlation technique that allows us to retrieve Green's functions among pairs of stations. We can now obtain Green's functions for the solid Earth from the portions of seismograms that do not contain earthquake signals. This owes to the existence of noise and the development of theory on diffuse wavefield generated by noise.

Extending the analysis to include the atmospheric layer will increase opportunities to detect signals that are potentially useful for quantitative analysis of solid-Earth processes. Volcanic eruption is an obvious process as it directly generates atmospheric waves. Direct tsunami observation is also possible as surface displacement of ocean is not small and there are atmospheric waves that are coupled to this phenomenon. Shallow earthquakes also emit waves into the atmosphere which may become a useful source of information. We have probably scratched only the surface of a large body of useful phenomena in the atmosphere and as we develop the understanding of waves in the whole Earth system, we may discover many things that seismologists have missed in the past 100 years.

References

Afraimvich EL, Perevalova NP, Plotnikov AV, and Uralov AM (2001) The shock-acoustic waves generated by earthquakes. *Annales of Geophysics* 19: 395–409.

Aki K (1957) Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bulletin of the Earthquake Research Institute* 35: 415–456.

Aki K (1965) A note on the use of microseisms in determining the shallow structures of the earth's crust. *Geophysics* 30: 665–666.

Aki K and Richards P (1980) *Quantitative Seismology*. New York: Freeman.

Arendt S and Fritts DC (2000) Acoustic radiation by ocean surface waves. *Journal of Fluid Mechanics* 415: 1–21.

Artru J, Ducic V, Kanamori H, Lognonne P, and Murakami M (2005) Ionospheric detection of gravity waves induced by Tsunamis. *Geophysical Journal International* 160(3): 840–848.

Artru J, Farges T, and Lognonne P (2004) Acoustic waves generated from seismic surface waves: Propagation properties determined from doppler sounding observations and normal-mode modelling. *Geophysical Journal International* 158(3): 1067–1077.

Artru J, Lognonne P, and Blanc E (2001) Normal modes modelling of post-seismic ionospheric oscillations. *Geophysical Research Letters* 28(4): 697–700.

Beauduin R, Lognonne P, Montagner JP, Cacho S, Karczewski JF, and Morand M (1996) The effects of the atmospheric pressure changes on seismic signals or how to improve the quality of a station. *Bulletin of the Seismological Society of America* 86(6): 1760–1769.

Bedard AJ (2000) Atmospheric infrasound. *Physics Today* 53: 32–37.

Ben-Menahem A (1975) Source parameters of the Siberia explosion of June, 1908, from analysis and synthesis of seismic signals at four stations. *Physics of the Earth and Planetary Interiors* 11: 1–35.

Beroza G and Jordan T (1990) Searching for slow and silent earthquakes using free oscillations. *Journal of Geophysical Research* 95: 2485–2510.

Bilitza D (2000) International reference ionosphere 2000. *Radio Science* 36(2): 261–275.

Blanc E (1985) Observations in the upper atmosphere of infrasonic waves from natural or artificial sources: A summary. *Annals of Geophysics* 3(6): 673–688.

Bolt B (1964) Seismic air waves from the great Alaska earthquake. *Nature* 202: 1095–1096.

Bolt BA and Tanimoto T (1980) Atmospheric oscillations after the May 18, 1980 eruption of Mount St. Helens. *EOS* 62(23): 529.

Boy JP, Hinderer J, and Gegout P (1998) Global atmospheric loading and gravity. *Physics of the Earth and Planetary Interiors* 109: 161–177.

Bromirski PD (2001) Vibrations from the 'Perfect Storm'. *Geochemistry Geophysics Geosystems* 2, doi:2000GC000119.

Bromirski PD and Duennebier FK (2002) The near-coastal microseism spectrum: Spatial and temporal wave climate relationships. *Journal of Geophysical Research* 107(B8): 2166 (10.1029/2001JB000265).

Bromirski PD, Flick RE, and Graham N (1999) Ocean wave height determined from inland seismometer data: Implications for investigating wave climate change in the NE Pacific. *Journal of Geophysical Research* 104(C9): 20753–20766.

Calais E, Haase JS, and Minster JB (2003) Detection of ionospheric perturbations using a dense GPS array in Southern California. *Geophysical Research Letters* 30: 1628–1631.

Calais E and Minster B (1995) GPS detection of ionospheric perturbations following the January 17, 1994, Northridge earthquake. *Geophysical Research Letters* 22(9): 1045–1048.

Calais E and Minster B (1998) GPS, earthquakes, the ionosphere, and the Space Shuttle. *Physics of the Earth and Planetary Interiors* 105(3–4): 167–181.

Campillo M and Paul A (2003) Long range correlations in the diffuse seismic coda. *Science* 299: 547–549.

Capon J (1972) Long-period signal processing results for LASA, NORSAR and ALPA. *Geophysical Journal of the Royal Astronomical Society* 31: 279–296.

Cevaloni G (1994) The explosion of the bolide over lugo-di-romagna (Italy) on 19 january 1993. *Planetary and Space Science* 42: 767–775.

Cook JC, Gorforth T, and Cook RK (1972) Seismic and underwater responses to sonic boom. *Journal of the Acoustical Society of America* 51(2): 729–741.

Cooper RIB and Longuet-Higgins MS (1951) An experimental study of the pressure variations in standing water waves. *Proceedings of the Royal Society of London Series A* 206: 424–435.

Crossley DJ, Jensen OG, and Hinderer J (1995) Effective barometric admittance and gravity residuals. *Physics of the Earth and Planetary Interiors* 90: 221–241.

Dahlen AF and Tromp J (1998) *Theoretical Global Seismology*. Princeton: Princeton University Press.

DasGupta A, Das A, Hui D, Bandyopadhyay KK, and Sivaraman MR (2006) Ionospheric perturbations observed by the GPS following the December 26th, 2004 Sumatra-Andaman earthquake. *Earth Planets Space* 58(2): 167–172.

Davies JB and Archambeau CB (1998) Modeling of atmospheric and ionospheric disturbances from shallow seismic sources. *Physics of the Earth and Planetary Interiors* 105: 183–199.

Davies K (1962) The measurement of ionospheric drifts by means of a Doppler shift technique. *Journal of Geophysical Research* 67: 4909–4913.

Davies K and Baker DM (1965) Ionospheric effects observed around the time of the Alaskan earthquake of March 28, 1964. *Journal of Geophysical Research* 70: 1251–1253.

Derode A, Larose E, Tanter M, *et al.* (2003) Recovering the Green's function from field–field correlations in an open scattering medium. *Journal of the Acoustical Society of America* 113: 2973–2976.

Dolenc D, Romanowitz B, Stakes D, McGill P, and Neuhauser D (2005) Observations of infragravity waves at the Montrey ocean bottom broadband station (MOBB). *Geochemistry Geophysics Geosystems* 6: Q09002 (doi:10.1029/2005GC000988).

Donn W and Naini B (1973) Sea wave origin of microbaroms and microseisms. *Journal of Geophysical Research* 78(21): 4482–4488.

Donn W and Posmentier ES (1964) Ground-coupled air waves from the great Alaskan earthquake. *Journal of Geophysical Research* 69: 5357.

Drob DP, Picone JM, and Garces M (2003) Global morphology of infrasound propagation. *Journal of Geophysical Research* 108(D21): 4680.

Duvall TL, Jefferies SM, Harvey JW, and Pomerantz M (1993) *Nature* 362: 430–432.

Dziewonski AM and Anderson DL (1981) Preliminary reference earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

Ekström G (2001) Time domain analysis of Earth's long-period background seismic radiation. *Journal of Geophysical Research* 106(B11): 26483–26493.

Friedrich A, Kruger F, and Klinge K (1998) Ocean-generated microseismic noise located with the Grafenberg array. *Journal of Seismology* 2: 47–64.

Fukao Y, Nishida K, Suda N, Nawa K, and Kobayashi N (2002) A theory of the Earth's background free oscillations. *Journal of Geophysical Research* 107: B9, 2206 (doi:10.1029/2001JB000153).

Garces MA (1997) On the volcanic waveguide. *Journal of Geophysical Research* 102(B10): 22547–22564.

Garces M, Hetzer C, Merrifield M, Willis M, and Aucan J (2003) Observations of surf infrasound in Hawaii. *Geophysical Research Letters* 30(24): 2264–2267 (doi:10.1029/2003GL018614).

Garces M, Caron P, Hetzer C, *et al.* (2005) Deep infrasound from the Sumatra earthquake and tsunami. *Eos Transactions, AGU* 86(35): 317–320.

Goldreich P and Keeley AD (1977) Solar seismology II. The stochastic excitation of the solar P-modes by turbulent convection. *Astrophysical Journal* 212: 243–251.

Grevemeyer I, Herber R, and Essen H-H (2000) Microseismological evidence for a changing wave climate in the northeast Atlantic Ocean. *Nature* 408: 349–352.

Gutenberg B (1958) Two types of microseisms. *Journal of Geophysical Research* 63: 595–597.

Hagerty MT, Kim WY, and Martysevich P (2002) Infrasound detection of large mining blasts in Kazakstan. *Pure and Applied Geophysics* 159: 1063–1079.

Harkrider DG (1964) Theoretical and observed acoustic-gravity waves from explosive sources in the atmosphere. *Journal of Geophysical Research* 69: 5295–5321.

Harkrider D and Press F (1967) The Krakatoa air-sea waves: An example of pulse propagation in coupled systems. *Geophysical Journal of the Royal Astronomical Society* 13: 149–159.

Harkrider D (1964) Theoretical and observed acoustic-gravity waves from explosive sources in the atmosphere. *Journal of Geophysical Research* 69: 5295.

Haubrich RA, Munk W, and Snodgrass FE (1963) Comparative spectra of microseisms and swell. *Bulletin of the Seismological Society of America* 53: 27–37.

Heki K, Otsuka Y, Choosakul N, Hemmakorn N, Komolmis T, and Maruyama T (2006) Detection of ruptures of Andaman fault segments in the 2004 great Sumatra earthquake with coseismic ionospheric disturbances. *Journal of Geophysical Research* 111: B09313 (doi:10.1029/2005JB004202).

Hinderer J and Crossley D (2000) Time variations in gravity and inferences on the Earth's structure and dynamics. *Surveys in Geophysics* 21: 1–45.

Hines CO (1960) Internal atmospheric gravity waves at ionospheric heights. *Canadian Journal of Physics* 38: 1441–1481.

Hines CO (1972) Gravity-waves in atmosphere. *Nature* 239: 73–78.

Igarashi K, Kaminuma S, Nishimuta I, *et al.* (1994) Ionosphere and atmospheric disturbances around Japan caused by the eruption of Mount-Pinatubo on 15 June, 1991. *Journal of Atmospheric and Terrestrial Physics* 56: 1227–1234.

Il-Young C, Jun M-S, Jeon J-S, and Min KD (2002) Analysis of local seismo-acoustic events in the Korean peninsula. *Geophysical Research Letters* 29: 10.1029/2001GL014060.

Ishihara Y, Tsukada S, Sakai S, Hiramatsu Y, and Furumoto M (2003) The 1998 Miyako fireballs trajectory determined from shock wave records of a dense seismic array. *Earth Planets and Space* 55: 9–12.

Iyemori T, Nose M, Han D, *et al.* (2005) Geomagnetic pulsations caused by the Sumatra earthquake on December 26, 2004. *Geophysical Research Letters* 32: L20807 (doi:10.1029/2005GL024083).

Johnson JB (2003) Generation and propagation of infrasonic airwaves from volcanic explosions. *Journal of Volcanology and Geothermal Research* 121: 1–14.

Kanamori H, Mori J, Anderson DL, and Heaton TH (1991) Seismic excitation by the space-shuttle columbia. *Nature* 349: 781–782.

Kanamori H, Mori J, and Harkrider D (1994) Excitation of atmospheric oscillations by volcanic eruptions. *Journal of Geophysical Research* 22: 21947–21961.

Kobayashi N and Nishida K (1998) Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature* 395: 357–360.

Lacoss RT, Kelly EJ, and Toksoz MN (1969) Estimation of seismic noise structure using seismic arrays. *Geophysics* 34(1): 21–38.

Landau LD and Lifshitz EM (1987) *Fluid Mechanics*. Elmsford, New York, USA: Pergamon Press.

Lanyi GE and Roth E (1988) A comparison of mapped and measured total ionospheric electron content using global positioning system and beacon satellite observation. *Radio Science* 23: 483–492.

Lay T, Kanamori H, Ammon CJ, *et al.* (2005) The great Sumatra-Andaman earthquake of 26 December 2004. *Science* 308(5725): 1127–1133 (doi: 10.1126/science.1112250).

Lee AW (1935) On the direction of approach of microseismic waves. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 149(866): 183–199.

LePichon A, Guilbert J, Vega A, Garces M, and Brachet N (2002) Ground-coupled air waves and diffracted infrasound from the Arequipa earthquake of June 23, 2001. *Geophysical Research Letters* 29: 1886–1889.

LePichon A, Guilbert J, Vallee M, Dessa JX, and Ulziibat M (2003) Infrasonic imaging of the Kunlun mountains for the

great 2001 China earthquake. *Geophysical Research Letters* 30: 1814.

Le Pichon A, Herry P, Mialle P, *et al.* (2005) Ceranna, Infrasound associated with 20042005 large Sumatra earthquakes and tsunami. *Geophysical Research Letters* 32: L19802 (doi:10.1029/2005GL023893).

Liu JY, Tsai YB, Chen SW, *et al.* (2006a) Giant ionospheric disturbances excited by the M9.3 Sumatra earthquake of 26 December 2004. *Geophysical Research Letters* 33: L02103 (doi:10.1029/2005GL023963).

Liu J-Y, Tsai Y-B, Ma K-F, *et al.* (2006b) Ionospheric GPS total electron content (TEC) disturbances triggered by the 26 December 2004 Indian Ocean tsunami. *Geophysical Research Letters* 111: A05303 (doi:10.1029/2005JA011200).

Lobkis O and Weaver R (2001) On the emergence of the Green's function in the correlations of a diffuse field. *Journal of the Acoustical Society of America* 110: 3011–3017.

Lognonne P, Clevede E, and Kanamori H (1998) Computation of seismograms and atmospheric oscillations by normal-mode summation for a spherical earth model with realistic atmosphere. *Geophysical Journal International* 135: 388–406.

Longuet-Higgins MS (1950) A theory of the origin of microseisms. *Philosophical Transactions of the Royal Society of London Series A* 243: 1–35.

Mannucci AJ, Wilson BD, Yuan DN, Ho CH, Lindqwister UJ, and Runge TF (1998) A global mapping technique for GPS-derived ionoispheric electron content measurements. *Radio Science* 33: 565–582.

Merriam JB (1992) Atmospheric pressure and gravity. *Geophysical Journal of the Royal Astronomical Society* 109: 488–500.

Müller T and Zürn W (1983) Observation of gravity changes during the passage of cold fronts. *Journal of Geophys* 53: 155–162.

Mutschlecner JP and Whitaker RW (2005) Infrasound from earthquakes. *Journal of Geophysical Research* 110(D1): D01108.

Najita K, Weaver PF, and Yuen PC (1973) A tsunami warning system using an ionospheric techniique. *Proceedings of the IEEE* 62: 563–567.

Najita K and Yuen PC (1979) Long-period oceanic Rayleigh wave group velocity dispersion curve from the doppler sounding of the ionosphere. *Journal of Geophysical Research* 84: 1253–1260.

Nawa K, Suda N, Fukao Y, Sato T, Aoyama Y, and Shibuya K (1998) Incessant excitation of the Earth's free oscillations. *Earth Planets Space* 50: 3–8.

Nawa K, Suda N, Fukao Y, *et al.* (2000) Incessant excitation of the Earth's free oscillations: global comparison of superconducting gravimeter records. *Physics of the Earth and Planetary Interiors* 120: 289–297.

Niebauer TM (1998) Correcting gravity measurements for the effects of local air pressure. *Journal of Geophysical Research* 93(B7): 7989–7991.

Nishida K, Kobayashi N, and Fukao Y (2000) Resonant oscillations between the solid earth and the atmosphere. *Science* 287: 2244–2246.

Nishida K, Fukao Y, Watada S, *et al.* (2005) Array observation of background atmospheric waves in the seismic band from 1 mHz to 0.5 Hz. *Geophysical Journal International* 162: 824–840, doi:10.1111/j.1365-246X.2005.02677.x.

Nishida K, Kobayashi N, and Fukao Y (2002) Origin of Earth's ground noise from 2 to 20 mHz. *Geophysical Research Letters* 29(10): 10.1029/2001/GL013862.

Occhipinti G, Lognonne P, Kherani EA, and Hebert H (2006) Three-dimensional waveform modeling of ionospheric signature induced by the 2004 Sumatra tsunami. *Geophysical Research Letters* 33: L20104 doi:10.1029/2006GL026865.

Okada H (2003) *The Microtremor Survey Method,* Geophysical Monograph Series, Tulsa, OK: Society of Exploration Geophysicists, No. 12.

Okihiro M, Guza R, and Seymour R (1992) Bound infragravity waves. *Geophysical Research Letters* 97(C7): 11453–11469.

Otsuka Y, Kotake N, Tsugawa T, *et al.* (2006) GPS detection of total electron content variations over Indonesia and Thailand following the 26 December 2004 earthquake. *Earth Planets Space* 58(2): 159–165.

Peltier WR and Hines CO (1976) Possible detection of tsunamis by a monitoring of ionosphere. *Journal of Geophysical Research* 81: 1995–2000.

Peterson J (1993). Observation and modeling of background seismic noise. US Geological Survey *Open-File Report* 93–322. Albuquerque, Mexico: US Geological Survey.

Phillips OM (1977) *The Dynamics of the Upper Ocean*. Cambridge: Cambridge University Press.

Posmentier ES (1967) A theory of microbaroms. *Geophysical Journal of the Royal Astronomical Society* 13: 487–501.

Rabbel W and Zschau J (1985) Static deformations and gravity changes at the Earth's surface due to atmospheric loading. *Journal of Geophysics* 56: 81–99.

Rhie J and Romanowitz B (2004) Excitation of Earth's continuous free oscillations by atmosphere-ocean-seafloor coupling. *Nature* 431: 552–556.

Rickett J and Claerbout J (1999) Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring. *The Leading Edge* 18: 957–960.

Roberts DH, Klobuchar JA, Fougere PF, and Hendrickson DH (1982) A large-amplitude travelling ionospheric disturbance produced by the May 18, 1980, explosion of Mount St. Helens. *Journal of Geophysical Research* 87: 6291–6301.

Row RV (1967) Acoustic-gravity waves in the upper atmosphere due to a nuclear detonation and an earthquake. *Journal of Geophysical Research* 72: 1599–1610.

Roult G and Crawford W (2000) Analysis of 'background' free oscillations and how to improve resolution by substracting the atmospheric pressure signal. *Physics of the Earth and Planetary Interiors* 121(3–4): 325–338.

Roux P, Sabra KG, Gerstoft P, Kupperman WA, and Roux A (2005a) Ambient noise cross correlation in free space: Theoretical approach. *Journal of the Acoustical Society of America* 117: 79–84 (doi:10.1121/1.1830673).

Roux P, Sabra KG, Gerstoft P, and Kupperman WA (2005b) P-waves from cross-correlation of seismic noise. *Geophysical Research Letters* 32: L19303 (doi:10.1029/2005GL023803).

Sabra KG, Gerstoft P, Roux P, Kupperman WA, and Fehler M (2005a) Surface wave tomography from microseisms in Southern California. *Geophysical Research Letters* 32: L14311 (doi:10.1029/2005GL02315).

Sabra KG, Gerstoft P, Roux P, and Kupperman WA (2005b) Extracting time-domain Green's function estimates from ambient seismic noise. *Geophysical Research Letters* 32: L03310 (doi:10.1029/2004GL021862).

Saito M (1988) Disper80: A subroutine package for the calculation of seismic normal mode solutions. In: Doornbos DJ (ed.) *Seismological Algorithms*. San Diego: Academic Press.

Santo T (1960) Investigations into microseisms using the observational data of many stations in Japan (Part I) – on the origin of microseisms. *Bulletin of the Earth Research Institute* 37: 307–325.

Sardon E, Rius A, and Zarraoa N (1994) Estimation of the transmitter and receiver differential biases and the ionospheric total electron-content from Global Positioning System observations. *Radio Science* 29: 577–586.

Schulte-Pelkum V, Earle P, and Vernon F (2004) Strong directivity of ocean-generated seismic noise. *Geochemistry Geophysics Geosystems* 5: Q03004, doi:10.1029/2003GC000520.

Shapiro N and Campillo M (2004) Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters* 31: L07614 (doi10.1029/2004GL019491).

Shapiro N, Campillo M, Stehley L, and Ritzwoller M (2005) High-resolution surface-wave tomography from ambient seismic noise. *Science* 307: 1615–1618.

Snieder R (2004) Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Physical Review E* 69: 046610.

Song YT, Ji C, Fu L-L, *et al.* (2005) The 26 December 2004 tsunami source estimated from satellite radar altimetry and seismic waves. *Geophysical Research Letters* 32: L20601 (doi:10.1029/2005GL023683).

Sorrells GG (1971) A preliminary investigation into the relationship between long-period noise and local fluctuations in the atmospheric pressure field. *Geophysical Journal of the Royal Astronomical Society* 26: 71–82.

Sorrells G, McDonald J, Der Z, and Herrin E (1971) Earth motion caused by local atmospheric pressure changes. *Geophysical Journal of the Royal Astronomical Society* 26: 83–98.

Sorrells G, Bonner J, and Herrin ET (2002) Seismic precursors to space shuttle shock fronts. *Pure and Applied Geophysics* 159: 1153–1181.

Spratt RS (1982) Modelling the effect of atmospheric pressure variations on gravity. *Geophysical Journal of the Royal Astronomical Society* 71: 173–186.

Suda N, Nawa K, and Fukao Y (1998) Earth's background free oscillations. *Science* 279: 2089–2091.

Sutton G and Barstow N (1996) Ocean bottom microseisms from a distant supertyphoon. *Geophysical Research Letters* 23(5): 499–502.

Takeuchi H and Saito M (1972) Seismic surface waves. In: Bolt BA (ed.) *Methods in Computational Physics*, vol. 11, pp. 217–295. New York: Academic Press.

Tanimoto T (1999) Excitation of normal modes by atmospheric turbulence: source of long-period seismic noise. *Geophysical Journal International* 136: 395–402.

Tanimoto T (2001) Continuous free oscillations: Atmosphere-solid Earth coupling. *Annual Review of Earth and Planetary Sciences* 29: 563–584.

Tanimoto T (2005) The oceanic excitation hypothesis for the continuous oscillations of the Earth. *Geophysical Journal International* 160(1): 10.1111/j.1365-246X.2005.02491.x.

Tanimoto T (2007) Excitation of normal modes by nonlinear interaction of ocean waves. *Geophysical Journal International* 168: 571–582, doi:10.1111/j.1365-246X.2006.03240.x.

Tanimoto T and Um J (1999) Cause of continuous oscillations of the Earth. *Journal of Geophysical Research* 104(B12): 28273–28739.

Tanimoto T, Um J, Nishida K, and Kobayashi N (1998) Earth's continuous oscillations observed on seismically quiet days. *Geophysical Research Letters* 25(10): 1553–1556.

Tanioka Y, Yudhicara, Kususose T, *et al.* (2006) Rupture process of the 2004 great Sumatra–Andaman earthquake estimated from tsunami waveforms. *Earth Plants Space* 58: 203–209.

Tennekes H and Lumley JL (1972) *A First Course in Turbulence*. Cambridge, MA, USA: MIT Press.

Van Dam TM and Wahr JM (1987) Displacement of the Earth's surface due to atmospheric loading: Effects on gravity and baseline measurements. *Journal of Geophysical Research* 92(B2): 1281–1286.

van Tiggelen BA (2003) Green function retrieval and time reversal in a disordered world. *Physical Review Letters* 91: 243904.

Virieux J, Garnier N, Blanc E, and Dessa JX (2004) Paraxial ray tracing for atmospheric wave propagation. *Geophysical Research Letters* 31: L20106.

Warburton RJ and Goodkind JM (1977) The influence of barometric-pressure variations on gravity. *Geophysical Journal of the Royal Astronomical Society* 48: 281–292.

Watada S (1995) *Near-Source Acoustic Coupling between the Atmosphere and the Solid Earth during Volcanic Eruptions*. PhD Thesis, California Institute of Technology, Pasadena.

Watada S and Masters G (2001) Oceanic excitation of the continuous oscillations of the Earth. *Eos Transactions AGU* 82(47): SS2A–0620.

Weaver RL and Lobkis OI (2000) Temperature dependence of diffuse field phase. *Ultrasonics* 38: 491–494.

Weaver R and Lobkis O (2001) Ultrasonics without a source. Thermal fluctuation correlations at MHz frequencies. *Physical Review Letters* 87: 134301.

Webb S (1998) Broadband seismology and noise under the ocean. *Reviews of Geophysics* 36: 105–142.

Webb S, Zhang X, and Crawford W (1991) Infragravity waves in the deep ocean. *Journal of Geophysical Research* 96(C2): 2723–2736.

Whipple FJW (1930) The great Siberian meteor and the waves, seismic and aerial, which it produced. *Royal Meterological Society Quarterly Journal* 56: 287–304.

Widmer-Schnidrig R (2003) What can superconducting gravimeters contribute to normal-mode seismology? *Bulletin of the Seismological Society of America* 93(3): 1370–1380.

Wielandt E and Steim JM (1986) A digital very-broad-band seismograph. *Annales of Geophysics* 4B: 227–232.

Woodhouse JH (1988) The calculation of the eigenfrequencies and eigenfunctions of the free oscillations of the earth and the sun. In: Doornbos DJ (ed.) *Seismological Algorithms: Computational Methods and Computer Programs*. London: Academic Press.

Yeh KC and Liu CH (1972) *Theory of Ionospheric Waves*. New York: Academic Press.

Young JM and Green GE (1982) Anomalous infrasound generated by the Alaskan earthquake of 28 March 1964. *Journal of the Acoustical Society of America* 71: 334–339.

Yuen PC, Weaver PF, Suzuki RK, and Furumoto AS (1969) Continous travelling coupling between seismic waves and the ionosphere evident in May 1968 japan earthquake data. *Geophysical Research Letters* 74: 2256–2264.

Müller T and Zürn W (1983) Observation of gravity changes during the passage of cold fronts. *Journal of Geophysics* 53: 155–162.

Zürn W and Widmer R (1995) On noise reduction in vertical seismic records below 2 mHz using local barometric pressure. *Geophysical Research Letters* 22(24): 3537–5340.

4.15 Episodic Aseismic Slip at Plate Boundaries

S. Y. Schwartz, University of California Santa Cruz, Santa Cruz, CA, USA

4.15.1 Seismic and Aseismic Slip at Plate Boundaries

Most plate boundaries accommodate far-field motions on localized zones of deformation through aseismic creep at depth and earthquake failure in the upper crust. Earthquake slip and the transition with depth from seismic to aseismic deformation can be understood in terms of the frictional response of materials. Realistic faults have a complicated dependence of friction on slip velocity, time, and slip distance known as rate- and state-dependent friction. If frictional resistance to fault movement decreases faster than stress dissipation due to movement, sliding occurs in sudden slips with associated stress drops (earthquakes), punctuated by periods of no motion as stress recharges. This motion is commonly referred to as stick-slip and the frictional behavior necessary for it as velocity weakening. If this condition is not met and fault strength does not decrease with slip, stable sliding occurs. The frictional condition for stable sliding is known as velocity strengthening. As temperature increases with depth, a transition from velocity-weakening to velocity-strengthening frictional behavior is believed to be responsible for the termination of interplate earthquakes and accommodation of plate motion through stable sliding or creep. The region of the plate boundary exhibiting velocity-weakening behavior and capable of generating earthquakes is often referred to as the seismogenic zone (**Figure 1**). The seismogenic zone may possess a second frictional transition to velocity strengthening (stable sliding) at its shallow up-dip edge and although earthquakes do not nucleate outside of the seismogenic zone, seismic rupture propagation has been observed to extend into both the up- and down-dip stable-sliding frictional regimes. The frictional property of these regions of the fault that can propagate but not nucleate seismic rupture is known as conditional stability.

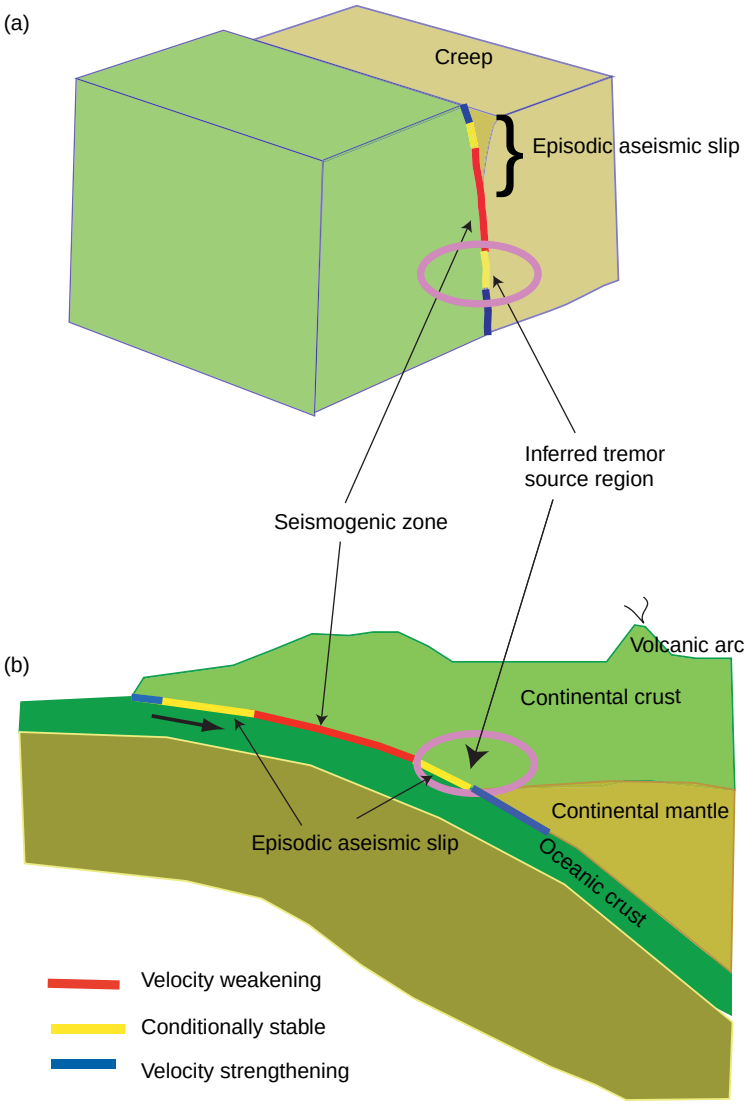


Figure 1 Cartoons of the seismogenic zone at (a) a vertical strike-slip fault and (b) a convergent plate boundary with frictional properties and regions hosting episodic aseismic slip labeled. Earthquakes nucleate in velocity-weakening materials, are capable of propagating into conditionally stable but not velocity-strengthening regions. EAS associated with strike-slip earthquakes is confined to shallow depth but occurs at both the down-and up-dip edge of the seismogenic zone at convergent margins. Seismic tremor appears to locate at comparable depths (below the seismogenic zone) in both environments.

On timescales of years, strain accumulates at plate-boundary seismogenic zones as stress from plate motions builds. When the stress reaches a failure threshold, strain release occurs via large earthquakes. However, several continental strike-slip faults, such as the San Andreas Fault, exhibit segments that release strain fairly continuously or creep both at the surface (Galehouse, 2002) and at seismogenic depths (Wesson, 1988; Schmidt *et al.*, 2005). Creeping segments were first recognized by offsets of cultural features and an abundance of microseismicity in the absence of sizeable earthquakes. Although the microseismicity may occur at a high rate, it does not contribute significantly to total fault slip and the majority of the strain in these fault segments is released through aseismic motions. Early instrumental measurements of aseismic motion using conventional geodetic techniques displayed

primarily linear slip with time (e.g., Galehouse and Lienkaemper, 2003). Indeed determinations of aseismic deformation rates are rare in general, due to the high detection threshold required to measure such slow processes, resulting in the widely held belief that aseismic motions occur at steady linear rates. Fortunately, the recent acquisition and integration of continuous global positioning system (CGPS), interferometric syththeic aperature radar (InSAR), borehole strainmeter, and creepmeter data at plate boundaries has improved the precision and long-term stability of deformation measurements revealing significant fluctuations in aseismic deformation rates. These new data have allowed the distinction of several types of aseismic slip processes at plate boundaries including (1) steady-state creep, (2) episodic slip events, (3) afterslip, (4) triggered slip, and (5) precursory slip, each having important implications for the earthquake process. Aseismic episodic slip events are recognized as periods of accelerated slip lasting between a few hours to a few years and have been documented to occur at convergent and strike-slip margins and continental faulting environments. Afterslip occurs as rapid acceleration of aseismic slip following a mainshock and is thought to be induced by unrelieved stresses within velocity strengthening or conditionally stable regions of the fault (Marone *et al.*, 1991). The concentration of afterslip at shallow depth and its logarithmic decay with time distinguish it from postseismic, viscoelastic relaxation where motion, or flow, is typically distributed deeper in the lower crust and upper mantle and continues for years to tens of years, decaying exponentially. Like afterslip, triggered aseismic slip also involves acceleration in slip rate following an earthquake, but it is produced by either dynamic stresses generated during the passage of large-amplitude surface waves or static stresses from the occurrence of nearby earthquakes on adjacent faults. Precursory slip, although controversial, has been suggested based on observations of accelerated slip, lasting hundreds of seconds to many years, prior to several earthquakes in a variety of tectonic settings.

Like aseismic motion, earthquakes display variations in their slip rate, having an average value of meters per second, as much as 10 orders of magnitude faster than plate motions and 4–9 orders of magnitude faster than episodic aseismic slip (EAS) rates. **Figure 2** shows a plot of characteristic slip distance versus slip duration for seismic and aseismic deformational processes. It shows a continuum of slip velocities with normal and slow earthquakes, both

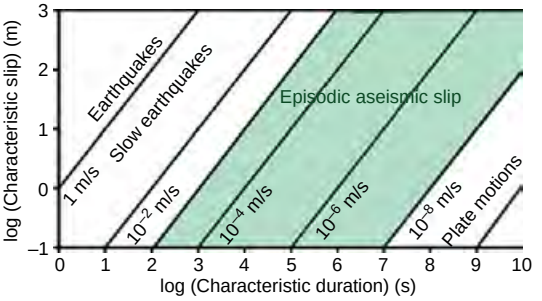


Figure 2 Plot of characteristic slip magnitude versus slip duration for various deformational processes. Normal and slow earthquakes both radiate seismic energy and occur at the fast end while plate tectonic motions occur at the slow end of the velocity spectrum. Most episodic aseismic slip (shaded region) occurs at intermediate velocities. Modified from Beroza GC and Jordan T (1990) Searching for slow and silent earthquakes. *Journal of Geophysical Research* 95: 2485–2510.

radiating seismic energy, at the fast end, and steady-state creep and plate motions at the slow end. The existence of slow earthquakes was postulated based on analysis of anomalous free oscillations of the Earth (Beroza and Jordan, 1990), anomalous excitation of long-period surface waves (Kanamori and Cipar, 1974; Cifuentes and Silver, 1989), weak seismic radiation at high frequencies or anomalous low-frequency radiation (Newman and Okal, 1998; Kanamori and Stewart, 1976; Perez-Campos *et al.*, 2003), or strong tsunami relative to seismic excitation (Pelayo and Wiens, 1992). Some examples of slow earthquakes include a predecessor to the great 1960 Chilean earthquake (Kanamori and Cipar, 1974; Linde and Silver, 1989), a successor to the 1978 Izu-Oshima, Japan earthquake (Sacks *et al.*, 1981), and the 1992 Nicaragua tsunami earthquake (Kanamori and Kikuchi, 1993). The relative slowness of these events has been attributed to unusual material properties along the fault interface (e.g., Bilek *et al.*, 2004) such that a large amount of energy is dissipated during deformation, making less available for seismic radiation.

EAS refers to deformational processes with rates intermediate between earthquakes and plate motions (**Figure 2**). This mode of strain release was first recognized from dilatational strain records in Japan (Sacks *et al.*, 1971) and has since been documented to occur in a variety of forms in almost all tectonic environments. These observations are allowing a much more complete understanding of strain accumulation and release at plate boundaries, with the rate of fault loading and

unloading varying in both space and time. Seismic slip at most plate boundaries accounts for only a fraction of plate tectonic motions so the inclusion of EAS can provide a better quantification of the seismic-moment budget. Earthquake slip on one fault segment changes the stress on adjacent segments and may accelerate or decelerate seismic activity. This abrupt stress change may also initiate other modes of strain release, such as EAS. Conversely, EAS also perturbs the surrounding stress field and may either increase or relieve stress on an adjacent fault segment bringing it closer to or farther from earthquake failure, respectively. Therefore, an improved understanding of EAS and its ability to trigger earthquakes is important for seismic hazard assessment.

This chapter uses distinct nomenclature to distinguish different manifestations of EAS, based on when it occurs in the earthquake cycle (i.e., preslip, interseismic slip, and afterslip), and whether it is better characterized as an intermediate-term change in deformation rate (episodic) or a shorter-term earthquake-like occurrence (event).

4.15.2 Modes of EAS

4.15.2.1 Afterslip

Aseismic slip following large earthquakes, or afterslip, was first documented for the 1966 Parkfield earthquake along the San Andreas Fault (Smith and Wyss, 1968). Since then it has been detected at many other strike-slip (Williams and Magistrale, 1989; Bucknam *et al.*, 1978) and subduction-zone thrust faults. Characteristics of afterslip include (1) location on the coseismic fault plane; (2) slip duration of days to months with a rapid initial rate of slip followed by a logarithmic decay; and (3) cumulative values that are a significant fraction of the average coseismic slip. Marone *et al.* (1991) proposed a model for earthquake afterslip where fault kinematics are prescribed by rate- and state-dependent friction laws. Unrelieved stresses on the fault induce aseismic slip after the earthquake within velocity-strengthening (or conditionally stable) regions of the fault. Their particular model consisted of a thin layer with velocity-strengthening properties, such as fault gouge, overlying a much thicker, velocity-weakening layer. Coseismic rupture initiates and propagates in the lower layer, is arrested by encountering the shallow velocity-strengthening layer, resulting in a stress concentration at the frictional transition that drives afterslip in the shallow layer. The existence of a slip deficit at shallow depth is required to drive the afterslip. Therefore, in their model, if coseismic rupture propagates to the surface, relaxing stress in the velocity-strengthening layer, no afterslip is predicted. In fact, such a phenomenon was observed for the 1992 Landers, California earthquake where significant coseismic slip extended to the surface (Wald and Heaton, 1994) and negligible afterslip was detected (Sylvester, 1993). Marone *et al*'s (1991) model is very well suited to vertical strike-slip faults that develop thick velocity-strengthening gouge layers as they mature. It has successfully reproduced the afterslip time histories for the 1966 Parkfield and 1987 Superstition Hills, California earthquakes and yielded estimates of the frictional transition depths and other important rheological properties of these fault zones from the logarithmic decay of the afterslip (Marone *et al.*, 1991).

With the increase in GPS monitoring of plate boundaries, numerous observations of afterslip have been made at shallow subduction-zone thrust faults (Pritchard and Simons, 2006, table 1). In this tectonic environment, afterslip is most commonly reported at the down-dip edge of the seismogenic zone (Hutton *et al.*, 2001; Melbourne *et al.*, 2002; Heki *et al.*, 1997; Yagi *et al.*, 2003; Bürgmann *et al.*, 2001). This is likely due to the location of most GPS stations on land, directly above the deep plate interface. These network configurations provide little resolution of postseismic deformation occurring offshore. An exception to this is GPS observations following the M_w 8.7 2005 Nias, Indonesia earthquake, which were made close enough to the trench to allow resolution of shallow afterslip. Eleven months of afterslip, equivalent to an M_w 8.2 earthquake, occurred both up- and down-dip of the mainshock asperity (Kreemer *et al.*, 2006; Hsu *et al.*, 2006). Although substantial afterslip was detected following the M_w 9.2 2004 Sumatra–Andaman earthquake, the lack of near-field deformation data made determination of specific locations of afterslip difficult. The best assessments place afterslip at shallow depth, occurring close to or possibly overlapping regions of coseismic slip and with larger displacements occurring in the northern section of the rupture, where coseismic slip was relatively small (Subarya *et al.*, 2006; Vigny *et al.*, 2006). Afterslip following the M_w 8.0 2003 Tokachi-Oki earthquake occurred within the same depth range as the coseismic slip (within the seismogenic zone), but at unique positions along strike, adjacent to but not overlapping the mainshock asperities (Matsubara *et al.*, 2005). In other words,

afterslip appears to fill in regions of slip deficit left by mainshocks as predicted by state- and rate-dependent frictional models (Marone *et al.*, 1991).

Afterslip is important to document and understand because it often accounts for as much slip as occurs in the mainshock it follows. This large amount of afterslip may help explain the observation that seismic moment at many plate boundaries is far less than the total moment expected if all plate motion is accommodated seismically. Studies of afterslip are also important to test rate- and state-dependent friction laws. Accurate estimation of the decay time of afterslip, as well as rigorous numerical tests of whether the observed decay is in fact logarithmic, and remains so over the entire observing period, will allow rate and state friction laws to be rigorously tested with field observations.

4.15.2.2 Preslip

Numerical and laboratory models of fault failure exhibit accelerating deformation prior to dynamic slip instabilities (e.g., Dieterich, 1979). However, observations of aseismic deformation-rate changes prior to earthquakes are quite rare. Roeloffs (2006) examined available data on aseismic deformation-rate changes prior to earthquakes and reported 10 credible accounts of accelerated slip, lasting a few minutes to more than 10 years, preceding earthquakes ranging in magnitude between 3.5 and 9.2. Roeloffs (2006) also documented the absence of such anomalies for seven well-instrumented events with magnitudes between 6.0 and 8.4 and estimated the maximum equivalent preslip moment that may have been released and gone undetected by existing geodetic instrumentation. Anomalous, slow motion immediately before great subduction-zone thrust earthquakes was suggested for the M_w 7.9 1944 Tonankai and M_w 8.0 1946 Nankaido, Japan (Linde and Sacks (2002) and references therein), the M_w 9.5 1960 Chilean (Kanamori and Cipar, 1974; Cifuentes and Silver, 1989), the M_w 9.2 1964 Alaskan (Hamilton and Shennan, 2005), and the M 9 1700 Cascadia (Shennan *et al.*, 1998) events. For these great earthquakes, aseismic preslip was suggested in order to account for slow deformation detected on leveling, tide gauge, and water-well level data and from microfossil analysis, but details of the slow-slip time histories were not determined. The relevant deformation data for the 1944 and 1946 events in southwest Japan were satisfied with 2 m of slow slip on the plate interface, down-dip of the coseismic slip. Preslip at the down-dip extension of the seismogenic zone was also proposed for the smaller M_w 7.7 1983 Japan Sea thrust earthquake, based on a borehole strainmeter recording of ~100 aseismic strain transient events (Linde *et al.*, 1988). More recently, precursory slip before the largest aftershock of the M_w 8.4 2001 Peru earthquake was detected on a continuous GPS station in Arequipa, Peru and shown to be consistent with slow slip along the plate interface (Melbourne and Webb, 2002). The lack of preslip detected before the mainshock constrains allowable preslip moment to be no greater than an M 7.6 earthquake. No preseismic deformation was detected prior to the M_w 8.3 2003 Tokachi-Oki earthquake, the first great subduction-zone thrust event to occur in Japan since installation of GPS Earth Observing Network (GEONET), a dense countrywide continuous GPS network. The minimum preslip detectable by this network was estimated to be equivalent to a M 7 earthquake (Roeloffs, 2006).

The absence of preseismic deformation-rate changes prior to the M_w 6.9 1989 Loma Prieta, CA; the M_w 7.3 1992 Landers, CA; the M_w 7.1 1999 Hector Mines, CA; the M_w 6.0 2004 Parkfield, CA; and M_w 7.6 1999 Chi Chi, Taiwan earthquakes, which were recorded by borehole strainmeters, dilatometers, and CGPS at various distances from the epicenter, constrains possible preslip moment to be less than equivalent magnitude 5.4, 4.8, 5, 3.2, and 6 earthquakes, respectively (Roeloffs, 2006). The paucity of positive preslip observations may in part be due to the existence of signals below the present detection thresholds. The expansion of continuous GPS networks, borehole strainmeters, and InSAR observations at plate boundaries should greatly improve the opportunity to detect and study preslip if and when it occurs.

4.15.2.3 Interseismic Slip Events

The interseismic pattern of surface deformation in the vicinity of active faults is generally attributed to strain accumulation on a locked shallow portion of the fault overlying a deeper continuously slipping region. Therefore, geodetic data collected between earthquakes can be used to infer the locking depth of faults as well as the pattern of strain accumulation on the locked portions (e.g., Bürgmann *et al.*, 2000; Norabuena *et al.*, 2004). Locked strike-slip faults produce a rather sharp decrease in horizontal surface velocity, reaching zero, as the fault is approached (**Figure 3(a)**, top). At convergent margins, where the

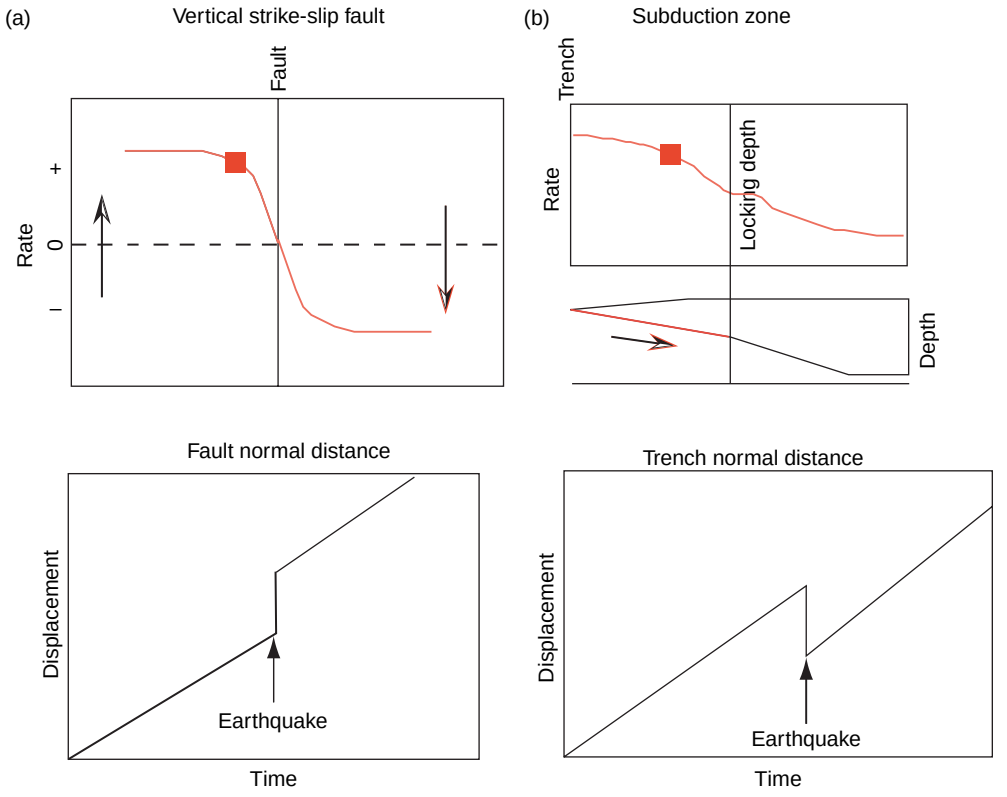


Figure 3 *Top row*: Schematic geodetic strain-accumulation patterns across (a) a vertical strike-slip fault and (b) a fully coupled subduction zone interface. The position of the fault and the horizontal projection of the depth of the transition from a locked to slipping fault is marked as the solid black lines. The red square corresponds to the location of a hypothetical GPS station for which displacement as a function of time is indicated in the bottom row. *Bottom row*: Horizontal displacement at the station located with the red square in row 1, showing strain accumulation and coseismic displacement associated with an earthquake, indicated by the arrow.

plate interface is locked, the downgoing slab drags the overlying plate with it in the direction of plate motion producing a horizontal velocity field above the coupled plate boundary that parallels the convergence direction and decreases with distance from the trench (**Figure 3(b)**, top). When the strain reaches the failure strength of the fault, an earthquake occurs, and sudden slip on the plate interface causes surface motions that are in the same (**Figure 3(a)**, bottom) and opposite directions (**Figure 3(b)**, bottom) of strain accumulation at strike-slip and subduction-zone margins, respectively. Earthquake motion is complete within seconds to minutes, depending on the size of the earthquake. Eventually, the strain-accumulation cycle begins again and surface displacements return to their previous directions and rates (**Figure 3**, bottom).

This strain accumulation and release pattern associated with the earthquake cycle is clearly visible in the daily position changes recorded at GPS station Arequipa (AREQ) in Peru above the South American subduction zone (**Figure 4(a)**). The linear northeast motion visible in the 500 days preceding the M_w 8.4 2001 Peru earthquake is parallel to convergence between the Nazca and South American plates and represents interseismic strain accumulation on the plate interface. On 23 June 2001 (day 0 in **Figure 4(a)**), a large displacement of 30–40 cm with the opposite sense of motion (southwest) indicates the sudden occurrence of slip on the plate interface, that is, an earthquake. This event was followed by afterslip for several months before the strain-accumulation pattern resumed at the same rate as prior to the earthquake. **Figure 4(b)** shows a similar pattern in daily longitudinal position changes at GPS station ALBH on Vancouver Island, overlying the Cascadia subduction zone; however, here the earthquake-like motion that occurs on day 230 is much smaller (~5 mm) and takes almost 20 days to complete before returning to the strain-accumulation signal. This slow, earthquake-like slip that occurs on the plate interface but generates no

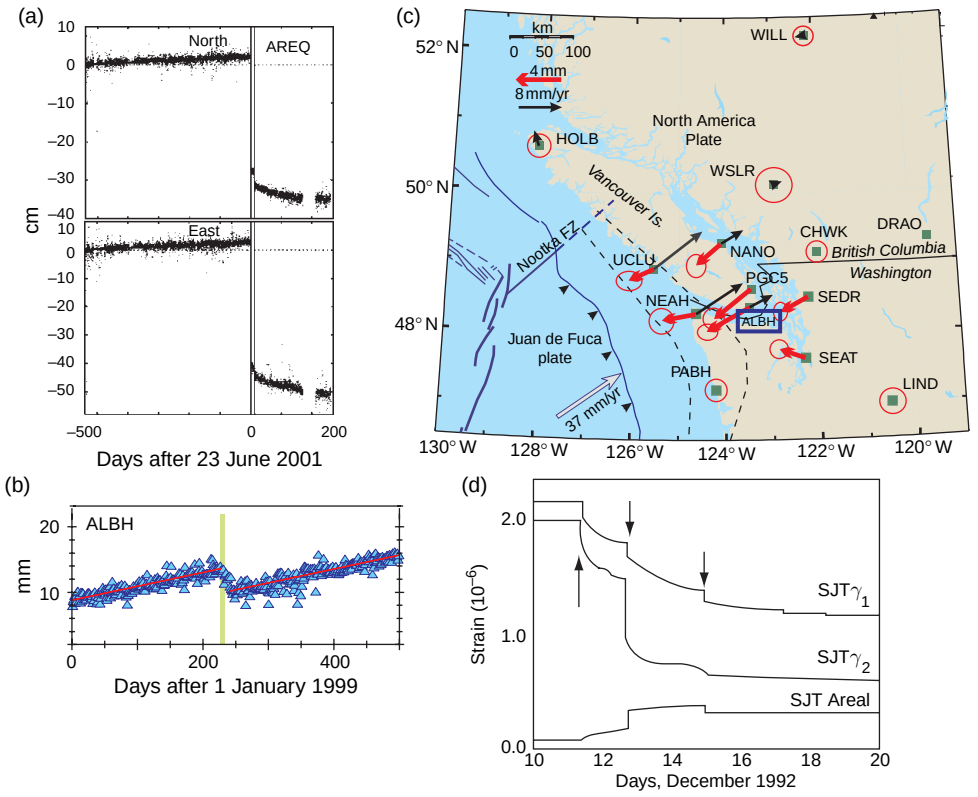


Figure 4 (a) Horizontal position solutions showing interseismic (slow linear northeast motion), coseismic (large, nearly instantaneous southwest motion), and postseismic deformation (slow logarithmic southeast motion) recorded at GPS station AREQ before and after the M_w 8.4 23 June 2001 Peru earthquake. (b) Longitudinal position solutions at GPS station ALBH, above the northern Cascadia subduction zone showing a slow-slip event (marked by the green line) that occurred in August 1999. (c) Map showing the location of the GPS stations that recorded the 1999 SSE. Red arrows show displacements caused by the SSE and black arrows indicate the interseismic strain-accumulation velocities. (d) Strain data from strainmeter SJT, located along the San Andreas Fault near San Juan Bautista, during 10 days in December 1992. The arrows indicate the occurrence of slow slip events on the fault parallel (γ_1), fault normal (γ_2) and areal strain components. (a) Adapted from Melbourne TI and Webb FH (2002) Precursory transient slip during the 2001 Mw = 8.4 Peru earthquake sequence from continuous GPS. *Geophysical Research Letters* 29(21): 2032 (doi:10.1029/2002GL015533). (c) Adapted from Dragert H, Wang K and James TS (2001) A silent slip event on the deeper Cascadia subduction interface. *Science* 292: 1525–1528. (d) Adapted from Linde AT, Gladwin M, Johnston M, Gwyther R and Bilham R (1996) A slow earthquake sequence on the San Andreas Fault. *Nature* 383: 65–68.

seismic radiation is often referred to as a slow-slip event (SSE). Surface velocities and displacements for the strain-accumulation period (black vectors) and this SSE (red vectors), determined at GPS stations operated by the Geological Survey of Canada and part of the Pacific Northwest Geodetic Array (PANGA), are shown in **Figure 4(c)**. The identification of SSEs in GPS displacement time series – that may contain other deformation signals or be contaminated by meteorological effects, local monument motion, and/or reference frame errors – is difficult and requires careful signal-processing techniques (e.g., McGuire and Segall, 2003). Borehole strainmeters can provide observations of crustal strain at much higher sensitivity in the critical SSE time range of months to minutes. **Figure 4(d)** shows strain changes detected in borehole strainmeters near the San Andreas Fault north of San Juan Bautista, CA during a 10-day period in December 1992. The large coherent strain changes evident between 11 and 15 December have been interpreted as a series of SSEs between 4 and 8 km depth on the fault (Linde *et al.*, 1996; Johnston and Linde, 2002).

Once SSEs are identified, associated strain or surface displacement data can be modeled, using elastic dislocation theory (Okada, 1985) to determine the SSE location on an assumed fault plane, average slip (d), seismic moment ($M_o = \mu A d$, where A is the fault area and μ is the shear modulus of the

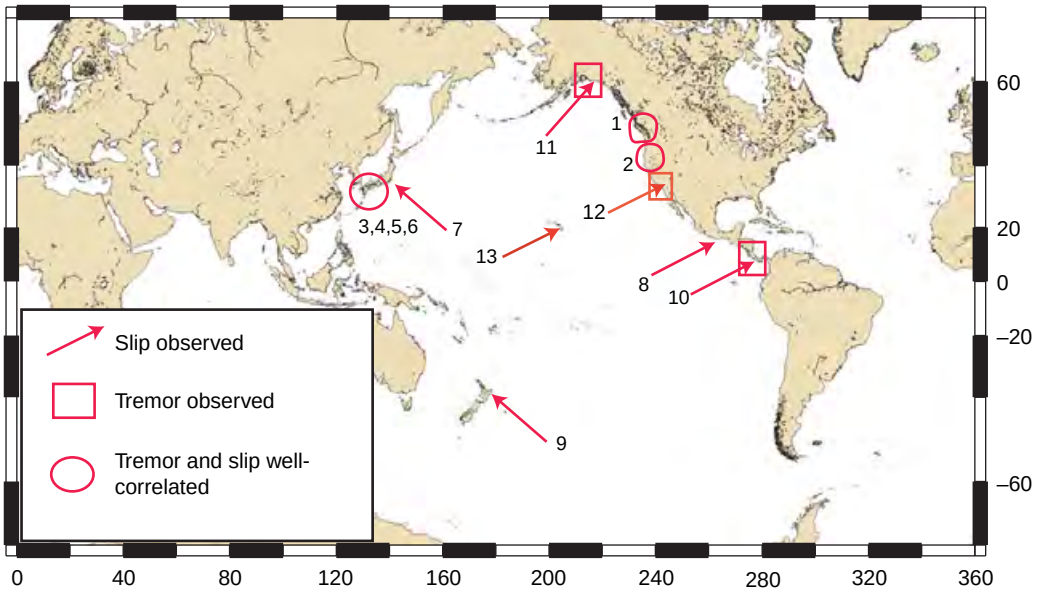


Figure 5 Map showing the location of slow-slip events discussed in this chapter. The numbers correspond to regions listed in **Table 1**. Circles and boxes represent regions where seismic tremor has been observed and is or is not well correlated with slow slip, respectively.

surrounding material) or equivalent earthquake moment magnitude ($M_w = 2/3 \log[M_o] - 6.07$, with M_o in Nm). In some areas, such as Japan and the Parkfield region of the San Andreas Fault with dense data coverage, deformation data have been inverted for the nonuniform distribution of slip or slip rate as a function of time on the assumed fault plane. Similar to coseismic slip during earthquakes, interseismic slow slip has been shown to vary over the fault plane with isolated patches of higher or faster slip.

Figure 5 shows the location of all well-documented interseismic SSEs that are discussed in the remainder of this chapter. The large majority of them occur at convergent margins. These plate boundaries generate the largest earthquakes and tend to be relatively well instrumented, so it is not surprising that the majority of SSEs locate here. Section 4.15.4 discusses relevant details of interseismic SSEs observed in each geographic region and **Table 1** summarizes this information.

4.15.2.4 Slip Episodes

Much of the transient deformation along strike-slip faults in California occurs as deviations in slip rate at depth from long-term rates (slip episodes) (Nadeau and McEvilly, 2004; Murray and Segall, 2005) or as shallow failure of the uppermost several hundred meters of the fault (Gladwin *et al.*, 1994), rather than as SSEs at seismogenic depths. Analysis of strainmeter and creepmeter data for several slip episodes in the Parkfield region of the San Andreas Fault constrained the depth of slip to the upper few hundred meters of the surface (Johnston and Linde, 2002). Similar results have been obtained for other portions of the creeping section of the San Andreas fault (Evans *et al.*, 1981) supporting the idea that most shallow slip episodes result from soil failure, may be triggered by increased rainfall (Roeloffs, 2001), and are therefore of little utility for understanding fault mechanics. However, it has been shown in some cases that shallow slip rates of creeping faults are modulated by slip at depth or slip on adjacent faults. For instance, surface slip accompanied a SSE near San Juan Bautista (Linde *et al.*, 1996; Johnston and Linde, 2002) and changes in surface slip rates were detected on creeping faults following nearby large earthquakes in several locations (Roeloffs, 2001). So deviations in shallow slip rate from long-term rates can be important indicators of faulting processes at depth within the seismogenic zone.

4.15.3 EAS and Seismic Tremor

EAS releases accumulated plate-boundary strain with durations of the order of minutes to years, too slow to generate seismic radiation. However, Rogers

Table 1 Episodic aseismic slip event characteristics

Geographic Region	Date (mo yr^{-1})	Size		τ (days)	Recurrence (years)	Migration (km day^{-1})	Tremor	References
		d (cm)	M_w					
1. Northern Cascadia	9/94	—	—	—	—	—	—	Dragert *et al.* (2001)
	12/95	—	—	—	1.3	—	—	Dragert *et al.* (2004)
	5/97	—	—	—	1.4	—	Yes	McGuire and Segall (2003)
	7/98	3	6.8	—	1.2	10–20	Yes	Melbourne *et al.* (2005)
	8/99	2–8	6.7–6.9	~50	1.1	6–15	Yes	Miller *et al.* (2002)
	12/00	3	6.7	—	1.3	—	Yes	Schmidt (2006)
	2/02	4	6.5	—	1.2	—	Yes	
	3/03	2–4	6.6	~40	1.1	5–10	Yes	
	7/04	3–4	6.8	~50	1.4	—	Yes	
	9/05	3–4	6.7	—	1.2	—	Yes	
2. Southern and Central Cascadia	Several zones of periodic SSEs (identified from single GPS stations) that correlate with seismic tremor occur in this region. Recurrence intervals range between 11 and 19 months and surface displacements are comparable to events in northern Cascadia							Brudzinski and Allen (2006) Szeliga *et al.* (2004)
3. Bungo Channel-Western Shikoku	Long term							Hirose *et al.* (1999)
	3/97	5–20	6.6–7.2	~300	—	Yes	—	Hirose and Obara (2005)
	8/03	11	6.8–7.0	~90	6	Yes	Yes	Miyazaki *et al.* (2003) Ozawa *et al.* (2004)
	Short term	Events identified by tilt changes approximately every 6 months lasting 5–7 days. Slip events always accompanied by tremor and the slip and tremor migrate together.						Hirose and Obara (2005) Obara *et al.* (2004) Obara and Hirose (2006)
	8/02	3.7	6.0	~5	0.5	Yes	Yes	
	8/03	3.9	6.1	~10	1	Yes	Yes	
	11/03	2.4	6.2	~7	0.25	Yes	Yes	
	2/04	2.1	6.0	~4	0.25	Yes	Yes	
	4/04	0.8	5.8	~5	0.17	Yes	Yes	
4. Eastern Shikoku	Events identified by tilt changes in 7/01, 11/01, 2/02, 5/02, 10/02, 6/03, each lasting several days to a week and strongly correlated with seismic tremor. Size of postulated SSEs are smaller than short-term events in W. Shikoku.							Obara and Hirose (2006)
5. Kii Peninsula	Events identified by tilt changes in 3/01, 9/01, 3/02, 9/02, 8/03, 1/04, 7/04, and 11/04 each lasting several days to a week and strongly correlated with seismic tremor. Size of postulated SSEs are smaller than short-term events in W. Shikoku.							Hirose and Obara (2006) Obara and Hirose (2006)
6. Tokai	Long term							Miyazaki *et al.* (2006)
	10/00	20–30	>7.0	~2000	>10	Yes	Yes	Ohta *et al.* (2004) Ozawa *et al.* (2002) Ozawa *et al.* (2005)

(*Continued*)

Table 1 (Continued)

Geographic Region	Date (mo yr^{-1})	Size		τ (days)	Recurrence (years)	Migration (km day^{-1})	Tremor	References
		d (cm)	M_w					Yamamoto *et al.* (2005)
	Short term	Slow slip with associated tremor is observed in Tokai with an approximately 6-month recurrence interval. Short-term SSEs appear to nucleate below the source area of the long-term Tokai SSE						Hirose and Obara (2006)
	11/04	1.8	5.8	4–5	~0.5	Yes	Yes	
	7/05	0.8	5.7	3	~0.5	Yes	Yes	
7. Boso	5/96	~5	6.4–6.5	7	—	Possible	No	Sagiya (2004)
	10/02	10–20	6.6	50	6	Yes	No	Ozawa *et al.* (2003)
8. Guerrero, Mexico	10/01	9–30	6.8–7.6	~200	~1	2	—	Franco *et al.* (2003)
	Events identified by campaign and continuous GPS and leveling observations in 95–96, 1998, 1999, 2000, early 2001, 2003 and 2004, with sizes that are slightly less than the 10/01 event and durations between 70 and 220 days. The apparent recurrence interval of SSEs in the region is ~1 year.							Kostoglodov *et al.* (2003)
								Larson *et al.* (2004)
								Lowry *et al.* (2001, 2005)
9. New Zealand	10/02	18	—	10	2–3	—	—	Beavan *et al.* (2007)
	1/04	35	7.0	~550	—	Up-dip	—	Douglas *et al.* (2005)
	Other events have been identified but at too few stations to accurately characterize.							Wallace and Beavan (2006)
10. Costa Rica	9/03	1.5	—	30	—	Down-dip	—	Protti *et al.* (2004)
	Postulated SSEs on 2/00, 4/00, and 6/00, recognized as correlated fluid-flow and seismic-tremor episodes measured on the ocean bottom lasting 2–3 weeks each.							Brown *et al.* (2005)
11. Alaska	1998	12–16	7.2	~1000	—	—	Yes	Ohta *et al.* (2006)
								Peterson *et al.* (2005)
12. SAF, California San Juan Bautista	12/92	0.3–2.5	5.0	5–7	—	—	No	Linde *et al.* (1996)
	SSEs in 4/96, 5/98, 1/03 and 3/04 determined from strainmeter recordings and estimated to have M_w from 4.9 to 5.0 and durations from 7 to 21 days.							Gladwin (2004)
								Gwyther *et al.* (2000)
								Johnston (1997)
								Johnston *et al.* (1996)
								Pollitz *et al.* (2006)
Parkfield	1993	Episodic slip rate changes lasting ~3 years with maximum slip velocity of 0.05 m yr^{-1}.						Gao *et al.* (2000)
								Gwyther *et al.* (1996)
								Langbein *et al.* (1999)
								Murray and Segall (2005)
13. Hawaii	9/98	—	5.6	A few days		—	—	Brooks *et al.* (2006)
	11/00	—	5.7		2.14	—		Cervelli *et al.* (2002)
	12/02	—	5.5		2.10	—		
	1/05	—	5.8		2.12	—		

d, average displacement; M_w, equivalent moment magnitude; τ, event duration.

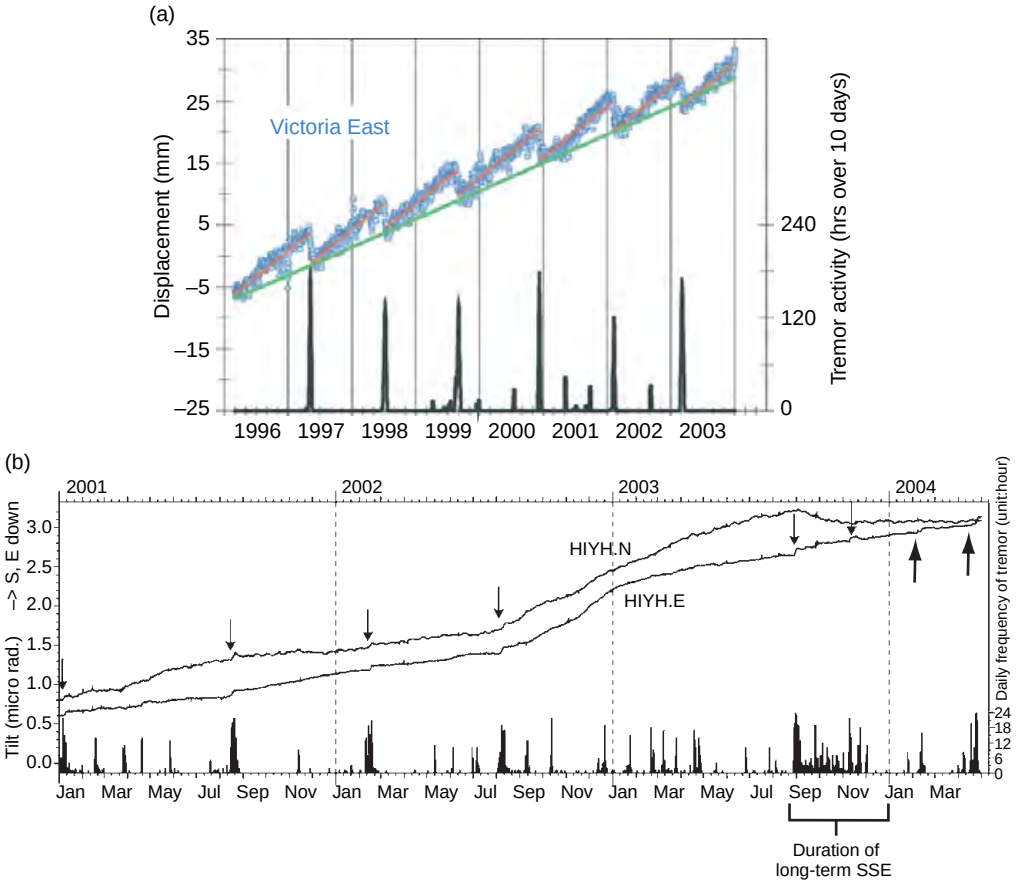


Figure 6 Correlated slow-slip and seismic-tremor activity in the northern Cascadia (a) and western Shikoku, Japan (b) subduction zones. Bursts in tremor activity, indicated as hours per unit time, strongly correlate with the SSEs apparent on GPS displacement (a) or tiltmeter (b) data. (a) Blue circles show east positions at GPS station ALBH, green line indicates the long-term eastward strain-accumulation signal which is interrupted by SSEs (causing motion reversals) about every 13–16 months. From Dragert H, Wang K, and Rogers G (2004) Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth, Planets and Space* 56: 1143–1150. (b) The north and east tilt records from Hinet station HIYH with arrows indicating short-term slow-slip events. The long-term Bungo channel slow-slip event began in August 2003 and corresponds to the highest concentration of tremor in the studied period. Modified from Obara K and Hirose H (2006) Non-volcanic deep low-frequency tremors accompanying slow slips in the southwest Japan subduction zone. *Tectonophysics* 417(1–2): 33–51.

and Dragert (2003) discovered that SSEs in the northern Cascadia subduction zone correlate with episodes of seismic tremor, a signal that is commonly associated with active volcanoes (**Figure 6(a)**). Peaks in seismic tremor activity have also been robustly tied to SSEs in the subduction zone of southwest Japan (**Figure 6(b)**) (Hirose and Obara, 2005) and correlations between SSEs and seismic tremor have been suggested at the Alaskan (Peterson *et al.*, 2005) and Costa Rican (Brown *et al.*, 2005) convergent margins. Subduction-zone tremor is characterized by long-duration (minutes to hours), high-amplitude seismic signal, without clear body-wave arrivals. Examples of seismic tremor observed during SSEs in Cascadia, southwest Japan, and southeast Alaska are shown in **Figure 7**. Tremor is identified by the coincidence of high-amplitude envelopes on several nearby stations (e.g., Obara, 2002; McCausland *et al.*, 2005). The plots on the right of **Figure 7** show envelope functions for the seismic records shown on the left. These envelopes show that increased amplitude of tremor can persist for over 30 min. The actual shape of these envelopes is correlated between stations to identify tremor as a relatively widespread phenomenon, rather than as a noise process isolated near a single station. Tremor duration varies between regions and episodes, and is often pulsed in nature (**Figure 7**). Tremor activity has traditionally been measured as the number of hours per day that

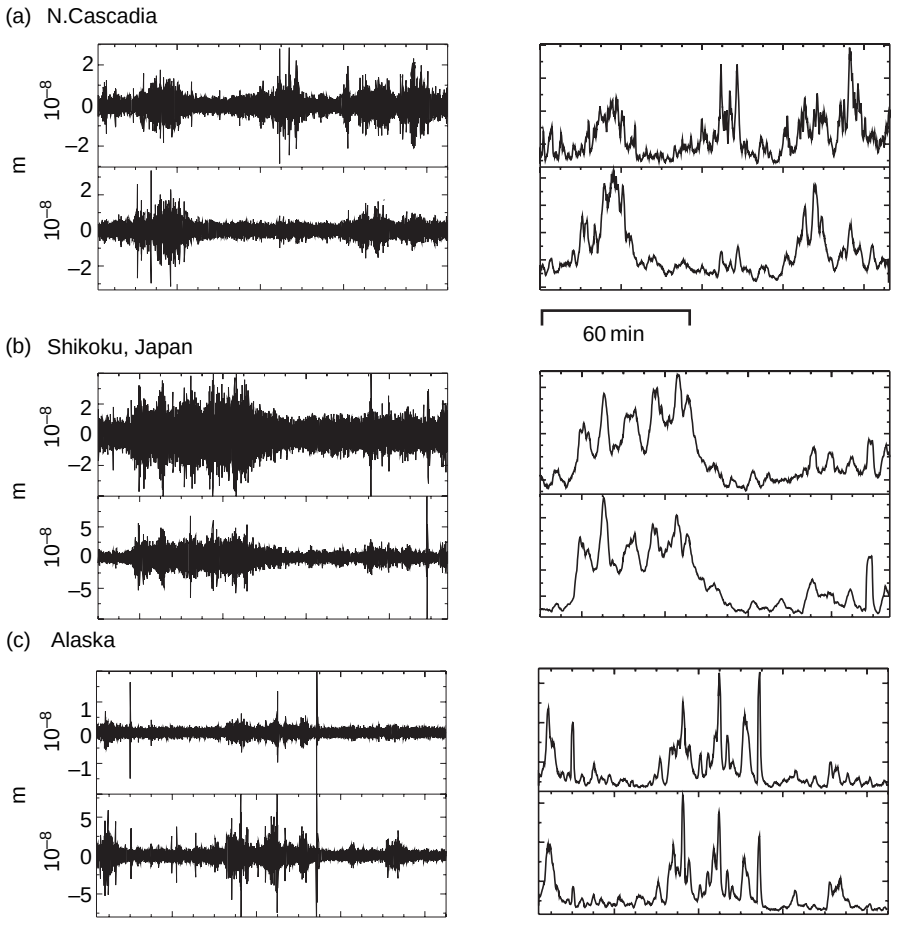


Figure 7 Horizontal records of seismic tremor (left) from two stations in (a) northern Cascadia, (b) western Shikoku, Japan, and (c) southeastern Alaska. All records are 8400 s long and band-pass filtered between 2 and 4 Hz. Note that amplitude scales vary between records. *Right*: Envelope-processed versions of the same records. These envelopes emphasize periods of tremor, which can be correlated between stations for identification and location.

contain tremor of any duration (**Figure 6**). The actual percentage of each hour filled by tremor varies significantly between different regions and tremor episodes.

Like volcanic tremor, which has been associated with the underground movement of fluids, subduction-zone tremor is generally characterized as low frequency, containing peak frequencies that are lower than local earthquake activity. Most tremor shows a rapid drop-off in frequency content above about 5–6 Hz (Kao *et al.*, 2005). Based on the spectral properties of subduction tremor and its presumed similarity to volcanic tremor, most generation models invoke the presence of fluids. Tremor depths in southwest Japan and Cascadia, between 25 and 40 km, coincide with pressure and temperature conditions that are suitable for slab dehydration, and tremor sources appear to cluster near regions of high vp/vs ratio consistent with the presence of fluids (Kurashimo and Naoshi, 2004; Matsubara *et al.*, 2005; Shelly *et al.*, 2006a). However, Shelly *et al.* (2006a) identify impulsive P- and S-arrivals embedded in tremor and used them to improve the location of tremor hypocenters. Their improved tremor locations formed a plane parallel to the inferred top of the subducting Philippine plate, more consistent with shear dislocation than fluid movement. By using well-located tremor events as templates, Shelly *et al.* (2006b) systematically searched tremor signals to detect matching waveforms at multiple stations. This technique revealed a nearly continuous sequence of detected events during times of active tremor, establishing that nonvolcanic tremor beneath Shikoku, Japan can be explained by a swarm of low-frequency earthquakes, with each event most likely caused by shear failure on the plate interface. In Cascadia, records of tremor do not exhibit such

impulsive arrivals. The quality discrepancy between Japanese borehole stations and surface stations used in North America may obscure the ability to see small-amplitude arrivals in the Cascadia traces. However, analyses of move-out curves for Cascadia tremor bursts suggest that multiple sources are active at the same time (McCausland *et al.*, 2005). This provides some additional evidence for a model in which multiple sources of shear failure interfere to create an emergent and long-duration tremor signal. Where studied, it appears that subduction tremor moves with S-wave velocities ($\sim$4 km s^{-1}), which is consistent with the observed concentration of energy on the horizontal components (Obara, 2002; Rogers and Dragert, 2003). Tremor has also been shown to propagate with slow slip in Cascadia and western Shikoku, Japan with propagation velocities between $\sim$5 and 15 km day^{-1}.

Tremors with characteristics similar to that recorded in subduction zones have been observed at the Parkfield (Ellsworth *et al.*, 2005) and Cholame (Nadeau and Dolenc, 2005) segments of the San Andreas Fault. Although the mechanisms behind subduction and continental fault tremors are not necessarily identical, similarity between the two phenomena argues for some relationship. If tremor is produced by the rupture of small asperities embedded within regions of accelerated aseismic slip (Shelly *et al.*, 2006a), then slow slip and tremor may be manifestations of the same phenomena and should be associated with one another regardless of tectonic environment. **Figure 5** and **Table 1** identify regions where seismic tremor is well correlated with SSEs or occurs with no clear association.

4.15.4 Global Observations of EAS

4.15.4.1 Convergent Margins

SSEs occur most frequently on the plate interface at convergent margins and exhibit a large variation in duration, magnitude, and recurrence behavior (**Table 1**). In fact, almost all subduction zones instrumented with a dense GPS network sufficient to detect SSEs have done so. One notable exception is the subduction zone off northern Honshu, where a very dense GPS network has recorded several episodes of afterslip, but no interseismic SSEs. Like afterslip, most observed interseismic slow slip locates at the down-dip extension of the seismogenic zone in the conditionally stable frictional regime (**Figure 1(b)**). The large range in behaviors exhibited by SSEs may provide clues to their generation. Documenting the behavior of these events in all subduction zones and comparing and contrasting their behaviors may contribute to a better understanding of how and why they occur and whether they perturb the local stress field in such a way as to bring the megathrust closer to failure in a great earthquake.

4.15.4.1.1 Cascadia

The Casadia subduction zone, extending 1000 km from British Columbia to northern California, has a history of generating $M_w > 8.0$ earthquakes approximately every 300–500 years with the last great earthquake occurring in 1700 (Satake *et al.*, 2003). Geodetic measurements made over the last decade have shown that the plate interface along this entire segment is locked from near the surface to a depth of about 20 km (Dragert *et al.*, 1994). This region of geodetic locking has been identified as the seismogenic zone. One of the earliest observations of a subduction-zone SSE was made along the northern portion of the Cascadia margin from CGPS data (Dragert *et al.*, 2001). These data revealed that in the fall of 1999 the contractional motions indicative of strain accumulation, at seven sites in southern Vancouver Island and western Washington, suddenly reversed their motion for a period of about 2 weeks at each station (**Figures 4(b)** and **4(c)**). Dragert *et al.* 2001) found that the surface displacement data associated with this excursion could be well fit by $\sim$2 cm of slow slip on a deep portion of the plate interface just below the seismogenic zone. This SSE had the equivalent moment of an $M_w = 6.7$ earthquake. Following this discovery careful analysis of the continuous GPS time series from this network dating back to 1992 revealed the occurrence of eight similar SSEs with an average recurrence interval of 13–16 months (Miller *et al.*, 2002). Each SSE lasted between 2 and 4 weeks at a particular station, migrated across the array and was consistent with the slip model obtained for the 1999 event, that is, a few centimeters of slip on the deep portion of the plate interface (25–45 km), yielding equivalent moment magnitudes of 6.5–6.8 (Dragert *et al.*, 2004). Inversion of surface displacements from the 1999 and 2003 Cascadia SSEs for models of the spatial and temporal variations in fault slip (McGuire and Segall, 2003; Melbourne *et al.*, 2005) indicated that like large earthquakes, SSEs also possess rupture complexity and concentrate their slip at isolated patches like asperities along the fault plane. Sufficient information to determine whether or not the same patches slip in

consecutive SSEs does not yet exist. **Table 1** lists the relevant parameters associated with 10 documented SSE events in the northern Cascadia margin.

Rogers and Dragert (2003) demonstrated a strong correlation between six Cascadia SSEs and seismic tremor (**Figure 6(a)**). Because of the correspondence and regular recurrence of tremor and SSEs in northern Cascadia, Rogers and Dragert (2003) termed this coupled phenomena episodic tremor and slip or ETS. Tremor and slip appear to migrate together in time, although the locations and migration patterns vary between episodes. Recent ETSs have shown bidirectional migration of slip and tremor sources during the multiweek episodes. Migration speeds varied between 5 and 15 km day^{-1} during the episodes. There appears to be no discernable migration in the vertical direction (Kao *et al.*, 2006).

CGPS and seismic data from northern California, Oregon, and southern Washington suggest that ETS occurs along the entire Cascadia margin. For northern California, Szeliga *et al.* (2004) detected eight SSEs, between 1997 and 2003 based on GPS position changes from a single station; the most recent five SSEs were accompanied by seismic tremor recorded by seismic stations of the Northern California Seismic Network. The average value of the horizontal surface displacement during these events (4–6 mm) is similar to the northern Cascadia SSEs; however, their recurrence interval is 9–12 months, shorter than the 13–16 months obtained in northern Cascadia. Although GPS station coverage is too sparse to correlate SSEs between stations for all but the northern Cascadia margin, several single GPS stations along this margin show sudden reversals of motion that are consistent with SSEs. In most cases, the correlation of sudden GPS motion reversals with seismic tremor episodes has been interpreted as strong evidence that they do represent slow slip on the subduction plate interface (Szeliga *et al.*, 2004; Brudzinski and Allen, 2006). In summary, the entire Cascadia margin appears to generate ETS with recurrence intervals that vary with location along the margin. Since all but the northern Cascadia SSEs are recorded at only a single GPS station, details of their locations, magnitudes, and migration patterns are unconstrained.

4.15.4.1.2 Southwest Japan

Southwestern Japan is a complex region of subduction where the Philippine Sea plate subducts beneath the Eurasian or Amurian plate along the Nankai and Sugura Troughs and beneath the North American plate along the Sagami Trough (**Figure 8(a)**). Great earthquakes and SSEs occur along most of this convergent margin. SSEs have been detected from network GPS observations in the Bungo Channel, Tokai, and Boso regions (**Figure 8** and **Table 1**). The operation of a very dense network of borehole tiltmeters in Japan since about 2000 (National Research Institute of Earth Science and Disaster Prevention or NIED Hi-net array) has allowed very small SSEs, not detected on surface GPS stations, to also be identified. Although these tilt changes are too small to be recognized as significant on their own, their strong correlation and migration with seismic tremor (**Figure 6(b)**) indicate that they are related phenomena and by association with ETS at Cascadia believed to reflect slow slip on the plate interface (Obara *et al.*, 2004). The SSEs detected by GPS typically last months to years, have equivalent moment magnitudes greater than 6.5, and recurrence intervals of more than 5 years and are referred to as long-term SSEs (**Table 1**). Short-term SSEs, detected with tiltmeters, are smaller, last only several days to a week and reoccur every 3–12 months (Obara *et al.*, 2004; Hirose and Obara, 2005; Obara and Hirose, 2006; Hirose and Obara, 2006). The best-documented short-term SSEs occurred beneath western Shikoku, with more poorly defined short-term slow slip evident in eastern Shikoku, the Kii Peninsula, and Tokai regions (**Figure 8(b)** and **Table 1**).

The first GPS deformation signal to be interpreted as a SSE was recorded by the Japanese nationwide GPS network, GEONET between 1996 and 1998 in the Bungo Channel region of southwestern Japan (Hirose *et al.*, 1999). A sudden reversal of surface displacements at several stations in the vicinity of the Bungo Channel, initiating in early 1997 and lasting ~300 days, suggested slow slip on the plate interface. Hirose *et al.* (1999) modeled the surface displacement pattern and found that it was best fit by variable slip down-dip of the seismogenic zone (Hyndman *et al.*, 1995). This model yielded an equivalent earthquake moment magnitude of 6.6. Similar slow-slip motion recurred in August 2003, as reported by Ozawa *et al.* (2004). They inverted the GPS observations for the spatial and temporal distribution of slow slip on the fault plane and found a very similar slip distribution to that of the 1997 SSE and an equivalent earthquake magnitude of $M_w = 7.0$. During both events, slip concentrated at the transition in plate coupling from strongly coupled in the northeast to weakly coupled in the southwest (Ito *et al.*, 1999; Ito and Hashimoto, 2004).

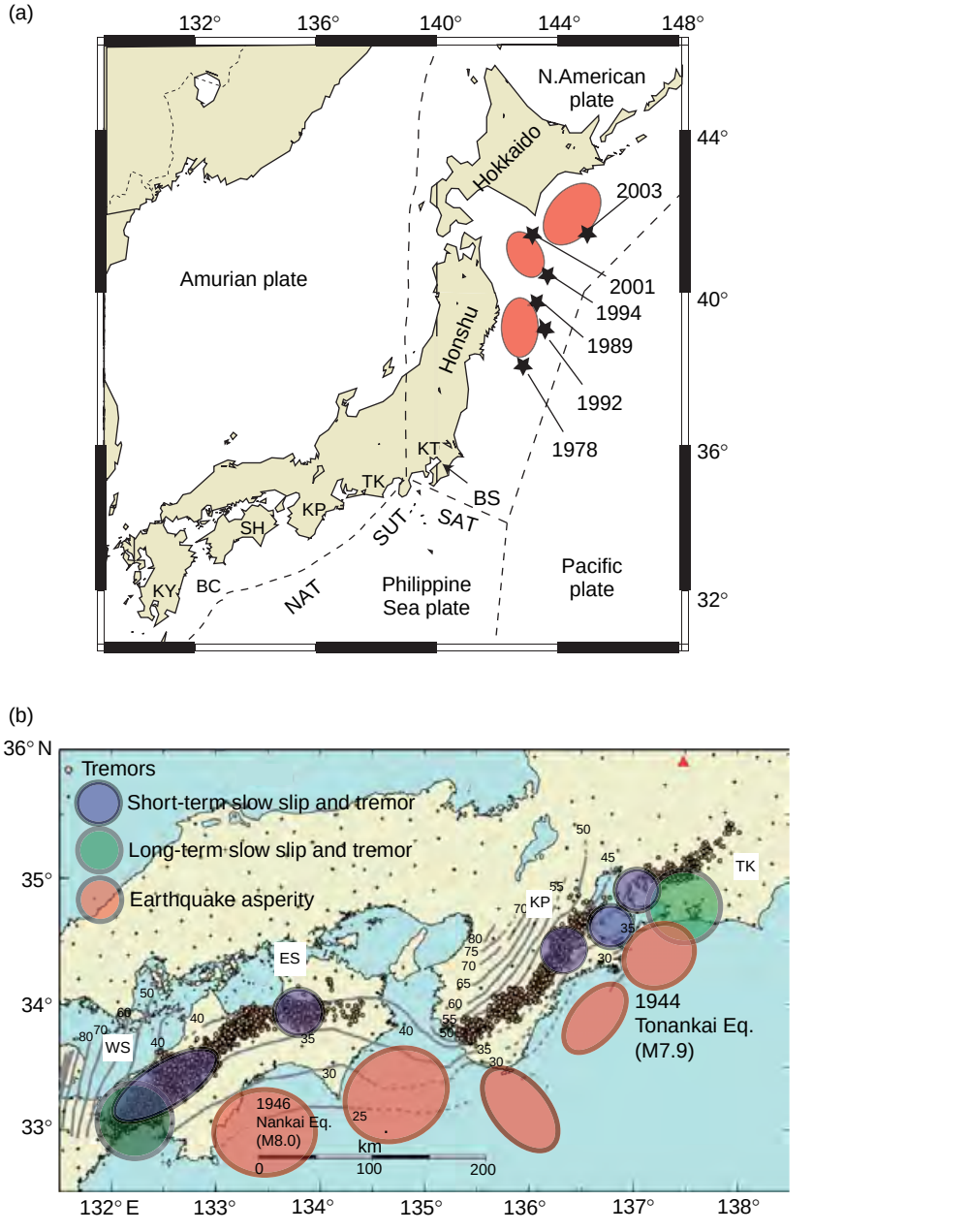


Figure 8 (a) Plate tectonic setting of Japan showing the location of six large to great earthquakes (stars) in northeast Japan that were followed by significant afterslip. Red ovals indicate regions of strong plate locking during the interseismic stage and concentrated slip during large earthquakes (asperities). Afterslip locates in weak regions around these patches. KY is Kyushu, BC, Bungo Channel; SH, Shikoku; KP, Kii Peninsula; TK, Tokai; KT, Kanto; BS, Boso; NAT, Nankai Trough; SUT, Sugura Trough; and SAT, Sagami Trough. (b) Spatial distribution of short- and long-term slow-slip events, great earthquake asperities, and seismic tremor in southwest Japan. WS and ES are western and eastern Shikoku, respectively. (b) Modified from Obara K and Hirose H (2006) Non-volcanic deep low-frequency tremors accompanying slow slips in the southwest Japan subduction zone. *Tectonophysics* 417(1–2): 33–51.

Continuous GPS data from GEONET stations in the Tokai region revealed a reversal in direction and change in deformation rate beginning in mid-2000 that persisted into 2006. These motions were shown to be consistent with slow slip on a deep (25–40 km) portion of the plate interface (Ozawa *et al.*, 2002; Ohta *et al.*, 2004; Miyazaki *et al.*, 2006). Miyazaki *et al.* (2006) inverted the GPS data in the time range

between January 2000 and November 2002 for the spatiotemporal distribution of slow slip on the fault plane. Their results indicated two slow subevents with slip in each that accelerated and decelerated several times but concentrated down-dip of the portion of the fault plane previously determined to be strongly coupled (Ohta *et al.*, 2004). The maximum slip attained in this time period was ~14 cm equivalent to an earthquake of $M_w = 6.8$; however, the SSE continued for several more years eventually attaining an M_w larger than 7.0. Analysis of tilt observations at a single station in the region between 1981 and the present revealed anomalous tilt associated with the recent Tokai SSE as well as a nearly identical anomaly between 1988 and 1990 (Yamamoto *et al.*, 2005). Yamamoto *et al.* (2005) interpreted the 1988–90 signal as an SSE and speculated that SSEs in the Tokai region occur repeatedly; however, they noted that the size and duration of the recent SSE far exceeds its predecessor and that variations in SSEs in this region must exist.

CGPS data in the time period between May and June 1996 revealed anomalous horizontal motions on the eastern Boso Peninsula that were interpreted as a SSE (Sagiya, 2004). Inversion of this data for the distribution of the slow slip indicated that approximately 50 mm of slip occurred at shallow depth (10–20 km) on the fault plane within a period of ~1 week, yielding an equivalent moment magnitude of 6.4. The source region of the SSE occurred in the same depth range as interseismic locking (Sagiya, 2004) and the 1923 Kanto earthquake asperity (Wald and Somerville, 1995), however, in a different location along strike, farther to the east. This is unusual in that all the other SSEs in southwest Japan located down-dip of regions of strong interseismic locking and coseismic rupture (**Figure 8(b)**). The source area of this SSE locates at the boundary between a strongly locked and more freely slipping region of the plate interface that corresponds to a large contortion in the downgoing Philippine Sea plate, perhaps due to its interaction with the westward-subducting Pacific Plate. A second SSE was documented in this region over a 50-day period beginning in October 2002 (Ozawa *et al.*, 2003). Ozawa *et al.* (2003) inverted the GPS data to obtain the time evolution of the 2002 Boso event and to compare it with characteristics of the SSE 6 years previously. They found that the magnitude of the 2002 event ($M_w = 6.6$) exceeded the 1996 event and although their slip areas overlapped, the centroid of slip during the 2002 event located farther to the south.

In addition to the long-term Bungo Channel SSEs in 1997 and 2003, eight episodes of short-term slow slip have been observed in western Shikoku between 2001 and 2005 (**Figure 6(b)**), **Table 1**). The longer duration (~1 year) 2003 SSE temporally overlaps two of these eight events. Dislocation modeling of the tilt data at several stations has allowed source parameters of five of the short-term SSEs to be determined (Obara *et al.*, 2004; Hirose and Obara, 2005; Obara and Hirose, 2006). In general the short-term SSEs locate at the down-dip edge of the plate interface at depths between 20 and 45 km, a bit deeper than the long-term 1997 and 2003 SSEs (**Figure 8(b)**). Their source areas appear to define several patches on the deep plate interface with each SSE activating a slightly different combination of these patches (Hirose and Obara, 2005). The average slip in each event is between 1 and 4 cm, yielding equivalent moment magnitudes between 5.8 and 6.2. While the long-term SSEs seem to repeat every 6 years, the recurrence interval for short-term SSEs is 2–6 months. The occurrence of short-term SSEs and correlated tremor has recently been discovered in the Tokai (Hirose and Obara, 2006), eastern Shikoku, and Kii Peninsula (Obara and Hirose, 2006) regions and approximate locations are indicated in **Figure 8(b)**.

4.15.4.1.3 Mexico

Subduction of the Cocos plate beneath the North American plate along the Middle American Trench in Mexico generates large to great earthquakes every 30–100 years. The last great earthquake that occurred along the Guerrero segment in southern Mexico was in 1911, making it a likely location for rupture in a future large earthquake. Data from a single CGPS station in the Guerrero region recorded transient surface displacements in early 1998 that were interpreted as slow slip on the plate interface (Lowry *et al.*, 2001). The CGPS data from this station, campaign data from several other stations in the area, as well as leveling data were consistent with moment approximately equivalent to an $M_w \geq 6.5$ earthquake that propagated along strike over several months. In 2001, seven permanent GPS stations in the Guerrero region recorded a reversal of surface motion consistent with an SSE beginning in October 2001, lasting approximately 7 months with an $M_w = 6.8$–7.5. The large range in size estimate for this event resulted from the use of different modeling techniques. Campaign and continuous GPS data collected in the Guerrero region between 1992 and early 2001

were inverted to obtain both the interseismic deformation pattern and the location of transient slip events during this period. The pattern of interseismic strain accumulation indicated a locked plate interface at depths shallower than 25 km and modeling of reversals of this motion indicated SSEs in 1995–96 ($M_w = 6.8$–7.1) as well as in 1998 ($M_w = 6.8$–7.1) that both occurred at depths below 25 km (Larson *et al.*, 2004). Lowry *et al.* (2005) detected additional SSEs along this plate boundary in 1999, 2000, 2001, 2003, and 2004, yielding a total of eight SSEs within 9 years, producing an average recurrence interval of approximately 1.1 years. Only the 2001–02 SSE had a well-determined location with up to 30 cm of slip on the plate interface concentrated below 25 km, below the region of strong interseismic coupling.

4.15.4.1.4 New Zealand

Since 2002, several SSEs of variable size and duration have been inferred to occur along the Hikurangi margin based on continuous GPS records (Beavan *et al.*, 2007). A reversal of surface displacements over a 10-day period in October 2002 at two continuous GPS stations on the northeast coast of the North Island of New Zealand was the first reported SSE in this region (Douglas *et al.*, 2005). The surface displacements were modeled by 18 cm of slip on the plate interface at the down-dip edge of the seismogenic zone. The largest SSE was recorded on seven continuous GPS stations to the south of the 2002 event beginning in early 2004 and lasting ~18 months (Wallace and Beavan, 2006). Inversion of the horizontal and vertical displacement field yielded a model of up to 35 cm of slip at the down-dip edge of the seismogenic zone. Smaller changes in surface displacements, detected on a few GPS stations in the intervening region between the larger SSEs, occur frequently (at least one each year), last about a month and have also been interpreted as SSEs on the plate interface at the down-dip end of the seismogenic zone (Beavan *et al.*, 2007).

4.15.4.1.5 Costa Rica

Three continuously recording GPS stations on the Nicoya Peninsula, Costa Rica recorded a transient deformation event lasting approximately 1 month in September–October 2003. The transient displacements were nearly opposite in direction to plate convergence and strain accumulation on the plate interface, and appeared to have initiated at the up-dip transition from a locked to more freely slipping interface and propagated landward (down-dip). The transient displacements have been interpreted as a slow underthrusting earthquake located primarily within the seismogenic zone (Protti *et al.*, 2004). The data are too limited to resolve details of the slip event, but a simple dislocation model suggests that the observations can be explained by an average of 1.5 ± 0.5 cm of slip on the plate interface. It is possible that this region of the plate boundary, part of which appears freely slipping in campaign GPS data (temporally averaged over nearly a decade), may in fact alternate between a locked state and SSEs.

Three SSEs in early 2000, lasting approximately 3 weeks each, were proposed in the Nicoya region, based on observations of correlated transient fluid-flow and seismic-tremor (Brown *et al.*, 2005). Brown *et al.* (2005) proposed a model where the poroelastic stress/strain field from dislocations along the shallow thrust fault force flow through fracture networks in the forearc and oceanic basement, generating seismic tremor recorded by ocean bottom seismometers (OBSs) collocated with the fluid-flow meters, and induce diffuse flow through shallow sediments producing fluid-flow transients. Although Brown *et al.* (2005) could not precisely locate the postulated SSEs, the magnitude and timing of the fluid-flow excursions argue for up-dip slip propagation extending to depths of only 1–2 km below the ocean floor; no constraints could be placed on the down-dip extension of slip. The GPS time series are too short to assess the repeat interval, if any, for Costa Rica events, but evidence for three SSEs in 2000 and 2003 suggest that this mode of strain release may be common along the Nicoya segment of the Costa Rica subduction zone.

4.15.4.1.6 Alaska

The eastern end of the Alaska-Aleutian subduction zone is the site of the great M_w 9.2 1964 Prince William Sound earthquake. This event ruptured a large region of the shallow plate interface above 30 km depth and had a long and complicated post-seismic deformation history (Cohen and Freymueller, 2004). Ohta *et al.* (2006) inverted GPS data for the period between 1997 and 2002 and found two distinct deformation patterns; data before 1998 and after 2001 showed surface displacements consistent with strain accumulation over the entire shallow plate interface while data between 1998 and 2001 indicated continued strain accumulation on the shallow plate interface and an SSE on the deeper interface in an isolated region north of Anchorage.

The region with the largest slip deficit or maximum locking lies beneath the southeast Kenai Peninsula and corresponds to the asperity that broke in the 1964 earthquake. During the period between 1998 and 2001 the surface velocities north of Anchorage changed compared to the other time periods and were consistent with the occurrence of an SSE. The best-fitting fault parameters for this SSE were found to be maximum and average slip rates of 55 and 40 mm yr^{-1} respectively, producing total slip of 12.0–16.5 cm over a fault area of $\sim 150 \times 150\,km^2$ for the 3-year event duration. The cumulative seismic moment was 1.1×10^{20} Nm, which corresponds to M_w 7.2 (Ohta *et al.*, 2006). The maximum slip area of the SSE locates at depths between 25 and 45 km, just below the seismogenic zone. Peterson *et al.* (2005) reported seismic tremor at stations spanning over 100 km along the Alaska/Aleutions trench over at least a few months of the SSE.

4.15.4.1.7 Northeast Japan

Strain release along the Pacific–North American plate interface in northeast Japan occurs as great earthquakes, many of which experienced significant amounts of afterslip, and continuous aseismic slip with no documented observations of interseismic SSEs. This is noteworthy due to the presence of dense networks of CGPS and tiltmeter stations in this region that are comparable to southwest Japan, where SSEs occur frequently. Suwa *et al.* (2006) inverted both the horizontal and vertical components of the GPS GEONET data to obtain the pattern of interseismic strain accumulation on the plate interface in northeastern Japan. They found a few patches of strong locking that extended to about 60 km in depth beneath north-central Honshu and to greater depth beneath Hokkaido (**Figure 8(a)**). A comparison of this pattern of interplate coupling with the asperity distribution for all earthquakes with $M_w > 7.5$ in this region, determined by Yamanaka and Kikuchi (2003, 2004) reveals a very close correspondence. This indicates that the regions of maximum coseismic slip during large earthquakes heal very quickly and become the loci of strain accumulation in the interseismic period. Significant afterslip that followed the 1994 Sanriku-Oki ($M_w = 7.7$) and 2003 Tokachi-Oki ($M_w = 8.0$) earthquakes was found to primarily occur in regions that slip relatively freely during interseismic periods. Afterslip following smaller earthquakes in this region has also been detected by strainmeters for the 1989 ($M_w = 7.4$) and 1992 ($M_w = 6.9$) Sanriku-Oki earthquakes (Kawasaki *et al.*, 1995, 2001), tide gauge and leveling for the 1978 Miyagi-Oki ($M_w = 7.4$) earthquake (Ueda *et al.*, 2001), and GPS for the moderate-sized 2001 ($M_w = 6.4$) Tokachi-Oki earthquake (Sato *et al.*, 2004). It appears that afterslip in the weakly coupled regions of the plate interface is a common occurrence following moderate to great earthquakes. The weakly coupled portions of the fault plane that host earthquake afterslip also generate a large number of small repeating earthquakes (Igarashi *et al.*, 2003). Repeating earthquakes have nearly identical waveforms recorded at the same stations, indicating that they rupture similar fault locations. Repeating earthquakes were first observed along the San Andreas Fault system in California and interpreted as the repeat rupture of small isolated asperities embedded within freely slipping regions (Nadeau and Johnson, 1998). Uchida *et al.*, (2003) contoured the distribution of repeating earthquakes on the plate boundary in northeastern Japan and found several concentrations of repeating earthquakes that they interpreted as freely slipping portions of the plate boundary. A comparison of the repeating earthquake clusters with the geodetically determined distribution of plate coupling confirms that they do occur where the plate interface is freely slipping. In summary, the plate interface in northeastern Japan appears to be composed of (1) regions that are accumulating strain at rates close to plate convergence that will become the loci of future large earthquakes (shaded red in **Figure 8(a)**), (2) regions that are continuously slipping at or near plate rates, and (3) regions that slip at intermediate rates somewhere between that of earthquakes and plate rates, but only when driven by coseismic slip in adjacent regions (these are regions of afterslip). Although several different modes of strain release are observed, there is no evidence that EAS-like events occur during the interseismic period in northeast Japan.

4.15.4.2 San Andreas Fault

Aseismic slip is a common mode of strain release within the San Andreas Fault system and is well documented along the Maacama, Concord, Hayward, Calaveras, and central San Andreas faults (Galehouse, 2002). This slip occurs as shallow subsurface creep and at seismogenic depth and includes both steady-state and episodic motions. The episodic nature of shallow creep appears to be caused primarily by soil failure (Johnston and Linde (2002) and references therein) and will not be discussed further in this chapter.

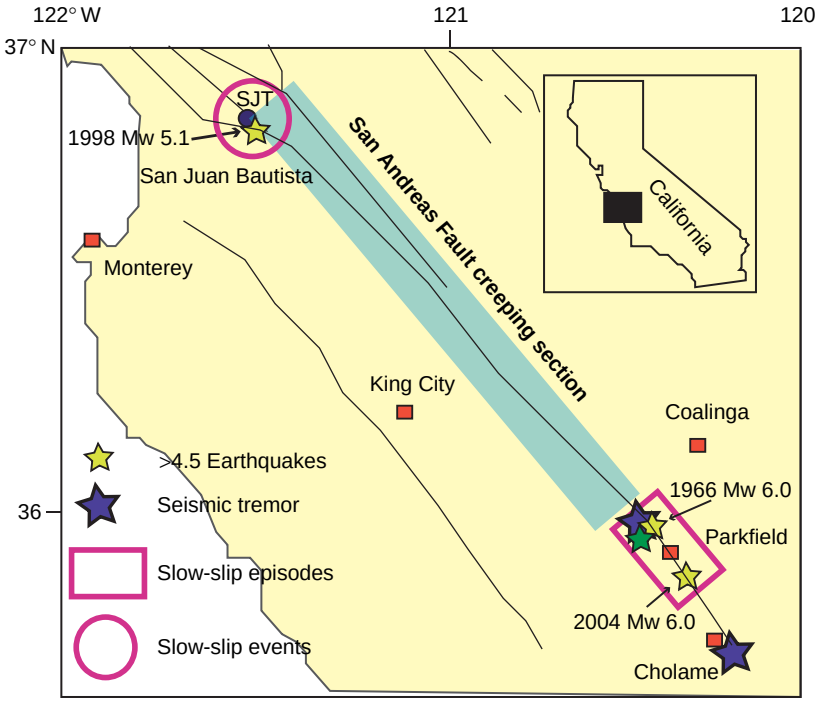


Figure 9 Map of central coastal California indicating the location of episodic aseismic slip associated with the creeping section of the San Andreas fault. The large circle and rectangle locate regions of slow-slip events near San Juan Bautista and changes in aseismic slip rates near Parkfield. The large blue stars mark the location where seismic tremor was recorded, the yellow stars mark the location of significant large earthquakes, the green star the location of ~M 4.5 earthquakes that preceded slip rate changes at Parkfield, and the circle labeled SJT shows the location of the borehole strainmeter for which data are shown in Figure 4.

Aseismic strain transients at seismogenic depth have been reported as slow-slip episodes, inferred from geodetically detected rate variations (Langbein *et al.*, 1999; Gao *et al.*, 2000) and characteristics of repeating microearthquakes (Nadeau and McEvilly, 1999, 2004) and SSEs inferred from borehole strainmeter data (Linde *et al.*, 1996). The locations of these EAS events are indicated in **Figure 9**. Several SSEs have been reported along the central San Andreas Fault near San Juan Bautista in 1992 (Linde *et al.*, 1996), 1996 (Johnston *et al.*, 1996; Johnston, 1997), 1998 (Gwyther *et al.*, 2000), 2003, and 2004 (Gladwin, 2004; Pollitz and Johnston, 2006); however, only the first of these is well documented. The December 1992 event was recorded by three borehole strainmeters and two surface creepmeters located near the fault. These data (at station SJT in **Figure 4(d)**) revealed a complicated series of strain changes that began abruptly, contained many subevents, and lasted for approximately 5 days. Forward modeling of these data indicated a slip sequence with complexities as great as those observed during normal earthquakes. The source processes consisted of several slip episodes scattered in the upper 8 km of ~50 km^2 fault plane with slip amplitudes and durations ranging from 0.3 to 2.5 cm and 0.6 to 43 h, respectively (Linde *et al.*, 1996). The total moment of this sequence was equivalent to an M_w 5 earthquake.

Episodic slow slip at Parkfield, CA was identified by abrupt changes in electronic distance meter (EDM) line length measurements, tensor strain, and a 30% increase in slip rate on the fault inferred from the recurrence interval of repeating earthquakes beginning in early 1993 and lasting until 1996 (Langbein *et al.*, 1999; Gao *et al.*, 2000; Nadeau and McEvilly, 1999). The deviations in surface deformation were successfully modeled as acceleration in slip rate at seismogenic depth on the San Andreas Fault (Gwyther *et al.*, 1996; Langbein *et al.*, 1999; Gao *et al.*, 2000). While previous studies assumed constant slip rates over fixed time intervals, Murray and Segall (2005) analyzed EDM data over a longer time period (mid-1980s to 2003) to image the time-dependent evolution of slip on the fault. They found episodic slip rate changes between 1993 and 1996, with the largest increases correlating with the occurrence of three $M \sim 4.5$ earthquakes (**Figure 9**), the largest

events to occur in the Parkfield region between 1966 and 2004. Slip was constrained to occur in the upper 8 km of the crust with a maximum estimated slip rate of 4.9 cm yr^{-1}, compared with a long-term slip rate of ~3.5 cm yr^{-1}. Nadeau and McEvilly (2004) determined slip rates as a function of time along the entire 175 km creeping section of the central San Andreas Fault (**Figure 9**) from analysis of repeating microearthquakes. They found considerable variability, with rate changes as large as 100% throughout the entire region, and some areas that possessed quasi-periodic rate changes with durations and recurrence periods of 1.7–3 and 2.5–3 years, respectively. Nadeau and McEvilly (2004) favorably compared these slip episodes to interseismic SSEs at convergent margins, concluding that both phenomena may share a common mechanism. However, only the 2000 Tokai SSE had a duration as long as these SAF slip episodes. No seismic tremor has been associated with the San Juan Bautista SSE or episodic slip on the central SAF. However, deep seismic tremor (20–40 km depth) was detected along the San Andreas Fault near Cholame, California, some 40 km south of Parkfield (Nadeau and Dolenc, 2005) and during a short period in May 2005 from borehole seismometers located within the San Andreas Fault Observatory at Depth (SAFOD) pilot drill hole (Ellsworth *et al.*, 2005) near Parkfield, CA (**Figure 9**).

4.15.4.3 Hawaii

A CGPS network on Kilauea volcano, Hawaii recorded transient displacements in November 2000 that were interpreted as a SSE (Cervelli *et al.*, 2002). Inversion of the GPS data indicated that slip occurred on a shallow dipping thrust fault at a depth of ~5 km beneath the southeast flank of the volcano. Slip attained a maximum rate of 6 cm day^{-1}, lasted for about 36 h, producing 8.7 cm of total slip with an equivalent moment magnitude of 5.7 (Cervelli *et al.*, 2002). A peak in rainfall 9 days before this SSE suggests a causal relationship. Cervelli *et al.* (2002) postulated that an elevated water table, resulting from nearly 1 m of rainfall in a short period of time, increased the pore pressure and decreased the effective normal stress on crustal faults triggering slip. Reexamination of 8 years of GPS data from Kilauea revealed three additional SSEs that were not associated with periods of increased rainfall (Brooks *et al.*, 2006). While source parameters for the new SSEs could not be uniquely determined, their GPS data are consistent with motion on a shallow reverse fault, similar to the model proposed by Cervelli *et al.* (2002) with comparable seismic moments. These four SSEs appear periodic with a recurrence interval of 774 ± 7 days. Brooks *et al.* (2006) suggested that a cycle of aquifer recharge, pressurization, slow slip, aquifer rupture, and re-pressurization might be responsible for the periodicity of the SSEs.

4.15.4.4 Oceanic Transform Faults

EAS has been invoked to explain anomalous seismic or geodetic behavior observed in several other tectonic settings as well. The rupture process of large oceanic transform faults has been shown to often involve slow precursory slip followed by compound rupture on spatially distinct faults (e.g., McGuire *et al.*, 1996; McGuire and Jordan, 2000). Support for this idea comes from observations of transform fault microseismicity. Forsyth *et al.* (2003) reported alternating slip between adjacent transform faults during an earthquake swarm recorded by OBS in the southern Pacific Ocean. Since static or dynamic triggering by such small earthquakes over relatively large distances (25 km) was deemed unlikely, Forsyth *et al.* (2003) suggested EAS to explain event coupling during these earthquake swarms. Observations of high foreshock to aftershock rates for events along the East Pacific Rise, compared to continental strike-slip faults, prompted McGuire *et al.* (2005) to invoke slow-slip transients as a mechanism for triggering these earthquakes. Similarly, Vidale and Shearer (2006) found that over 70 earthquake swarms in southern California did not follow typical foreshock–mainshock–aftershock behavior and concluded that EAS was the most probable explanation for the observed steady rate of seismicity.

4.15.5 EAS and Seismicity

Afterslip associated with moderate to large earthquakes suggests a relationship between EAS and higher-velocity rupture. In addition, models of EAS generation that appeal to small perturbations in the stress field (e.g., Shen *et al.*, 2005) imply that minute stress changes related to nearby earthquakes can trigger slow slip. Conversely, one of the motivations behind increased study of intereseismic EAS, is the potential triggering effect these events may have for normal earthquakes. Many individual observations suggesting a triggering relationship between EAS

and seismic behavior have been made including (1) the occurrence of moderate to large earthquakes associated with EAS, (2) changes in seismicity rate preceding or following EAS, and (3) changes in slip rates following moderate to large earthquakes; however, a global pattern has yet to emerge and more work is clearly warranted.

The most extensive work on seismicity patterns and EAS was performed for the multiyear Tokai SSE of 2000. Several authors reported a relative quiescence in shallow (<50 km) seismicity prior to slip in Tokai (Yamamoto *et al.*, 2005; Matsumura, 2006). During the SSE, Yoshida *et al.* (2006) reported variations in slab and crustal seismicity concurrent with changes in slip velocity. The rate of slab earthquakes ($M > 1.1$) increased at the initiation of slow slip, decreased during a period of slower slip, and rose again coincident with an increase in slip. The fact that this correlation is made using seismicity down to $M = 1.1$ suggests that a highly complete catalog may be necessary for tracking any subtle changes in seismicity resulting from slow slip.

Hirose *et al.* (1999) suggested that the 1996–97 Bungo Channel SSE was affected by changes in stress due to the occurrence of two large M_w 6.7 earthquakes near Hyuganada, 100 km to the southwest, about 3 months prior to the SSE. These two large earthquakes were accompanied by significant afterslip. Although Miyazaki *et al.*, (2003) showed that the regions of afterslip and slow slip did not overlap, they still may have been related through static stress transfer. It is possible that stress perturbation caused by the Hyuganada events contributed to the longer duration and larger moment of the 1996–97 compared with the 2003 SSE, but this possibility is difficult to constrain. Additionally, a small swarm of earthquakes (maximum $M_j = 4.9$) occurred north of the Bungo Channel in April 1997, just after the initiation of slow slip (Hirose *et al.*, 1999; Miyazaki *et al.*, 2003).

Ozawa *et al.* (2003) noted a seismic swarm off the coast of Boso Peninsula on the upper surface of the Philippine Sea plate (20–30 km) concurrent with the regions 2002 SSE. A similar swarm occurred during the 1996 SSE, suggesting a consistent relationship between microseismicity and slow slip (Sagiya, 2004). Ozawa *et al.* (2003) suggested that this swarm activity increased stress in the neighboring region and caused the observed slow slip. Consequently, these authors used other peaks in seismicity to propose additional slip events in 1983 and 1991–92 (consistent with the inferred 6–7 year recurrence interval).

A cursory analysis of seismicity patterns of moderate to large earthquakes ($M > 3$ in southwest Japan and $M > 4$ elsewhere) associated with SSEs at six convergent margins (Cascadia, western Shikoku and Boso, Japan, Guerrero, Mexico, New Zealand, and Costa Rica) revealed a few isolated correlations, but no consistent cause-and-effect relationship (Schwartz and Rokosky, 2007). The most robust pattern occurred in Costa Rica where Brown *et al.* (2005) reported a rise in cumulative moment 20 days after each of three, correlated fluid-flow and seismic-tremor episodes that occurred in 2000 and were interpreted as SSEs. A series of earthquakes to the northwest of the 2003 geodetically inferred SSE also produced a precipitous rise in cumulative seismic moment and event number following this slip event (Schwartz and Rokosky, 2007).

Well-documented observations of EAS along the San Andreas Fault have also been associated with anomalous seismic behavior. The 1992 San Juan Bautista SSE was accompanied by several small earthquakes (M_L 3.1–3.7), not typical of seismicity patterns before or after this event (Linde *et al.*, 1996). The largest recent earthquake (M_w 5.1) along the San Juan Bautista segment of the San Andreas Fault (**Figure 9**) immediately preceded the 1998 SSE in this region (Uhrhammer *et al.*, 1999). Nadeau and McEvilly (2004) found a correlation between the occurrence of $M > 3.5$ earthquakes and pulses of accelerated slip that they inferred from repeating microearthquakes. They reported a six- to sevenfold increase in seismic activity near the onset of accelerated slip episodes compared with background rates. They could not determine whether accelerated slip preceded or followed earthquake productivity increases; however, they concluded that the quasi-periodic recurrence of accelerated slip episodes and increased seismicity indicated a triggering relationship. Murray and Segall (2005) also reported a correlation between accelerated slow slip and seismicity during the 1993–96 Parkfield, California slip episode. They found that two of the largest episodic slip rate increases were preceded by the occurrence of three $M \sim 4.5$ earthquakes (**Figure 9**), suggesting that the earthquakes may have triggered slow slip.

In Hawaii, an increase in microseismicity was associated with all four of the SSEs inferred to occur there, with slow slip clearly preceding the seismicity increase for the largest SSE of 2005 (Segall *et al.*, 2006). Assuming that the increased earthquake activity resulted from static stress changes induced by slow slip, Segall *et al.* (2006) constrained the 2005 SSE to occur on the same, nearly horizontal plane as the earthquakes. The authors suggested that triggering of microseismicity by slow slip may be a common

behavior that goes undetected in regions with low levels of background seismicity (e.g., Cascadia). Perhaps microseismicity triggered by SSEs has not gone undetected and is responsible for generating seismic tremor commonly recorded in the Cascadia and southwestern Japan subduction zones. Based on the planar location of relocated low-frequency earthquakes embedded within tremor in southwest Japan, Shelly *et al.* (2006a) speculated that tremor may represent patches of more accelerated shear failure on the plate interface surrounded and driven by broad regions of slow slip.

4.15.6 Mechanics of EAS

EAS at convergent margins, either occurring in the interseismic stage or as afterslip, has primarily been detected at the down-dip edge of the seismogenic zone with slip continuing into the deeper conditionally stable portion of the plate interface (**Figure 1(b)**). Notable exceptions include the 1996, 2002 Boso and 2003 Costa Rica interseismic SSEs and afterslip following several large earthquakes in northeastern Japan and the 2004 Sumatra and 2005 Nias, Indonesia earthquakes. EAS along the SAF also occurs within the seismogenic zone. In these cases, however, EAS occurs in complementary locations along strike of strongly locked patches (coseismic asperities) or large earthquake hypocenters. No EAS has been located in strongly locked regions of the plate interface. This observation indicates that different frictional properties control the occurrence of fast versus slow slip. Rate- and state-dependent frictional modeling supports this, finding that SSEs require either spatially or temporally variable friction. Shibazaki and Iio (2003) were able to simulate SSEs by introducing temporal variations in their rate- and state-dependent frictional modeling such that steady-state friction behaves as velocity weakening at low slip velocity, and velocity strengthening at high slip velocity. Their simulated SSEs had slip velocities of tens of meters per year, propagated horizontally at rates of $\sim$10 km yr^{-1}, and persisted for tens of years. Yoshida and Kato (2003) and Liu and Rice (2005) successfully simulated SSEs with spatially variable friction using block-spring models and three-dimensional numerical simulations, respectively. In the block-spring models, slip events occurred only in regions where friction was near the stability limit (Yoshida and Kato, 2003). In the numerical modeling, SSEs appeared spontaneously at the down-dip edge of the seismogenic zone, which Liu and Rice (2005) attributed to the transition from unstable to stable frictional properties. These simulated strain transients had slip velocities that were meters to tens of meters per year (1–2 orders of magnitude faster than plate velocities). Resulting slip was confined to the down-dip frictional stability transition and migrated along strike at tens of km per year over tens of years. Only the last characteristic is unlike most observations of SSEs in subduction zones in that it is longer than the observations. Transitions in frictional properties from velocity weakening to velocity strengthening seem to be required to generate SSEs. While we expect this transition at the down-dip edge of seismogenic zones, where most SSEs occur, it appears likely that such transitions also exist within seismogenic zones (**Figure 10**) and are the loci of SSE nucleation.

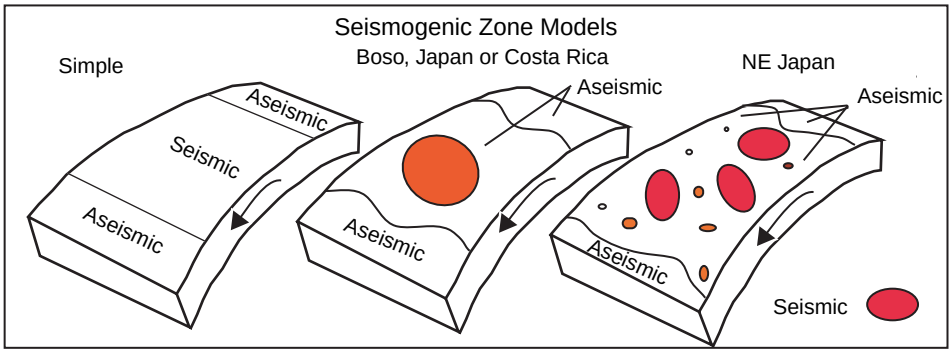


Figure 10 Cartoon of the plate interface at different types of convergent margin seismogenic zones. The simple model contains only two frictional transitions from velocity-weakening (seismic) to conditionally stable or velocity-strengthening (aseismic) behavior at its up- and down-dip limits. The Boso or Costa Rica (middle) and northeast Japan (right) models have these transitions within the seismogenic zone. Slow-slip events (SSEs) nucleate at these frictional transitions with most occurring at the deeper transition. The extreme frictional heterogeneity of the northeast Japan subduction zone may prevent SSEs from initiating within its seismogenic zone.

The subduction zones where no interseismic SSEs occur at or below the seismogenic zone (NE Japan, Boso, and Costa Rica) share some common features that may be important in understanding the generation of slow slip. In particular, the seismogenic zones at these plate boundaries, defined by the 350°C isotherm (believed to represent the frictional transition from velocity weakening to strengthening behavior) and/or the down-dip limit of underthrusting earthquakes or geodetic locking, extend to depths below where the continental Moho intersects the downgoing plate. The three subduction zones that have the best located SSEs at the down-dip edge of the seismogenic zone (Cascadia, SW Japan and Guerrero, Mexico) have relatively hot plate interfaces where temperatures reach 350°C at shallow depths, above where the continental Moho intersects the subducting plates. Although it is unclear which attributes of these hot subduction zones are most important for slow-slip generation, reaching the frictional transition at relatively low pressure or where oceanic crust of the downgoing plate is in contact with crust of the overlying plate appears to be required for deep slow slip.

The Boso and Costa Rica subduction zones have interseismic SSEs that locate at relatively shallow depths within their seismogenic zones. At Boso, slow slip occurs between 10 and 20 km depth adjacent to a locked patch, believed to be the likely location of the future Kanto earthquake (Sagiya *et al.*, 1995). In Costa Rica it occurs at depths between 15 and 25 km, in a region geodetically determined to be freely slipping at the down-dip edge of a strongly coupled segment of the plate interface (Norabuena *et al.*, 2004). As mentioned above, slow-slip nucleation most likely occurs at frictional transitions within the seismogenic zone at depths above where the continental Moho intersects the downgoing plate.

In northeast Japan, afterslip is very common following great, large- and even moderate-sized earthquakes and locates both within and below the seismogenic zone. However, interseismic SSEs have not been identified anywhere at this margin. As discussed previously, afterslip is believed to take place in velocity-strengthening regions, driven by stress concentrations that remain at frictional transitions after earthquakes. Afterslip occurs within the seismogenic zone of northeast Japan indicating the presence of such frictional transitions, so why doesn't slow slip nucleate at these shallow transitions? Previous work has established that the seismogenic zone in northeast Japan is extremely heterogeneous in its frictional properties, consisting of both large and small patches with velocity-weakening behavior embedded within regions with velocity-strengthening behavior (eg., Igarashi *et al.*, 2003; Uchida *et al.*, 2003). Perhaps this extreme degree of frictional heterogeneity prevents large-enough expanses of velocity-strengthening material from existing to sustain significant slow slip to be detected. Numerical simulations by Kuroki *et al.* (2004) showed that SSEs had large nucleation zones, lending support to this idea. **Figure 10** illustrates this concept, showing seismogenic zones with increasing levels of frictional heterogeneity. The large regions of velocity-strengthening material (labeled aseismic in **Figure 10**) postulated to exist within the Boso and Costa Rica seismogenic zones may host slow slip, while comparable regions are absent within the northeast Japan seismogenic zone. It is also plausible that the large amount of afterslip following earthquakes in northeast Japan relieves nearly all the stress in the velocity-strengthening regions of the plate interface, leaving little to be released in SSEs.

Schwartz and Rokosky (2007) investigated scaling relations for well-documented observations of SSEs. The moment magnitude (M_w), which is based on the equivalent seismic moment (M_o), and the source duration (τ) were considered to be the best-determined parameters from modeling studies. Unlike seismic moment and earthquake source duration, that follow the scaling relation $M_o \propto \tau^3$, implying constant earthquake stress drop (Kanamori and Brodsky, 2004), they found $M_o \propto \tau^{0.7}$ for all data and $M_o \propto \tau$ if the Tokai and Alaska SSEs, with less well-determined, nearly an order of magnitude larger source durations, were eliminated. Well-characterized afterslip fell off this relationship and in general had shorter durations for equivalent seismic moments. Although the 1992 San Juan Bautista SSE had a much longer duration for its equivalent seismic moment, the 2000 Hawaii SSE fits this relationship well emphasizing the difference between SSE and earthquake behavior.

4.15.7 Recurrence of EAS

Most SSEs are recurrent, with repeat intervals that range from many years (long-term SSEs in the Bungo Channel, Boso, and possibly Tokai, Japan) to a few months (short-term SSEs in the Bungo Channel, Japan). The recurrence interval for SSEs in northern Cascadia is well established at 13–16 months. Miller *et al.* (2002) noted that this repeat interval encompasses the 14.5 month periodicity of the Earth's

Chandler wobble (wobble of Earth's rotation axis caused by mass redistribution at the Earth's surface) but dismissed a causative relationship due to the small size of the stress perturbations generated. Shen *et al.* (2005) quantitatively investigated this relationship by calculating the stress changes induced by the Chandler wobble at the down-dip plate interface in northern and southern Cascadia, Guerrero, Mexico, Bungo Channel, and Boso Peninsula, Japan and comparing them with SSE cyclicity. They found that 14 out of 20 SSEs occurred just prior to, and 3 occurred synchronous with, induced stress changes attaining their maxima. They concluded that although these stress perturbations were very small (between 0.2 and 1 kPa), the long period (14.5 months) of the modulating stress was critical in reducing the triggering threshold and inducing slow slip. In addition, they suggest that the perturbations trigger SSEs rather than earthquakes because the triggering threshold is lowered in the region of the fault plane with frictional conditions that favor SSEs.

Lowry (2006) hypothesized that periodic slow slip is a resonant fault response to stress perturbations caused by the Chandler wobble and/or other climatic loading phenomena. However, he noted that on average, hydrologic stress perturbations are at least a factor of 2 greater than those generated by the Chandler wobble. He noted that for slip to be triggered by hydrologic or other climatic forcing it must respond preferentially to longer-period stress oscillations. Perfettini and Schmittbuhl (2001) found that slip resulting from an applied stress was amplified provided that the frictional conditions were near the stability transition and the period of loading was close to a critical value. Lowry (2006) called this critical period for slip amplification, the natural resonant period of the fault and showed that it depended on frictional parameters of the fault plane and the plate convergence velocity. He speculated that environmental loading by the ocean, atmosphere, and tides provided a broad-enough spectrum of forcing signals to potentially match the resonant periods of all fault planes hosting SSEs.

4.15.8 Conclusions

The increasing density of CGPS and other geodetic observations worldwide has revealed EAS to be a ubiquitous process that constitutes a fundamental mode of strain release at plate boundaries. For a long time the aseismic component of plate motions has been assumed to take place primarily as steady-state motion; however, evidence is accumulating to support a model where a large fraction of aseismic slip at plate boundaries occurs episodically rather than continuously. These episodic aseismic motions appear to have a significant influence on other, probably interrelated tectonic processes, such as the generation of seismic tremor, pore-fluid pressurization and expulsion at the seafloor (Brown *et al.*, 2005; Davis *et al.*, 2006) and earthquake nucleation. It is too early in the study of aseismic strain transients to make overarching conclusions about their behavior, but it is certain that continued study of this phenomenon is critical to attaining a meaningful understanding of seismogenic zones. The Plate Boundary Observatory (PBO) component of Earthscope, established to study the strain field resulting from active deformation between the Pacific and North American plates, is well poised to accomplish this in western North America. PBO is in the process of deploying a dense array of continuously operating GPS receivers and strainmeters in the western United States favorably positioned to record strain transients associated with the Cascadia and Alaskan subduction zones and the San Andreas Fault. Geodetic data collected at PBO stations with stable monumentation and long time series promise to be useful in improving our understanding of EAS at these locations.

Progress in developing and using other geodetic techniques capable of detecting and studying EAS at plate boundaries is also being made. For example, Lambert *et al.* (2006) demonstrated that terrestrial, absolute gravity measurements can be used in conjunction with space-based measurements to detect short-term gravity variations associated with SSEs. There are undoubtedly many more efforts to establish and upgrade monitoring capabilities at active plate boundaries, all of which will contribute to a fast growing database of EAS events to study, resulting in new discoveries and a deeper understanding of this important mode of strain release.

Acknowledgments

The author thanks Julie Rokosky for her collaboration and important contributions to EAS research at UCSC. The author is grateful to Kazushige Obara, Wendy McCausland and Chloe Peterson for providing tremor waveforms from Japan, Cascadia and Alaska, Michael Brudzinski for providing a copy of his paper prior to publication and Tim Dixon, Emily

Brodsky and Greg Beroza for enlightening conversations about episodic aseismic slip. This work was supported by NSF through award EAR-0506463.

References

Beroza GC and Jordan T (1990) Searching for slow and silent earthquakes. *Journal of Geophysical Research* 95: 2485–2510.

Beavan J, Wallace L, Fletcher H, and Douglas A (2007) Slow slip events on the Hikurangi subduction interface, NZ. *Journal of Geodesy* (in press).

Bilek SL, Lay T, and Ruff LJ (2004) Radiated seismic energy and earthquake source duration variations from teleseismic source time functions for shallow subduction zone thrust earthquakes. *Journal of Geophysical Research* 109: B09308 (doi:10.1029/2004JB003039).

Brooks BA, Foster JH, Bevis M, Frazer LN, Wolfe CJ, and Behn M (2006) Periodic slow earthquakes on the flank of Kilauea volcano, Hawaii. *Earth and Planetary Science Letters* 246: 207–216.

Brown KM, Tryon MD, DeShon HR, Dorman LM, and Schwartz SY (2005) Correlated transient fluid pulsing and seismic tremor in the Costa Rica subduction zone. *Earth and Planetary Science Letters* 238: 189–203.

Brudzinski MR and Allen R (2006) Segmentation in episodic tremor and slip all along Cascadia. *EOS Transactions of American Geophysical Union*, *Fall Meeting Supplement* 87(52): T53G-05.

Bucknam RC, Plafker G, and Sharp RV (1978) Fault movement (afterslip) following the Guatamala earthquake of February 4, 1976. *Geology* 6: 170–173.

Bürgmann R, Schmidt D, Nadeau RM, *et al.* (2000) Earthquake potential along the northern Hayward fault, California. *Science* 289(5482): 1178–1182.

Bürgmann R, Kogan MG, Levin VE, Scholz CH, King RW, and Steblov GM (2001) Rapid aseismic moment release following the 5 December 1997 Kronotsky Kamchatka earthquake. *Geophysical Research Letters* 28: 1331–1334.

Cervelli P, Segall P, Johnson K, Lisowski M, and Miklius A (2002) Sudden aseismic fault slip on the south flank of Kilauea volcano. *Nature* 415: 1014–1018.

Cifuentes IL and Silver PG (1989) The 1960 Chilean earthquake. *Journal of Geophysical Research* 94: 665–680.

Cohen SC and Freymueller JT (2004) Crustal deformation in southcentral Alaska: The 1964 Prince William Sound earthquake subduction zone (2004). *Advances in Geophysics* 47: 1–63.

Davis EE, Becker K, Wang K, Obara K, Ito Y, and Kinoshita M (2006) A discrete episode of seismic and aseismic deformation of the Nankai trough subduction zone accretionary prism and incoming Philippine Sea plate. *Earth and Planetary Science Letters* 242: 73–84.

Dieterich JH (1979) Modeling of rock friction 1. Experimental results and constitutive equations. *Journal of Geophysical Research* 84: 2161–2168.

Douglas A, Beavan J, Wallace L, and Townend J (2005) Slow slip on the northern Hikurangi subduction interface, New Zealand. *Geophysical Research Letters* 32, doi:10.1029/2005GL023607.

Dragert H, Hyndman RD, Rogers GC, and Wang K (1994) Current deformation and the width of the seismogenic zone of the northern Cascadia subduction thrust. *Journal of Geophysical Research* 99(B1): 653–658.

Dragert H, Wang K, and James TS (2001) A silent slip event on the deeper Cascadia subduction interface. *Science* 292: 1525–1528.

Dragert H, Wang K, and Rogers G (2004) Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth, Planets and Space* 56: 1143–1150.

Ellsworth WL, Luetgert JH, and Oppenheimer DH (2005) Borehole array observations of non-volcanic tremor at SAFOD. EOS *Transactions of American Geophysical Union, Fall Meeting Supplement* 86(52): T21A-0443.

Evans KF, Burford RO, and King GCP (1981) Propagating episodic creep and aseismic slip behavior of the Calaveras Fault north of Hollister, California. *Journal of Geophysical Research* 86: 3721–3735.

Forsyth DW, Yang Y, Mangriotis MD, and Shen Y (2003) Coupled seismic slip on adjacent oceanic transform faults. *Journal of Geophysical Research* 30, doi:10.1029/2002GL016454.

Franco SI, Kostoglodov V, Larsen KM, Manea VC, Manea M, and Santiago JA (2005) Propogation of the 2001–2002 silent earthquake and interplate coupling in the Oaxaca subduction zone, Mexico. *Earth, Planets and Space* 57: 973–985.

Galehouse JS (2002) Data from theodolite measurements of creep rates on San Francisco Bay Region faults, California: 1979–2001, *US Geological Survey Open File Report*, 02–225.

Galehouse JS and Lienkaemper JJ (2003) Inferences drawn from two decades of alignment array measurements of creep on faults in the San Francisco Bay region. *Bulletin of the Seismological Society of America* 93: 2415–2433.

Gao SS, Silver PG, and Linde AT (2000) Analysis of deformation data at Parkfield, California: Detection of a long-term strain transient. *Journal of Geophysical Research* 105: 2955–2968.

Gladwin MT, Gwyther RL, Hart RHG, and Breckenridge K (1994) Measurements of the strain field associated with episodic creep events on the San Andreas fault near San Juan Bautista, California. *Journal of Geophysical Research* 99: 4559–4565.

Gladwin MT (2004) Maintenance, data archive and analysis of existing low-frequency GTSM installations in California, National Earthquake Hazard Reduction Program annual project summary. US Geological Survey, Reston, VA.

Gwyther RL, Gladwin MT, Mee M, and Hart RHG (1996) Anomalous tensor strain at Parkfield during 1993–1994. *Geophysical Research Letters* 23: 2425–2428.

Gwyther RL, Thurber CH, Gladwin MT, and Mee M (2000) Seismic and aseismic observations of the 12th August 1998 San Juan Bautista, California M5.3 earthquake, paper presented at 3rd San Andreas Fault Conference, Stanford University Stanford, CA.

Hamilton S and Shennan I (2005) Late Holocene great earthquakes and relative sea level change at Kenai, southern Alaska. *Journal of Quaternary Science* 20: 95–111.

Heki K, Miyazaki S, and Tsuji H (1997) Silent fault slip following an interplate thrust earthquake at the Japan Trench. *Nature* 386: 595–598.

Hirose H, Hirahara K, Kimata F, Fujii N, and Miyazaki S (1999) A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, southwest Japan. *Geophysical Research Letters* 26(21): 3237–3240 (doi:10.1029/1999GL010999).

Hirose H and Obara K (2005) Repeating short- and long-term slow slip events with deep tremor activity around the Bungo channel region, southwest Japan. *Earth, Planets and Space* 57: 961–972.

Hirose H and Obara K (2006) Short-term slow slip and correlated tremor episodes in the Tokai region, central Japan. *Geophysical Research Letters* 33: L17311 (doi:10.1029/2006GL026579).

Hsu Y-J, Simons M, Avouac J-P, *et al.* (2006) Frictional afterslip following the 2005 Nias–Simeulue earthquake, Sumatra. *Science* 312: 1921–1926.

Hutton W, DeMets C, Sanchez O, Suarez G, and Stock J (2001) Slip kinematics and dynamics during and after the 1995 October 9 Mw = 8.0 Colima–Jalisco earthquake, Mexico, from GPS geodetic constraints. *Geophysical Journal International* 146: 637–658.

Hyndman RD, Wang K, and Yamano M (1995) Thermal constraints on the seismogenic portion of the southwestern Japan subduction thrust. *Journal of Geophysical Research* 100: 15373–15392.

Igarashi T, Matsuzawa T, and Hasegawa A (2003) Repeating earthquakes and interplate aseismic slip in the northeastern Japan subduction zone. *Journal of Geophysical Research* 108: 2249 (doi:10.1029/2002JB001920).

Ito T, Yoshioka S, and Miyazaki S (1999) Interplate coupling in southwest Japan deduced form inversion analysis of GPS data. *Physics of the Earth and Planetary Interiors* 115: 17–34.

Ito T and Hashimoto M (2004) Spatiotemporal distribution of interplate coupling in southwest Japan from inversion of geodetic data. *Journal of Geophysical Research* 109: B02315 (doi:10.1029/2002JB002358).

Johnston MJS (1997) Near-silent earthquakes on the San Andreas Fault system. *Eos Transaction*, *American Geophysical Union Fall Meeting Supplement* 78(46): 156.

Johnston MJS, Gwyther RL, Linde AT, Gladwin MT, Myren GD, and Mueller RJ (1996) Another slow earthquake on the San Andreas fault triggered by a M4.7 earthquake on April 19, 1996. *Eos Transaction*, *American Geophysical Union Fall Meeting Supplement* 77(46): 515.

Johnston M and Linde A (2002) Implications of crustal strain during conventional, slow, and silent earthquakes. In: Lee WHK, Kanamori H, Jenning P, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology*, *International Geophysics Series*, Vol. 81, pp. 589–605. New York: Elsevier.

Kanamori H and Brodsky EE (2004) The physics of earthquakes. *Reports on Progress in Physics* 67: 1429–1496.

Kanamori H and Cipar J (1974) Focal process of the great Chilean earthquake May 22, 1960. *Physics of the Earth and Planetary Interiors* 9: 127–136.

Kanamori H and Kikuchi M (1993) The 1992 Nicaragua earthquake: A slow tsunami earthquake associated with subducted sediments. *Nature* 361(6414): 714–716.

Kanamori H and Stewart GS (1976) Mode of strain release along the Gibbs fracture zone, Mid-Atlantic Ridge. *Physics of the Earth and Planetary Interiors* 11: 312–332.

Kao H, Shan S, Dragert H, Rogers G, Cassidy JF, and Ramachandran K (2005) A wide depth distribution of seismic tremors along the northern Cascadia margin. *Nature* 436: 841–844.

Kao H, Shan S, Dragert H, *et al.* (2006) Spatial-temporal patterns of seismic tremors in northern Cascadia. *Journal of Geophysical Research* 111: B03309 (doi:10.1029/2005JB003727).

Kawasaki I, Asai Y, Tamura Y, *et al.* (1995) The 1992 Sanriku-Oki, Japan, ultra-slow earthquake. *Journal of Physics of the Earth* 43: 105–116.

Kawasaki I, Asai Y, and Tamura Y (2001) Space-time distribution of interplate moment release including slow earthquakes and the seismogeodetic coupling in the Sanriku-Oki region along the Japan trench. *Tectonophysics* 330: 267–283.

Kostoglodov V, Singh SK, Santiago JA, *et al.* (2003) A large silent earthquake in the Guerrero seismic gap, Mexico. *Geophysical Research Letters* 30: 1807 (doi: 10.1029/2003GL017219).

Kreemer C, Blewitt G, and Maerten F (2006) Co- and postseismic deformation of the 28 March 2005 Nias M_w 8.7 earthquake from continuous GPS data. *Geophysical Research Letters* 33: L07307 (doi: 10.1029/2005GL025566).

Kurashimo E and Naoshi H (2004) Low V_p and V_p/V_s zone beneath the northern Fossa Magna basin, central Japan, derived from a dense array observation. *Earth, Planets and Space* 56(12): 1301–1308.

Kuroki H, Ito HM, Takayama H, and Yoshida A (2004) 3-D simulation of the occurrence of slow slip events in the Tokai region with a rate- and state-dependent friction law. *Bulletin of the Seismological Society of America* 94: 2037–2050.

Lambert A, Courtier N, and James TS (2006) Long-term monitoring by absolute gravimetry: Tides to postglacial rebound. *Journal of Geodynamics* 41: 307–317.

Langbein J, Gwyther RL, Hart RHG, and Gladwin MT (1999) Slip-rate increase at Parkfield in 1993 detected by high-precision EDM and borehole tensor strainmeters. *Geophysical Research Letters* 26: 2529–2532.

Larson KM, Kostoglodov V, and Lowry A *Geophysical Research Letters* (2004) Crustal deformation measurements in Guerrero, Mexico. *Journal of Geophysical Research* 109: B04409 (doi:10.1029/2003JB002843).

Linde AT, Gladwin M, Johnston M, Gwyther R, and Bilham R (1996) A slow earthquake sequence on the San Andreas Fault. *Nature* 383: 65–68.

Linde AT and Sacks SI (2002) Slow earthquakes and great earthquakes along the Nankai trough. *Earth and Planetary Science Letters* 203: 265–275.

Linde AT and Silver PG (1989) Elevation changes and the great 1960 Chilean earthquake: Support for aseismic slip. *Geophysical Research Letters* 16: 1305–1308.

Linde AT, Suyehiro K, Satoshi M, Sacks SI, and Takagi A (1988) Episodic aseismic earthquake precursors. *Nature* 334(6182): 513–515.

Liu Y and Rice JR (2005) Aseismic slip transients emerge spontaneously in three-dimensional rate and state modeling of subduction earthquake sequences. *Journal of Geophysical Research* 110: B08307 (doi:10.1029/2004JB003424).

Lowry AR (2006) Resonant slow fault slip in subduction zones forced by climatic load stress. *Nature* 442(7104): 802–805.

Lowry AR, Larson KM, Kostoglodov V, and Bilham R (2001) Transient fault slip in Guerrero, southern Mexico. *Geophysical Research Letters* 28: 3753–3756.

Lowry AR, Larson KM, Kostoglodov V, and Sanchez O (2005) The fault slip budget in Guerrero, Southern Mexico. *Geophysical Journal International* 200: 1–15.

Marone CJ, Scholtz CH, and Bilham R (1991) On the mechanics of afterslip. *Journal of Geophysical Research* 96: 8441–8452.

Matsubara M, Yagi Y, and Obara K (2005) Plate boundary slip associated with the 2003 off-Tokachi earthquake based on small repeating earthquakes. *Geophysical Research Letters* 32: L08316 (doi: 10.1029/2004GL022310).

Matsumura S (2006) Seismic activity changes progressing simultaneously with slow-slip in the Tokai area. *Tectonophysics* 417(1–2): 5–15.

McCausland W, Malone S, and Johnson D (2005) Temporal and spatial occurrence of deep non-volcanic tremor: From Washington to Northern California. *Geophysical Research Letters* 32(24): L24311 (doi:10.1029/2005GL024349).

McGuire JJ, Ihmle PF, and Jordan TH (1996) Time-domain observations of a slow precursor to the 1994 Romanche transform earthquake. *Science* 274: 82–85.

McGuire JJ and Jordan TH (2000) Further evidence for the compound nature of slow earthquakes: The Prince Edward Island earthquake of April 28, 1997. *Journal of Geophysical Research* 105: 7819–7828.

McGuire JJ and Segall P (2003) Imaging of aseismic fault slip transients recorded by dense geodetic networks. *Geophysical Journal International* 155: 778–788.

McGuire JJ, Boettcher MS, and Jordan TH (2005) Foreshock sequences and short-term earthquake predictability on east pacific rise transform faults. *Nature* 434: 457–461.

Melbourne TI and Webb FH (2002) Precursory transient slip during the 2001 Mw = 8.4 Peru earthquake sequence from continuous GPS. *Geophysical Research Letters* 29(21): 2032 (doi:10.1029/2002GL015533).

Melbourne TI, Webb FH, Stock JM, and Riegber C (2002) Rapid postseismic transients in subduction zones from continuous GPS. *Journal of Geophysical Research* 107: B102241 (doi:10.1029/2001JB000555).

Melbourne TI, Szeliga WM, Miller MM, and Santillan VM (2005) Extent and duration of the 2003 Cascadia slow earthquake. *Geophysical Research Letters* 32(4): L04301 (doi:10.1029/2004GL021790).

Miller MM, Melbourne T, Johnson DJ, and Sumner WQ (2002) Periodic slow earthquakes from the Cascadia subduction zone. *Science* 295: 2423.

Miyazaki S, McGuire JJ, and Segall P (2003) A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array. *Journal of Geophysical Research* 108, doi:10.1029/2001JB000456.

Miyazaki S, Seagall P, McGuire JJ, Kato T, and Hatanaka Y (2006) Spatial and temporal evolution of stress and slip rate during the 2000 Tokai slow earthquake. *Journal of Geophysical Research* 111: B03409 (doi: 10.1029/2004JB003426).

Murray JR and Segall P (2005) Spatiotemporal evolution of a transient slip event on the San Andreas Fault near Parkfield, California. *Journal of Geophysical Research* 110: (doi:10.1029/2005JB003651).

Nadeau R and Dolenc D (2005) Nonvolcanic tremors deep beneath the San Andreas Fault. *Science* 307: 389.

Nadeau R and Johnson LR (1998) Seismological studies at Parkfield VI: Moment releas rates and estimates of source parameters for small repeating earthquakes. *Bulletin of the Seismological Society of America* 88: 790–814.

Nadeau RM and McEvilly TV (1999) Fault slip rates at depth from recurrence intervals of repeating microearthquakes. *Science* 285: 718–721.

Nadeau RM and McEvilly TV (2004) Periodic pulsing of characteristic microearthquakes on the San Andreas Fault. *Science* 303: 220–222.

Newman AV and Okal EA (1998) Teleseismic estimates of radiated seismic energy: The E/M_o discriminant for tsunami earthquakes. *Journal of Geophysical Research* 103(11): 26885–26898.

Norabuena E, Dixon TH, Schwartz SY, *et al.* (2004) Geodetic and seismic constraints on some seismogenic zone processes in Costa Rica. *Journal of Geophysical Research* 109: B11403 (doi:10.1029/2003JB002931).

Obara K (2002) Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* 296: 1679–1681.

Obara K, Hirose H, Yamamizu F, and Kasahara K (2004) Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone. *Geophysical Research Letters* 31: doi:10.1029/2004GL020848.

Obara K and Hirose H (2006) Non-volcanic deep low-frequency tremors accompanying slow slips in the southwest Japan subduction zone. *Tectonophysics* 417(1–2): 33–51.

Ohta Y, Kimata F, and Sagiya T (2004) Reexamination of the interplate coupling in the Tokai region, central Japan, based on the GPS data in 1997–2002. *Geophysical Research Letters* 31: L24604 (doi:10.1029/2004GL021404).

Ohta Y, Freymueller JT, Hreinsdóttir S, and Suito H (2006) A large slow slip event and the depth of the seismogenic zone in the south central Alaska subduction zone. *Earth and Planetary Science Letters* 247(1–2): 108–116.

Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America* 75: 1135–1154.

Ozawa S, Murakami M, Kaidzu M, *et al.* (2002) Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan. *Science* 298: 1009–1012.

Ozawa S, Miyazaki S, Hatanaka Y, Imakiire T, Kaidzu M, and Murakami M (2003) Characteristic silent earthquakes in the eastern part of the Boso peninsula, central Japan. *Geophysical Research Letters* 30: 1283 (doi:10.1029/2002GL019381).

Ozawa S, Hatanaka Y, Kaidzu M, Murakami M, Imakiire T, and Ishigaki Y (2004) Aseismic slip and low-frequency earthquakes in the Bungo Channel, southwestern Japan. *Geophysical Research Letters* 31: 7609 (doi:10.1029/2003GL016665).

Ozawa S, Murakami M, Masaru K, and Yuki H (2005) Transient crustal deformation in Tokai region, central Japan until May 2004. *Earth, Planets and Space* 57(10): 909–915.

Pelayo AM and Wiens DA (1992) Tsunami earthquakes: Slow thrust-faulting events in the accretionary wedge. *Journal of Geophysical Research* 97(11): 15321–15337.

Perez-Campos X, McGuire JJ, and Beroza GC (2003) Resolution of the slow earthquake/apparent stress paradox for oceanic transform fault earthquakes. *Journal of Geophysical Research* 108, (doi:1029/2002JB002312).

Perfettini H and Schmittbuhl J (2001) Periodic loading on a creeping fault: Implications for tides. *Geophysical Research Letters* 28(3): 435–438.

Peterson C, Christensen D, and McNutt S (2005) Episodic tremor in the Alaska/Aleutian subduction zone. *Eos Transaction, American Geophysical Union Fall Meeting Supplement* 85: G51B-0830.

Pollitz FF and Johnston MJS (2006) Direct test of static stress versus dynamic stress triggering of aftershocks. *Geophysical Research Letters* 33,(doi:10.1029/2006GL026764).

Pritchard ME and Simons M (2006) An aseismic slip pulse in northern Chile and along-strike variations in seismogenic behavior. *Journal of Geophysical Research* 111, (doi:1029/2006JB004258).

Protti M, Gonzales V, Kato T, *et al.* (2004) A creep event on the shallow interface of the Nicoya Peninsula, Costa Rica seismogenic zone. *Eos Transaction, American Geophysical Union Fall Meeting Supplement* 85: S441D-07.

Roeloffs E (2001) Creep rate changes at Parkfield, California, 1966–1999: Seasonal, precipitation induced, and tectonic. *Journal of Geophysical Research* 106: 16525–16547.

Roeloffs EA (2006) Evidence for aseismic deformation rate changes prior to earthquakes. *Annual Review of Earth and Planetary Sciences* 34: 591–627.

Rogers G and Dragert H (2003) Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* 300: 1942–1943.

Sacks IS, Linde AT, Snoke JA, and Suyehiro S (1981) A slow earthquake sequence following the Izu–Oshima earthquake of 1978 In: Simpson DW and Richards PG (eds.) *Maurice Ewing Series*, Vol 4: *Earthquake Prediction: An International Review*, Washington, DC: AGU.

Sacks IS, Suyhiro S, Evertson DW, and Yamagishi Y (1971) Sacks–Evertson strainmeter: Its installation in Japan and some preliminary results concerning strainsteps. *Papers in Meteorology and Geophysics* 22: 195–208.

Sagiya T, Yoshimura A, Iwata E, *et al.* (1995) Establishement of permanent GPS observation network and crustal deformation monitoring in the southern Kanto and Tokai areas. *Bulletin of Geographical Survey Institute* 41: 105–118.

Sagiya T (2004) Interplate coupling in the Kanto District, central Japan, and the Boso Peninsula silent earthquake in May 1996. *Pure and Applied Geophysics* 161: 2327–2342.

Sato T, Imanishi K, Kato N, and Sagiya T (2004) Detection of a slow slip event from small signal in GPS data. *Geophysical Research Letters* 31: L05606 (doi:10.1029/2004GL019514).

Satake K, Wang K, and Atwater BF (2003) Fault slip and seismic moment of the 1700 Cascadia earthquake inferred from Japanese tsunami descriptions. *Journal of Geophysical Research* 108(B11): 2535 (doi:10.1029/2003JB002521).

Schmidt D (2006) The 2005 Cascadia ETS event inferred from PBO tensor strainmeters and GPS. *Geophysical Journal International Fall Meeting Supplement*, 87(52): T41A–1545.

Schmidt DA, Burgmann R, Nadeau RM, and d'Alessio M (2005) Distribution of aseismic slip rate on the Hayward fault inferred from seismic and geodetic data. *Journal of Geophysical Research* 110, (doi:10.1029/2004JB003397).

Schwartz SY and Rokosky JM (2007) Slow slip events and seismic tremor at circum-Pacific subduction zones. *Reviews of Geophysics* (in press).

Segall P, Desmarais EK, Shelly D, Miklius A, and Cervelli P (2006) Earthquakes triggered by silent slip events on Kilauea Volcano, Hawaii. *Nature* 442(7098): 71–74.

Shelly DR, Beroza GC, Satoshi I, and Nakamula S (2006a) Low-frequency earthquakes in Shikoku, Japan and their relationship to episodic tremor and slip. *Nature* 442(7099): 188–191.

Shelly DR, Beroza GC, and Ide S (2006b) Low frequency earthquake swarms and non-volcanic tremor under Shikoku, Japan. *Geophysical Journal International Fall Meeting Supplement*, 87(52): T41A–1546.

Shen Z, Wang Q, Burgmann R, Wan Y, and Ning J (2005) Pole-tide modulation of slow slip events at circum-Pacific subduction zones. *Bulletin of the Seismological Society of America* 95(5): 2009–2015.

Shennan I, Long AJ, Rutherford MM, Innes JB, Green FM, and Walker KJ (1998) Tidal marsh stratigraphy, sea-level change and large earthquakes- II: submergence events during the last 3500 years at Netarts Bay, Oregon, USA. *Quatenary Science Reviews* 17: 365–394.

Shibazaki B and Iio Y (2003) On the physical mechanism of silent slip events along the deeper part of the seismogenic zone. *Geophysical Research Letters* 30(9): 1489 (doi:10.1029/2003GL017047).

Smith SW and Wyss M (1968) Displacement on the San Andreas Fault subsequent to the 1966 Parkfield earthquake. *Bulletin of the Seismological Society of America* 58: 1955–1973.

Subarya C, Chlieh M, Prawirodirdjo L, *et al.* (2006) Plate-boundary deformation associated with the great Sumatra–Andaman earthquake. *Nature* 440: 46–51.

Suwa Y, Miura S, Hasegawa A, Sato T, and Tachibana K (2006) Interplate coupling beneath NE Japan inferred from three-dimensional displacement field. *Journal of Geophysical Research* 111: B04402 (doi:1029/2004JB003203).

Sylvester AG (1993) Investigation of nearfield postseismic slip following the Mw 7.3 Landers earthquake sequence of 28 June 1992, California. *Geophysical Research Letters* 20(11): 1079–1082.

Szeliga W, Melbourne T, Miller M, and Santillian V (2004) Southern Cascadia episodic slow earthquakes. *Geophysical Research Letters* 31: L16602 (doi:10.1029/2004GL020824).

Uchida N, Matsuzawa T, Igarashi T, and Hasagawa A (2003) Interplate quasi-static slip off Sanriku, NE Japan, estimated from repeating earthquakes. *Geophysical Research Letters* 30, (doi:10.1029/2003GL017452).

Ueda H, Ohtake M, and Sato H (2001) Afterslip of the plate interface following the 1978 Miyagi-Oki Japan earthquake, as revealed from geodetic measurement data. *Tectonophysics* 338: 45–57.

Uhrhammer R, Gee LS, Murray MH, Dreger D, and Romanowicz B (1999) The Mw 5.1 San Juan Bautista, California earthquake of 12 August 1998. *Seismological Research Letters* 70(1): 10–18.

Vidale JE and Shearer PM (2006) A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. *Journal of Geophysical Research* 111, (doi:10.1029/2005JB004034).

Vigny C, Simons WJF, Abu S, *et al.* (2006) Insight into the 2004 Sumatra–Andaman earthquake from GPS measurements in southeast Asia. *Nature* 436: 201–206.

Wald DJ and Heaton TH (1994) Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake. *Bulletin of the Seismological Society of America* 84(3): 668–691.

Wald DJ and Somerville PG (1995) Variable slip rupture model for the great 1923 Kanto, Japan earthquake: Geodetic and body waveform analysis. *Bulletin of the Seismological Society of America* 85: 159–177.

Wallace LM and Beavan J (2006) A large slow slip event on the central Hikurangi subduction interface beneath the Manawatu region, North Island, New Zealand. *Geophysical Research Letters* 33: L11301 (doi:10.1029/2006GL026009).

Wesson RL (1988) Dynamics of fault creep. *Journal of Geophysical Research* 93: 8929–8951.

Williams PL and Magistrale HW (1989) Slip along the Superstition Hills fault associated with the 24 November 1987 Superstition Hills, California, earthquake. *Bulletin of the Seismological Society of America* 79: 390–410.

Yagi Y, Kikuchi M, and Nishimura T (2003) Co-seismic slip, post-seismic slip, and largest aftershock associated with the 1994 Sanriku–haruka-oki, Japan, earthquake. *Bulletin of the Seismological Society of America* 30(22): 2177 (doi:10.1029/2003GL018189).

Yamamoto E, Matsumura S, and Ohkubo T (2005) A slow slip event in the Tokai area detected by tilt and seismic observation and its possible recurrence. *Earth, Planets and Space* 57: 917–923.

Yamanaka Y and Kikuchi M (2003) Source process of the recurrent Tokachi-Oki earthquake on September 26, 2003, inferred from teleseismic body waves. *Earth, Planets and Space* 55: e1–e24.

Yamanaka Y and Kikuchi M (2004) Asperity map along the subduction zone along northeastern Japan inferred from regional seismic data. *Journal of Geophysical Research* 109(7), (doi:1029/2003JB002683).

Yoshida S and Kato N (2003) Episodic aseismic slip in a two-degree-of-freedom block-spring model. *Geophysical Research Letters* 30(13): 1681 (doi:10.1029/2003GL017439).

Yoshida A, Hososno K, Tsukakoshi T, Kobayashi A, Takayama H, and Wiemer S (2006) Change in seismic activity in the Tokai region related to weakening and strengthening of the interplate coupling. *Tectonophysics* 417: 17–31.

4.16 Global Seismicity: Results from Systematic Waveform Analyses, 1976–2005

G. Ekström, Columbia University, Palisades, NY, USA

4.16.1 Introduction

Our knowledge of the geographical patterns of earthquake activity away from populated areas evolved with the global installation of seismographic stations during the first half of the twentieth century. In their classic 1941 paper, *Seismicity of the Earth*, and their 1949 book, Gutenberg and Richter (1941, 1949) documented the then-current understanding of seismic zones around the world, and some 25 years later the spatial features of earthquake activity they described were given a comprehensive explanation in the context of plate tectonics. Integral to this explanation were the development of a kinematic model of the earthquake source in terms of fault motion and techniques of inferring the geometry of faulting from features in seismic recordings, initially from the first motions of P-waves.

While the fundamental aspects of global seismicity are well understood, the determination of source characteristics for individual earthquakes (i.e., location, magnitude, and source mechanism) remains important. The details of the seismicity of different regions are highly variable, and neither seismicity patterns nor focal mechanisms can be predicted with much precision from a simple characterization in terms of convergent, divergent, or transcurrent plate motion. Higher-order descriptions of the deformation field are needed, and these are best constrained by comprehensive investigations of the patterns of earthquake activity and strain release.

Through the 1970s, the dominant mode of seismic recording was analog, and the determination of a single earthquake focal mechanism required a significant research effort. Access to the seismograms was the initial challenge, and any quantitative analysis required manual digitization of the seismograms. The low dynamic range of the analog recordings also generally restricted the analysis for teleseismic distances to earthquakes with magnitudes greater than 6. The development and deployment of digital instruments with low noise in the mid-1970s provided the initial conditions for a more comprehensive and quantitative analysis of global seismicity.

Concurrent with technical progress in the high-fidelity recording of seismic waves, theoretical and computational advances in the calculation of theoretical waveforms made possible new methods of source analysis, in particular those based on the direct comparison and correlation of observed and predicted long-period waveforms. A very successful approach was that developed by Dziewonski and his colleagues at Harvard University, who in 1981 published the centroid moment tensor (CMT) algorithm (Dziewonski *et al.*, 1981) for the systematic analysis of global and regional seismicity. The ongoing systematic application of the CMT method to data from the Global Seismographic Network (GSN) has led to the accumulation of the most comprehensive catalog of global earthquake focal mechanisms available, spanning the years 1976 to the present. This chapter reviews this systematic analysis effort and some aspects of global seismicity as reflected in the CMT catalog. The review is not comprehensive, and the references have been limited to those directly associated with the evolution of the CMT project. Section 4.16.2 provides a brief summary of the basic principles of the CMT method, as well as a description of how the algorithm has evolved since 1981. Section 4.16.3 presents a few representative maps and diagrams illustrating some of the primary characteristics of global seismicity. Section 4.16.4 describes recent discoveries of previously undetected

earthquakes and suggests some possible future directions for the monitoring and study of global seismicity.

4.16.2 The CMT Method

The CMT method (Dziewonski *et al.*, 1981) is based on the linear relationship that exists between the six independent elements of the zeroth-order moment tensor describing the earthquake source and the elastic vibrations of the Earth generated by the earthquake. This relationship holds as long as the earthquake source dimension is small relative to the wavelength of the seismic wave considered and the duration of the source is small relative to the period of the seismic wave. An additional component of the CMT algorithm is the consideration and estimation of the source centroid, which represents the spatial and temporal center of moment of the earthquake.

The CMT concept is general and can in principle be used with any recorded seismogram, given that the point-source approximation holds. A necessary condition is, of course, that synthetic waveforms corresponding to the observed seismograms can be generated with sufficient fidelity to allow a quantitative comparison. In the CMT algorithm, complete seismograms are compared in a least-squares sense, requiring that the phase and amplitude of the waveforms can be matched very well. In their original paper, Dziewonski *et al.* (1981) used normal-mode summation for a spherical Earth to generate the synthetic seismograms necessary to retrieve the 'moment tensor' elements and the earthquake centroid. They found that the body-wave portion of long-period ($T > 45$ s) seismograms could be fit well using spherical-Earth synthetic waveforms, but that surface waves, which are significantly distorted by propagation through Earth's laterally heterogeneous shallow structure, could not be fit well.

In a follow-up paper, Dziewonski and Woodhouse (1983) extended the CMT method to consider very-long-period ($T > 135$ s) surface waves in the inversion for the moment tensor. Following the development of tomographic models of the Earth's upper mantle, the calculation of synthetic seismograms in the CMT processing was modified to include corrections for laterally varying Earth structure (Dziewonski *et al.*, 1984; Woodhouse and Dziewonski, 1984). These corrections were particularly useful for improving the phase alignment of observed and synthetic waveforms, and resulted in improved moment tensor determinations.

Even with better tomographic models of the upper mantle, the prediction of intermediate-period (35–150 s) surface waves was not sufficiently good to allow direct comparison of observed and synthetic surface-wave phases at regional or teleseismic distances. However, following the development of global phase-velocity maps for Love and Rayleigh waves in this period range (Ekström *et al.*, 1997), Arvidsson and Ekström (1998) extended the CMT algorithm to include intermediate-period surface waves in the inversion for source parameters. This development, in turn, allowed the CMT algorithm to be applied to earthquakes of a smaller magnitude. Since 2004, the routine CMT analysis has included the fitting of intermediate-period surface-wave seismograms in the analysis, significantly increasing the number of earthquakes that can be successfully analyzed each year.

The routine determination of moment tensors has benefited from the increasing number of broadband seismographic stations deployed globally. The earliest year for which the CMT method has been applied systematically is 1976 (Ekström and Nettles, 1997). In that year, the experimental High-Gain Long-Period (HGLP) seismographic network was operating, as were a small number of Standard Research Observatory (SRO) and International Deployment of Accelerometer (IDA) stations. In total, fewer than a dozen stations were available for any single event. With the development first of the Global Digital Seismographic Network (GDSN) operated by the United States Geological Survey (USGS), and subsequently the GSN of the Incorporated Research Institutions of Seismology (IRIS) and the USGS, both the number of stations and the quality of the data have improved. These improvements have, in turn, led to better and more robust moment tensor determinations. In our current CMT analysis, seismograms from more than 100 stations are routinely analyzed for each earthquake.

4.16.3 Aspects of Global Seismicity

Figure 1 shows a map of focal mechanisms for shallow earthquakes in the CMT catalog for the period 1976–2005. While many focal mechanisms are obscured by focal mechanisms of later, nearby earthquakes, the map conveys some of the essential aspects of the pattern of global seismicity. The plate

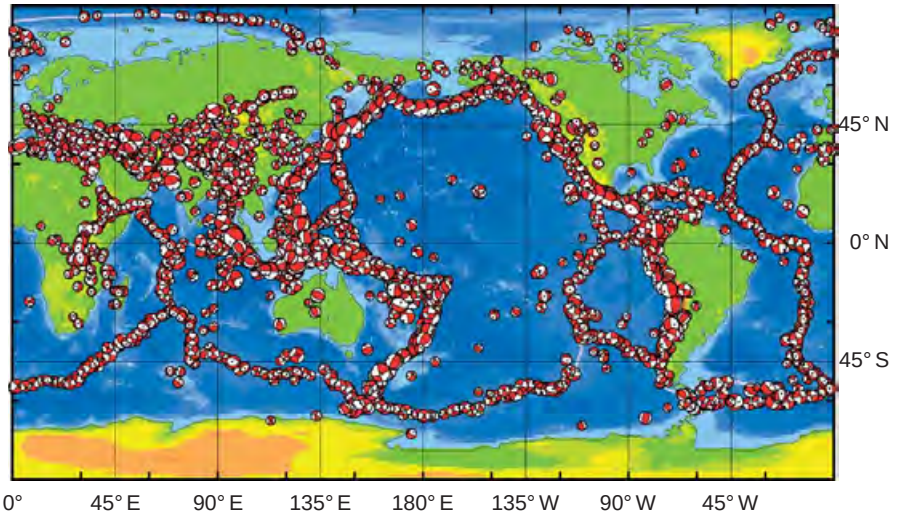


Figure 1 Map showing focal mechanisms for shallow earthquakes ($h < 70$ km) in the CMT catalog, 1976–2005. The earthquakes are shown in lower-hemisphere projection, and the diameter of the symbols increases linearly with the moment magnitude. Plate boundaries are shown, but are nearly entirely covered by focal mechanisms.

boundaries are clearly delineated by earthquake activity, except in areas of very slow plate motion, or near fast-spreading ridges, where the lithosphere is too weak to support earthquakes of significant magnitude. The consistency of focal mechanisms along the plate boundaries is also evident, particularly in subduction zones and along oceanic transforms. While some portions of the continents are devoid of $M > 5$ earthquakes, it is evident that over the 30-year period shown, many continental intraplate regions have experienced significant seismicity. The regional variability of seismic patterns (both in terms of frequency and style of faulting) within oceanic plates is also striking.

Seismically active plate boundaries involving at least one oceanic plate are normally characterized by a relatively narrow zone of tectonic deformation, reflected in a narrow zone of earthquakes. **Figure 2** shows focal mechanisms from the Alaska–Aleutian Islands subduction zone, an area of very active seismicity and the site of several $M \geq 7.5$ earthquakes since 1976. The vast majority of the earthquakes are consistent with the shallow underthrusting of the Pacific Plate beneath the North American Plate. In

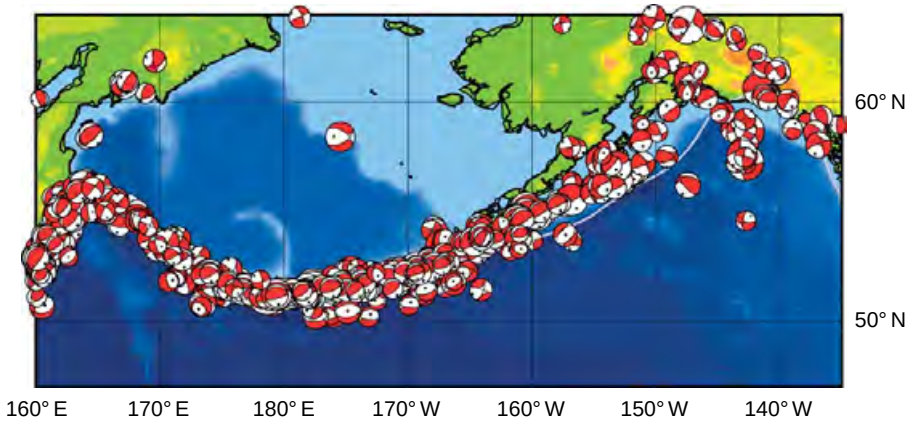


Figure 2 Map showing the Alaska–Aleutian subduction zone, with focal mechanisms for shallow earthquakes from the CMT catalog for the period 1976–2005. While seismic activity along the central and western part of the subduction zone is focused on the plate boundary, several intraplate earthquakes in the eastern zone indicate deformation within both the North American and Pacific Plates.

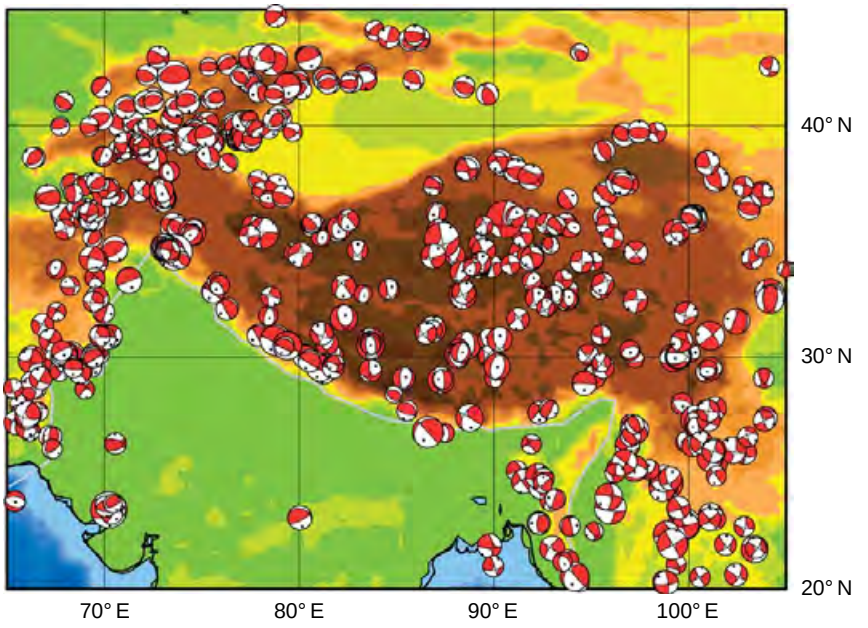


Figure 3 Map showing the India–Eurasia collision zone, with focal mechanisms for shallow earthquakes from the CMT catalog for the period 1976–2005. While earthquakes at the northern and southern edges of the Tibetan plateau show mainly reverse motion associated with north–south compression, most of the earthquakes on the plateau reflect east–west extension.

the western portion of the subduction zone, the greater predominance of strike-slip focal mechanisms reflects the increasing obliquity of plate motion towards the west. In the eastern portion of the subduction zone, the pattern of seismicity is more distributed, with earthquakes associated with the plate collision occurring both in the interior of Alaska and within the Pacific Plate.

The most dramatic example of geographically distributed seismicity occurring in response to plate convergence is that associated with the collision of India and Eurasia. Figure 3 shows focal mechanisms for shallow earthquakes in this zone. The geographical association of earthquakes with topography is evident, and the zone of activity extends more than 1000 km north from the Himalayan front. Within the Tibetan plateau, earthquake focal mechanisms primarily represent not the relative plate motion, but rather the strains associated with the gravity-controlled dynamics of the elevated plateau. The vast majority of focal mechanisms indicate strike-slip and normal faulting consistent with east–west extension.

Figure 4 shows the cumulative global moment of earthquakes since 1976. As is well known, most of the seismic moment (and corresponding strain) in a region occurs in the largest earthquakes, while small earthquakes, individually and collectively, account for only a small fraction of the cumulative moment. This division of cumulative seismic moment is illustrated by the contribution provided by earthquakes with $M < 6.5$ to the total, represented by the area

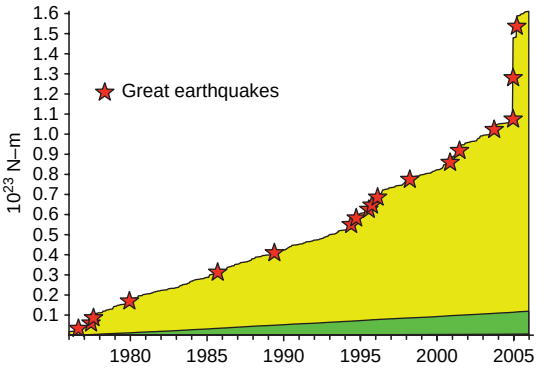


Figure 4 Cumulative moment of all earthquakes in the CMT catalog. Red stars indicate the times of earthquakes with moment magnitude $M \geq 8.0$. The field shaded yellow reflects the cumulative moment of earthquakes with $M \geq 6.5$. The green field reflects cumulative moment of earthquakes with $5.3 \leq M \leq 6.5$, and orange (barely visible) reflects earthquakes with $M \leq 5.3$. The contribution of the December 2004 Sumatra earthquake to the total cumulative moment is seen as the largest step in the curve, which would be even more prominent if the $M = 9.3$ estimate for the size of this event determined by Tsai *et al*. (2005) had been used.

shaded green in **Figure 4**. Over the 30-year period of the CMT catalog, 16 earthquakes have moment magnitudes of 8.0 or above. The largest earthquake to occur during the last 40 years is the 2004 Sumatra earthquake, which is easily identified in **Figure 4** with the largest step increase in cumulative moment. This earthquake is included with its point-source magnitude of 9.0; considering the spatial and temporal extent of the source in the CMT modeling leads to a moment magnitude of 9.3 (Tsai *et al.*, 2005), which would more than double the amplitude of the moment step in **Figure 4**.

The impact of the 2004 Sumatra earthquake can also be seen in **Figure 5**, which shows the cumulative number of earthquakes in the CMT catalog since 1976. While the primary trend in **Figure 5** is the gradual increase in the number of earthquakes with $M < 5.3$ that we have been able to analyze as a result of improvements in global station coverage and the CMT algorithm, it is also evident that the seismic activity in the months following the 2004 Sumatra earthquake is the most intense seen during the last 30 years.

Most earthquakes occur along plate boundaries at shallow depths, and this is also where a maximum in cumulative seismic moment is observed. **Figure 6** shows the depth distribution of cumulative seismic moment for the period 1976–2005. A minimum in moment occurs near 300 km depth, a second local minimum near 500 km depth, and a local maximum

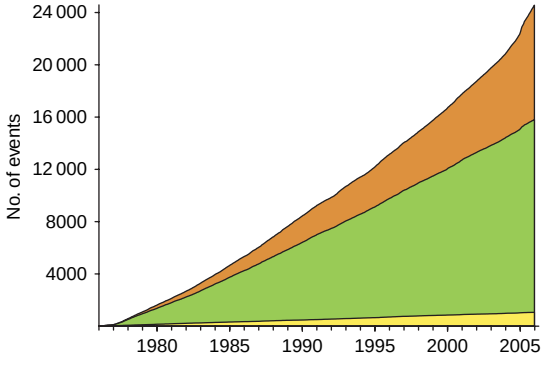


Figure 5 Cumulative number of earthquakes in the CMT catalog since 1976. Groups of earthquakes of different magnitudes are color coded as in **Figure 4** (orange, $M \leq 5.3$; green, $5.3 \leq M \leq 6.5$; yellow, $M \geq 6.5$). The largest increase in number of earthquakes in the catalog is seen for events with $M \leq 5.3$, especially since 2004 when intermediate-period surface waves were included in the analysis. The intense aftershock activity following the December 2004 Sumatra earthquake is reflected in the steep slope of the curve in the months following the earthquake.

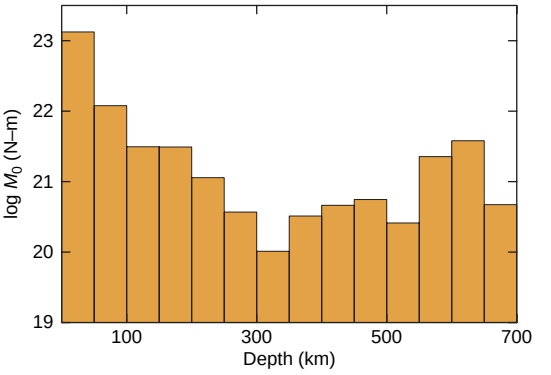


Figure 6 Graph showing the cumulative moment of earthquakes in the CMT catalog, 1976–2005, for different depths. The earthquakes have been binned in 50 km thick layers. Most of the cumulative moment is located in the top 50 km. A minimum occurs in the depth range 300–350 km, and a secondary minimum occurs below 500 km. A second maximum occurs at the bottom of the upper mantle, at 600–650 km.

at the bottom of the upper mantle at 650 km. While all deep earthquakes can be associated with current subduction or the likely location of recently subducted lithosphere, though not always unambiguously, the nature of seismic faulting at depth and its relationship to the physical state, stress, and strain of the mantle rock is not resolved.

Figure 7 shows an interesting example of seismic activity in the subducted slab beneath Chile and Argentina. **Figure 7(a)** shows, in map view, the distribution of earthquakes along the plate interface and in the subducted Nazca slab. One of the deepest earthquakes is located to the east of the adjacent earthquakes, and in cross-section (**Figure 7(b)**), it is clear that a spatial separation exists between the main group of deep events and this 'outlier earthquake'. The outlier also exhibits a compression axis with a different orientation than the adjacent events, which have focal mechanisms consistent with down-dip compression. The significance of outliers such as this one is hard to assess. If deep earthquakes are markers of subducted lithosphere, we must invoke some contortion of the subducting slab, or the existence of a second slab or slab fragment, in order to explain the existence of outlying earthquakes.

A further illustration of the complexity of deep seismicity is provided by the earthquakes beneath Fiji and Tonga, the world's most productive zone of deep seismicity. **Figure 8(a)** shows the earthquakes in this region in map view; **Figure 8(b)** shows the seismicity projected onto a vertical plane with a

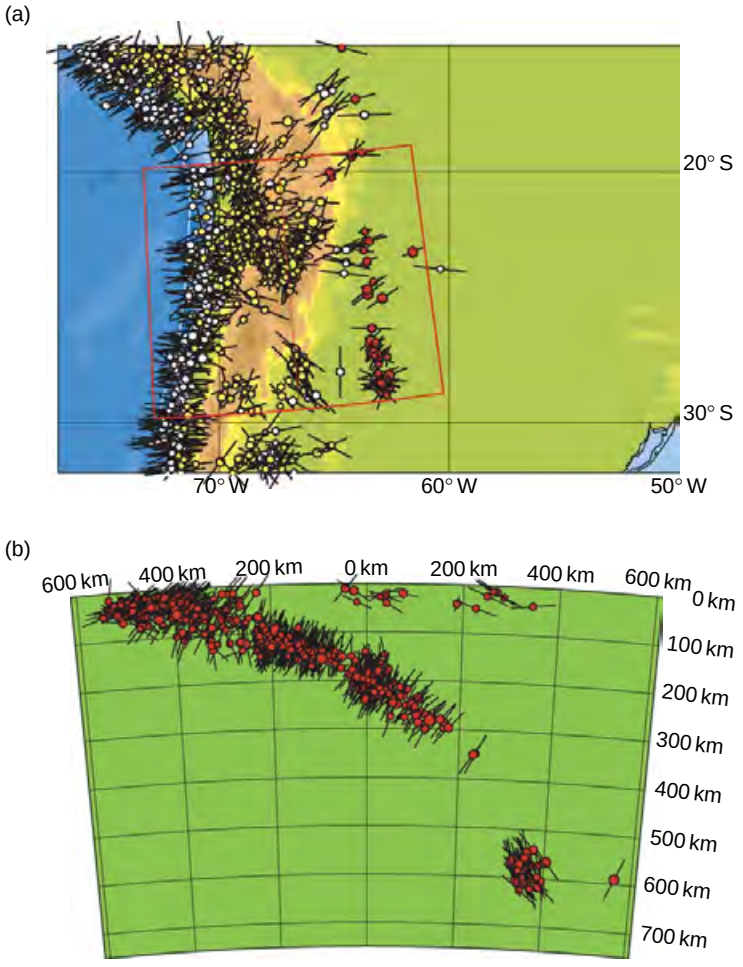


Figure 7 (a) Compression axes of CMT moment tensors in the western part of South America, projected into the horizontal plane. The earthquakes are color-coded with respect to depth, with white symbols for shallow earthquakes and red symbols indicating the deepest events. The red box indicates the area shown in cross-section in (b). The viewing direction is shown by the thin red line in the center of the box. (b) Vertical cross-section aligned with a viewing direction in the plane of the earthquakes associated with the subduction of the Nazca plate beneath South America. Compression axes of CMT focal mechanisms are shown, projected into the plane of the cross-section. At the bottom of the subduction zone, one outlying event has a compression axis that is rotated by 70° with respect to the down-dip slab direction suggested by the main seismicity at this depth. This event is also displaced horizontally by nearly 200 km from the inferred location of the subducted slab.

strike close to that of the deep seismicity. In this view, it is clear that the earthquakes show complex spatial patterns, including both lineations in the seismicity and large volumes without significant seismicity. The deepest earthquakes, toward the south, are surrounded by a large area lacking significant earthquakes during the last 30 years. Although most of the compression axes shown in **Figure 8(b)** are consistent with down-dip compression, there are several clear and systematic deviations from this basic pattern which require more detailed explanations in terms of the stresses and deformation in the subducted lithosphere.

4.16.4 Recent Discoveries and Future Directions

The study of global seismicity depends on the initial detection and location of seismic events. Several agencies are responsible for the monitoring of seismic activity on a global scale, in particular the National Earthquake Information Service (NEIS) of the USGS, the International Seismological Centre (ISC), and the International Monitoring System (IMS) in Vienna. All of these organizations rely primarily on high-frequency seismic signals for the detection of seismic events, leaving open the

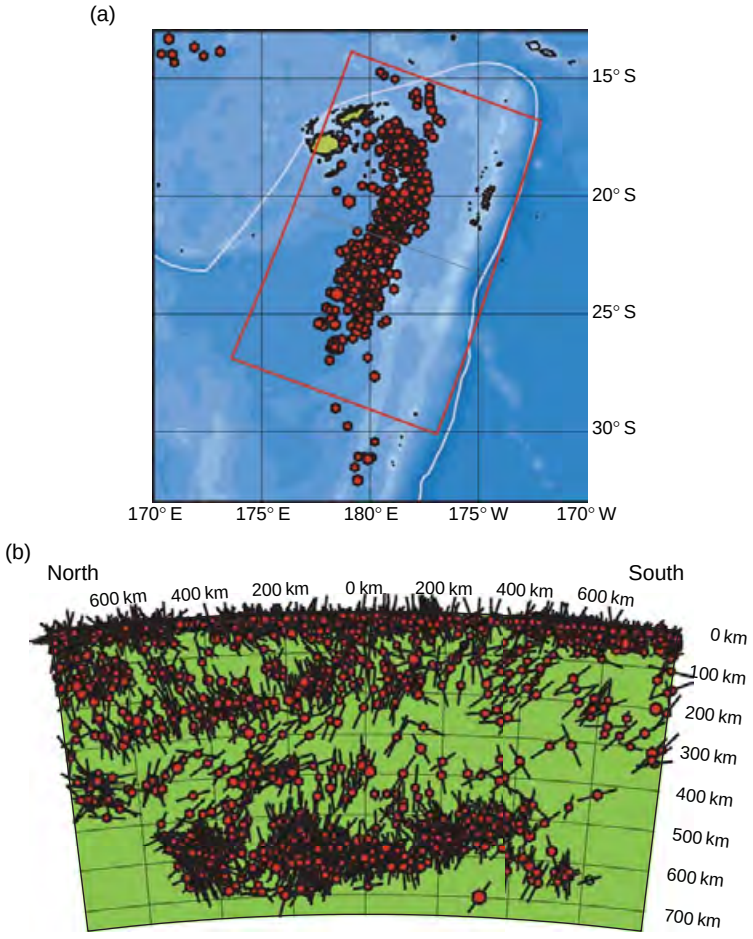


Figure 8 (a) Map view of deep seismicity beneath Fiji-Tonga. The red box indicates the area viewed in cross-section in the bottom panel. (b) Vertical cross-section aligned with a viewing direction perpendicular to the strike of the deep seismicity associated with the Tonga subduction zone. The compression axes of the focal mechanism are projected onto the vertical plane, and show a very complex pattern of internal deformation in the plane of the slab. Some spatial trends in the locations and stress axes can be followed for hundreds of kilometers along the slab.

possibility that earthquakes unusually deficient in radiated short-period energy may escape detection.

In recent work, we have applied a new method of array processing of continuous long-period vertical data from the GSN (Ekström *et al.*, 2003; Ekström, 2006) to search for sources of long-period Rayleigh waves. The algorithm was applied to data filtered in the period range 40–150 s, in which seismic noise is low and knowledge of Earth's structure is sufficiently good to allow accurate deconvolution of surface-wave dispersion effects for arbitrary ray paths. Following the application of the detection and location algorithm to data for the period 1993–2003, we found nearly 25 000 seismic sources with magnitudes $M > 4.5$, where the magnitude M is a good estimate of the moment magnitude. The vast majority of these events are standard earthquakes that are also found in the event catalogs of the NEIC, ISC, and the IMS. Approximately 5% of the events are, however, not listed in any of these catalogs, and are therefore considered 'new'.

Figure 9 shows the locations of new earthquakes for the period 1993–2003. Most of the events occur in the Southern Hemisphere along the ridge-transform plate boundaries. Very few events are located in subduction zones. The events near plate boundaries may be earthquakes that have unusual rupture characteristics, or they may have gone undetected for other reasons (e.g., high microseismic noise or nodal radiation to many stations). Several other events are located in areas where earthquakes of this magnitude ($M > 4.5$) are unexpected, or where they should have

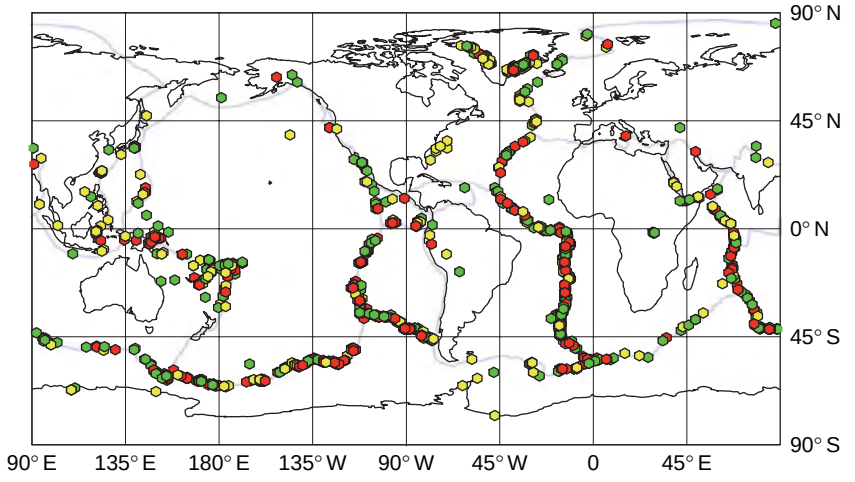


Figure 9 Map showing the locations of 1301 previously undetected earthquakes for the period 1993–2003 located using the surface-wave algorithm of Ekström (2006). The smallest earthquake has an estimated magnitude of $M = 4.6$, and the largest, $M = 5.9$. The quality of detection is indicated by the color of the symbol, with red indicating the best quality, green very good quality, and yellow good quality.

been detected if they represent standard earthquake sources. Most prominent in this category are the events on Greenland, which we have identified with a new phenomenon called 'glacial earthquakes' (Ekström *et al.*, 2003, 2006). These events radiate surface waves with polarities and amplitudes consistent with a mass-sliding source, rather than with tectonic stress release. **Figure 10** shows an example of a teleseismic record section from one of the glacial earthquakes, clearly demonstrating that signals for

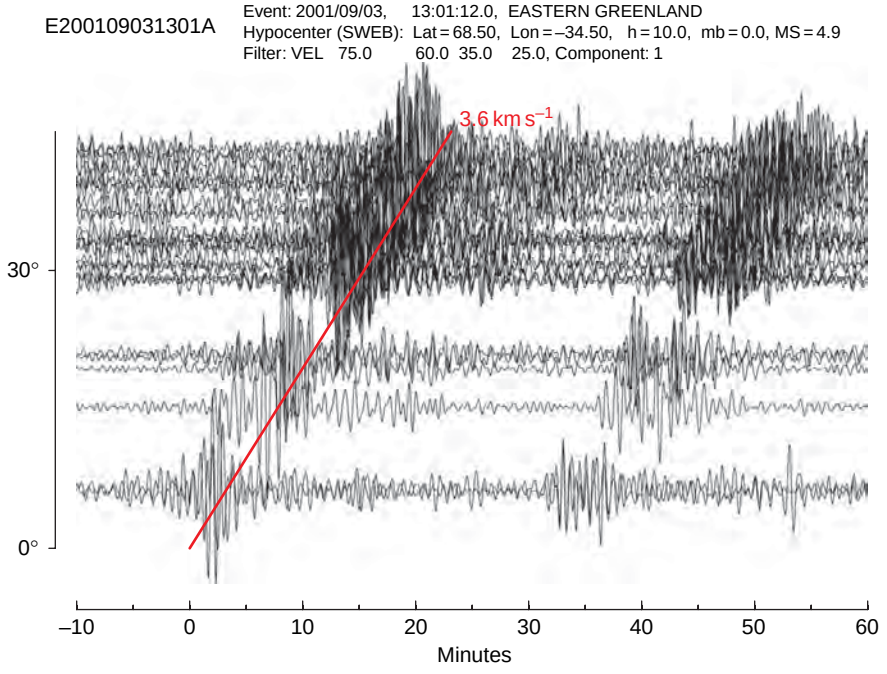


Figure 10 Record section for a recently detected glacial earthquake, showing the long-period Rayleigh wave arrivals at global stations. Approximately 30 min after the main Rayleigh wave arrivals (which triggered the detection), a second arrival is seen with the same moveout, suggesting that a second, smaller event occured in the same location.

these events are large and can easily be identified at long periods. Several events off the east coast of the United States may represent submarine land-slides or slumps. Other groups of new events are spatially associated with volcanos, and yet others (e.g., in the Middle East and Central Asia) lack, at this time, an obvious association or explanation.

These findings, and other recent discoveries of new events and processes that generate seismic radiation, indicate that the spectrum of seismic sources is richer than we had previously realized. These new events may be important for our understanding not only of lithospheric tectonics, but potentially also of other phenomena such as magma motion and subaereal and submarine mass wasting. With the aid of the powerful infrastructure provided by the GSN, and by many new regional broadband seismographic networks, it seems likely that a focused effort to investigate the global seismic wave field in the period range 20–40 s for seismic events would be very fruitful. Events with $M > 4.0$ routinely generate large surface waves in this period range, and it seems possible that, with recent advances in the characterization of crustal structure and shallow upper-mantle elastic and anelastic structure, these waves could be properly migrated and stacked for array processing and event detection.

Acknowledgments

The author thanks Adam Dziewonski, the founder and for many years the leader of the CMT Project, and Meredith Nettles, the third current member of the CMT Project team. The seismic waveforms used in the CMT Project come primarily from stations of the GSN operated by IRIS and the USGS; the waveforms are distributed by the IRIS Data Management Center. Additional waveforms come from several other networks, including the Geoscope, Mednet, GEOFON, and the US and Canadian national seismic networks. We are grateful to everyone involved in the collection and distribution of data from these networks. The CMT catalog and associated results are accessible at the website www.globalcmt.org. The CMT Project has been continuously funded by the National Science Foundation since its inception, and is currently supported by award EAR-0639963.

References

Arvidsson R and Ekström G (1998) Global CMT analysis of moderate earthquakes, $M_w \geq 4.5$, using intermediate period surface waves. *Bulletin of the Seismological Society of America* 88: 1003–1013.

Dziewonski AM, Chou T-A, and Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research* 86: 2825–2853.

Dziewonski AM, Franzen JE, and Woodhouse JH (1984) Centroid-moment tensor solutions for July–September 1983. *Physics of the Earth and Planetary Interiors* 34: 1–8.

Dziewonski AM and Woodhouse JH (1983) An experiment in the systematic study of global seismicity: Centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981. *Journal of Geophysical Research* 88: 3247–3271.

Ekström G (2006) Global detection and location of seismic sources by using surface waves. *Bulletin of the Seismological Society of America* 96: 1201–1212.

Ekström G and Nettles M (1997) Calibration of the HGLP seismograph network and centroid-moment tensor analysis of significant earthquakes of 1976. *Physics of the Earth and Planetary Interiors* 101: 219–243.

Ekström G, Nettles M, and Abers GA (2003) Glacial earthquakes. *Science* 302: 622–624.

Ekström G, Nettles M, and Tsai VC (2006) Seasonality and increasing frequency of Greenland glacial earthquakes. *Science* 311: 1756–1758.

Ekström G, Tromp J, and Larson EWF (1997) Measurements and global models of surface wave propagation. *Journal of Geophysical Research* 102: 8137–8157.

Gutenberg B and Richter CF (1941) Seismicity of the Earth. *Geological Society of America Special Paper 34*. New York: Geological Society of America.

Gutenberg B and Richter CF (1949) *Seismicity of the Earth*. Princeton: Princeton University Press.

Tsai VC, Nettles M, Ekström G, and Dziewonski AM (2005) Multiple CMT source analysis of the 2004 Sumatra earthquake. *Geophysical Research Letters* 32: L17304 (doi:10.10292005GL023813).

Woodhouse JH and Dziewonski AM (1984) Mapping the upper mantle: Three dimensional modelling of earth structure by inversion of seismic waveforms. *Journal of Geophysical Research* 89: 5953–5986.

4.17 Tsunamis

K. Satake, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

4.17.1 Introduction

A tsunami is an oceanic gravity wave generated by submarine or coastal geological processes such as earthquakes, landslides, or volcanic eruptions. Most tsunamis are caused by large shallow earthquakes along subduction zones. Tsunami is a Japanese word meaning 'harbor wave', and is sometimes called seismic sea wave or, erroneously, tidal wave. Tsunamis are usually small and barely noticed in deep ocean, but become large and cause damage when they approach coasts or harbors. The recent 2004 tsunami in the Indian Ocean demonstrated that a tsunami can travel at trans-oceanic distances and cause damage far from the source.

Tsunamis have many aspects that are studied by researchers in various fields. Their generation is related to geological processes. The propagation and offshore observation are the subjects of oceanography, and basic hydrodynamics are needed to understand the characteristics. The coastal behaviors, such as run-up on beaches or resonance in bays, are mostly studied by coastal engineers. For tsunami hazard reduction, officials in local and national governments, civil defense, and emergency managements are responsible and play important roles.

In this chapter, the wide range of tsunami phenomena is reviewed. Section 4.17.2 describes the tsunami caused by the 2004 Sumatra–Andaman earthquake, the most recent and the worst tsunami disaster that affected the countries around the Indian Ocean. Other examples of recent, historical and prehistoric tsunamis are reviewed in Section 4.17.3. Section 4.17.4 reviews the tsunami observation systems: instrumental measurements, field surveys, and studies of historical and geological data. The tsunami magnitude scales and catalogs are also described. Section 4.17.5 reviews the tsunami generation, by earthquakes, landslides, and volcanic eruptions, with some emphasis on the studies of tsunami sources. Section 4.17.6 discusses the tsunami propagation, from its hydrodynamic background to numerical computations. Section 4.17.7 discusses the tsunami warning systems, both basin-wide and local, and hazard mitigation methods such as probabilistic hazard assessments and hazard maps.

4.17.2 The 2004 Indian Ocean Tsunami

A giant earthquake ($M_w = 9.1$–9.3) occurred off Sumatra Island on 26 December 2004 (Lay *et al.*, 2005; Stein and Okal, 2005). This earthquake generated tsunami which devastated the shores of Indian Ocean (**Figure 1**); more than 200 000 people lost their lives. The number of victims, dead and missing together, is the largest in Indonesia (163 795), followed by Sri Lanka (35 399), India (16 389), and Thailand (8 345) (International Federation of Red Cross and Red Crescent Societies, 2005). Within 30 min of the earthquake, the tsunami devastated Banda Aceh and other coastal villages of Sumatra Island (Borrero, 2005a). The tsunami arrived at

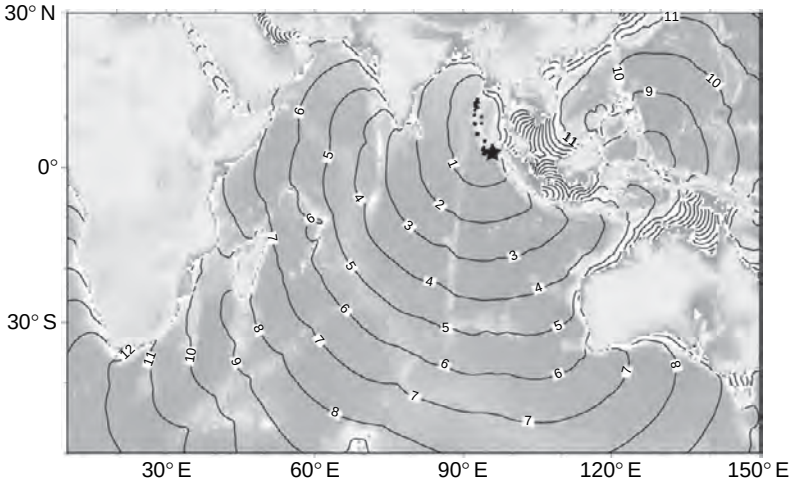


Figure 1 Tsunami propagation in the Indian Ocean. Each curve indicates wavefront at 1 h increments. Black star and circles indicate the epicenters of mainshock and aftershocks within 1 day, respectively (data by USGS).

Thai, Sri Lankan, and Indian coasts in about 2 h after the earthquake. The tsunami propagated farther and arrived at the eastern coast of Africa several hours after the earthquake to cause 298 fatalities in Somalia. The tsunami was also observed in the Atlantic and Pacific Oceans (Titov *et al.*, 2005a).

To document the 2004 tsunami, many scientists from all over the world visited the affected coasts (e.g., Borrero, 2005b; Liu *et al.*, 2005; Satake *et al.*, 2006a; Tsuji *et al.*, 2006). The measured tsunami heights in Sumatra Island, particularly in the Aceh province, were mostly larger than 20 m with the maximum above 30 m. The tsunami heights along the Andaman Sea coast were highly variable; 5–15 m in Thailand (**Figure 2**) but less than 3 m in Myanmar. The tsunami heights were up to 5 m on India's Andaman Islands. In Sri Lanka, the tsunami heights were 5–15 m. The tsunami height distribution is consistent with the damage distribution and indicates that the source of the large and damaging tsunami was concentrated in the southern 700 km section of the aftershock zone. Sedimentological characteristics of the modern tsunami deposits were examined (Jaffe *et al.*, 2006; Moore *et al.*, 2006), with hopes that they provide valuable information for the studies of paleo-tsunami deposits.

The tsunami was instrumentally recorded on coastal tide gauges in Indian Ocean (Merrifield *et al.*, 2005; Nagarajan *et al.*, 2006). This tsunami was also recorded by other instruments, some were for the first time. The tsunami propagation in deep water was

Figure 2 Tsunami damage at Ban Nam Ken, about 100 km north of Phuket, Thailand.

captured by satellite altimeter (Gower, 2005). The tsunami signal was also recorded on horizontal components of broadband seismic stations near coasts (Yuan *et al.*, 2005) or hydrophones (Hanson *et al.*, 2007).

Seismological analyses of the Sumatra–Andaman earthquake indicate that the source region extended from off Sumatra through the Nicobar to Andaman Islands with the total length of 1200–1300 km, and the total duration of rupture was about 500 s (Ishii *et al.*, 2005; Tsai *et al.*, 2005). The fault slip was the largest near off the northern Sumatra, followed by off Nicobar Islands (Ammon *et al.*, 2005). Fault slip around Andaman Islands was estimated to be small from seismological analysis and speculated to be slow (Bilham, 2005; Lay *et al.*, 2005).

Analyses of satellite images indicate that the sea level changes were observed in Sumatra, Nicobar, and Andaman Islands, suggesting that the source length was as long as 1600 km (Meltzner *et al.*, 2006; Tobita *et al.*, 2006). Field investigations indicate that the largest subsidence was 3 m at Great Nicobar Island, while the small islands west off Middle Andaman uplifted about 1.5 m (Kayanne *et al.*, 2007; Malik and Murty, 2005). Tide gauge record at Port Blair in South Andaman shows a slow subsidence with a duration of at least 15 min but possibly less than 35 min (Singh *et al.*, 2006).

The tsunami source was estimated from tsunami arrival times as 600–700 km (Lay *et al.*, 2005), but later revised to be ~900 km (Neetu *et al.*, 2005) extending to the south of Andaman Islands. The satellite altimeter data support longer, more than 1000 km long, tsunami source (Fine *et al.*, 2005b; Song *et al.*, 2005; Hirata *et al.*, 2006). The tide gauge data do not support tsunami source beneath Andaman Islands (Tanioka *et al.*, 2006; Fujii and Satake, 2007), although Piatanesi and Lorito (2007) estimated slip on the deeper part of slab, east of Andaman Islands. Joint inversion of tide gauge and satellite data prefer the source lengh of ~900 km (Fujii and Satake, 2007) (**Figures 3** and **4**). Seno and Hirata (2007) speculated that the fast rupture was followed by slow slip at shallower part near the trench.

The source of the previous 1881 earthquake around Car Nicobar was studied from tide gauge records around Bay of Bengal (Ortiz and Bilham, 2003). The potential tsunami hazard in Indian Ocean was pointed out before the 2004 tsunami, with the computer simulation of the 1833 earthquake off Sumatra (Cummins, 2004). Future tsunami forecast models in the region were examined by Geist *et al.* (2007).

4.17.3 Other Examples

In this section, other examples of tsunamis are briefly described in reverse-chronological order. Their sources are shown in **Figure 5**. We start with the recent and historical tsunamis that were instrumentally recorded, then prehistoric tsunamis that were geologically recorded.

4.17.3.1 Historical Tsunamis

4.17.3.1.1 1998 Papua New Guinea tsunami

Another recent tsunami hazard was from an earthquake ($M_w = 7.1$) along the northern coast of New Guinea Island on 17 July 1998. The tsunami heights were as large as 15 m around Sissano Lagoon near the epicenter, and the total casualties were reported to be more than 2000. The large tsunami (>5 m), however, was limited to a small region (~40 km) along the coast (Kawata *et al.*, 1999) and the amplitudes on tide gauges around Japan are <10 cm (Satake and Tanioka, 2003), suggesting a local effect of tsunami generation or amplification. Post-tsunami marine surveys revealed bathymetry features that focused tsunami energy toward Sissano Lagoon and possible sources of submarine landslide (Tappin *et al.*, 1999). Numerical simulations show that a source in addition to the earthquake fault motion, probably from submarine landslides, is needed to reproduce the locally large tsunami heights (Matsuyama *et al.*, 1999; Geist, 2000; Titov and Gonzalez, 2001; Synolakis *et al.*, 2002; Satake and Tanioka, 2003).

4.17.3.1.2 1993 Hokkaido tsunami

The most recent tsunami disaster in Japan was that in Okushiri Island in Japan Sea, caused by the Southwest Hokkaido earthquake on 12 July 1993 ($M_w = 7.8$). The tsunami attacked the island within a few minutes after strong ground shaking, and caused more than 200 casualties. The tsunami heights were 5–10 m around the island, while maximum tsunami run-up of more than 30 m occurred near a small valley. Numerical computations of tsunamis were made to reproduce the measured tsunami heights, recorded tsunami waveforms on tide gauges, as well as seismic or geodetic data (Shuto and Matsutomi, 1995; Takahashi *et al.*, 1995; Tanioka *et al.*, 1995; Titov and Synolakis, 1997).

4.17.3.1.3 1960 Chilean tsunami

The largest ($M_w = 9.5$) earthquake in the twentieth century occurred offshore southern Chile on 22 May 1960. The tsunami caused damage on the Chilean coast, including more than 1000 casualties, then propagated across the Pacific Ocean. About 15 h later, the tsunami reached the Hawaiian Islands and took 61 lives. It continued across the Pacific, reaching Japan in about 23 h, and caused 142 casualties. The large (~5 m) tsunami in Japan was due to both focusing and resonance effects. Because of the sphericity of the Earth and bathymetry in the Pacific Ocean, energy was focused toward the Japanese Islands (Satake, 1988; Shuto, 1991). Some bays along the Pacific coast of Japan have a shape whose characteristic period coincided with the predominant period (about 1 h) of the trans-Pacific tsunami (Abe, 2000). This tsunami motivated to form an international tsunami warning system in the Pacific.

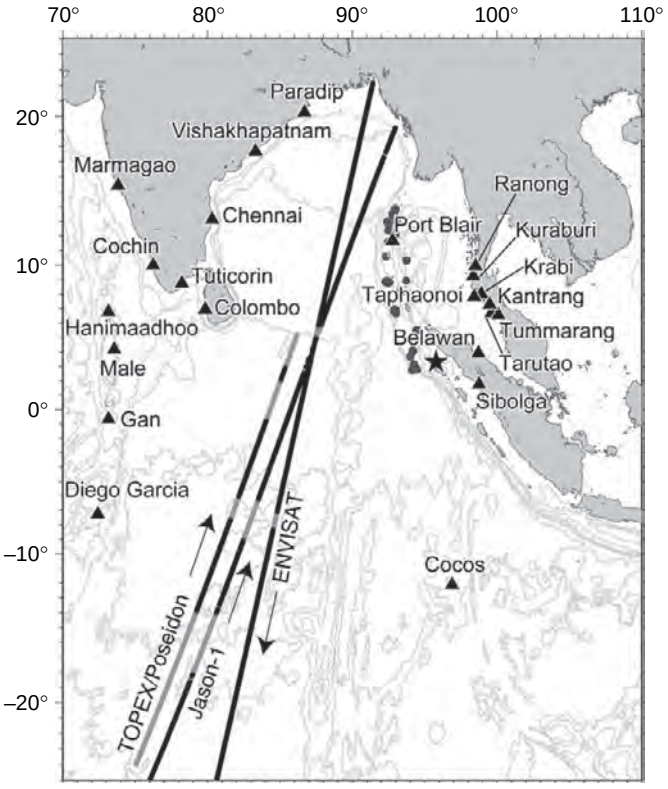


Figure 3 Locations of instrumental observation for the 2004 tsunami: tide gauge stations (triangles) and satellite path (solid line). From Fujii Y and Satake K (2007) Tsunami source model of the 2004 Sumatra–Andaman earthquake inferred from tide gauge and satellite data. *Bulletin of the Seismological Society of America* 97: S192–S207.

4.17.3.1.4 1958 Lituya Bay tsunami

The largest water run-up was recorded in Lituya Bay, Alaska. A large strike-slip earthquake ($M_w = 7.9$) on 10 July 1958 triggered a huge rockslide (700–900 m wide, 1000 m long, and 90 m thick, or the total volume of about 3×10^7 m^3) in this 10 km long bay. The rockslide generated water waves that surged up the opposite slope, and stripped the trees up to 520 m altitude (Miller, 1960; Lander and Lockridge, 1989). The large waves were confined in the bay; hence, this is not considered as a typical tsunami.

4.17.3.1.5 1946 Aleutian tsunami

The Aleutian earthquake of 1 April 1946 was very unusual; it generated huge tsunamis in the Aleutians and Hawaii, even though its surface wave magnitude M_s was only 7.4. This is a typical example of 'tsunami earthquake', which generates much larger tsunamis than expected from seismic waves (Kanamori, 1972). The Scotch Cap lighthouse on a 10 m cliff on Unimak Island, about 100 km from the epicenter, was totally demolished by the tsunami (Lander and Lockridge, 1989). The tsunami traveled southward through the Pacific Ocean and reached the Hawaiian Islands in the morning of 1 April. Without any warning, the tsunami took 159 lives in the Hawaiian Islands. A Pacific tsunami warning system was introduced after this tsunami. Recent studies indicate that slow earthquake rupture process and large submarine landslide are responsible to far-field and near-field tsunamis, respectively (Okal *et al.*, 2002, 2003).

4.17.3.1.6 1896 Sanriku tsunami

The 1896 Sanriku earthquake is another example of 'tsunami earthquake' (Kanamori, 1972; Tanioka and Satake, 1996b). The earthquake ($M_s = 7.2$) was not great in size and the ground shaking was weak (2–3 on the Japanese intensity scale, corresponding to IV–V on modified Mercalli scale). However, the following tsunami was very destructive; the run-up height was >10 m for about 200 km of Sanriku coast with the maximum of 38 m, and resulted 22 000 casualties, the worst tsunami hazard in Japan in the last few centuries.

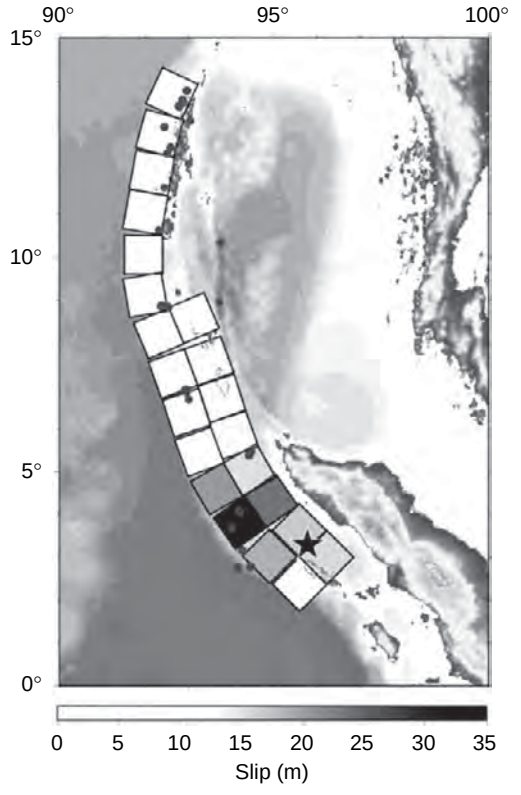


Figure 4 Slip distribution on fault for the 2004 Sumatra–Andaman earthquake by an inversion of tsunami waveforms recorded on tide gauges and sea surface height measured by satellite. From Fujii Y and Satake K (2007) Tsunami source model of the 2004 Sumatra–Andaman earthquake inferred from tide gauge and satellite data. *Bulletin of the Seismological Society of America* 97: S192–S207.

4.17.3.1.7 1883 Krakatoa tsunami

The 1883 eruption of Krakatoa, Indonesia, is one of the largest volcanic eruptions recorded in human history. The sound waves generated by the eruption were heard over 4000 km across the Indian Ocean. The eruption refaced the island of Krakatoa; the northen portion of the island disappeared and a caldera up to 270 m deep replaced it. The eruption also generated tsunamis that destroyed nearby coastal villages and took about 34 000 lives (Simkin and Fiske, 1983; Nomanbhoy and Satake, 1995). The maximum tsunami height along the Sunda Strait was about 15 m. Tide gauge disturbances were observed in India, New Zealand, and even at San Francisco, but those at very far places are interpreted as sea waves coupled by air waves from the eruption (Press and Harkrider, 1966; Pelinovsky *et al.*, 2005).

4.17.3.2 Prehistoric Tsunamis

4.17.3.2.1 1700 Cascadia tsunami

Geological and historical studies show the evidence of past earthquakes and tsunamis. Evidence of a great prehistoric earthquake that occurred along the Cascadia subduction zone has been recently found both near and far from the source (Atwater *et al.*, 2005). Along the Pacific coast of North America, geological and tree-ring evidence indicates that coseismic subsidence took place about 300 years ago. In addition, tsunami deposits found throughout the Pacific Northwest indicate that this earthquake

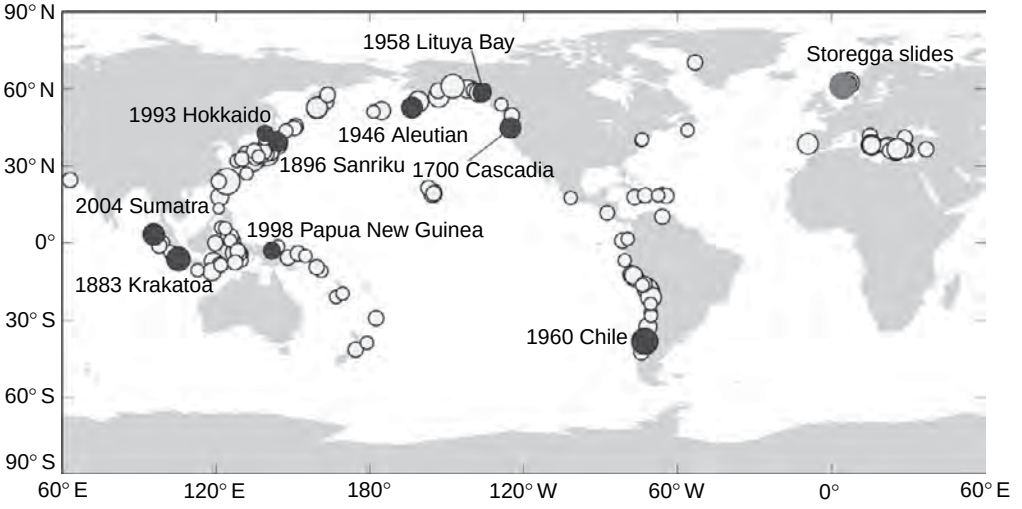


Figure 5 Distribution of tsunamis in the world. Open circles are historical events with intensity ≥2 and validity ≥3 from Historical Tsunami Data Base (see Section 4.17.4.6) (Institute of Computational Mathematics and Mathematical Geophysics 2005). Solid circles are those described in Sections 4.17.2 and 4.17.3.

generated a sizable tsunami. In Japan, historical records show a widespread tsunami with heights of 1–5 m struck the Pacific coast. From these records, the date and size of the earthquake was inferred to be 26 January 1700 and M_w = 8.7–9.2, respectively (Satake *et al.*, 1996, 2003).

4.17.3.2.2 Hokkaido tsunami in seventeenth century

Tsunami deposits show that the southern Kuril trench repeatedly produced tsunamis larger than those recorded in the last few centuries in Japan. Prehistoric sand sheets extend as much as 3 km inland along the 200 km of Hokkaido's Pacific coast, where the recent tsunamis from 1952 to 2003 Tokachi-oki or 1894 to 1973 Nemuro-oki earthquakes ($M_w \sim 8$) penetrated less than 1 km from the coast. The time intervals between the extensive sand sheets average about 500 years, and the most recent event occurred in the seventeenth century. These anomalous tsunamis, characterized by large inundation area and long recurrence interval, are best explained by earthquakes that rupture multiple segments, including Tokachi-oki and Nemuro-oki, of the Kuril subduction zone (Nanayama *et al.*, 2003; Satake *et al.*, 2005b).

4.17.3.2.3 Tsunami from the Storegga slide

The Storegga slide in the Norwegian Sea is the largest submarine landslide ever mapped. The Storegga slide is estimated to have taken place thrice: once at 30–50 ka and twice at 6–8 ka, with the total volume of $5.6 \times 10^3\,km^3$ (Bugge *et al.*, 1988). The gigantic submarine landslide probably caused a tsunami. Tsunami deposits inferred from this event have been found on the coasts of Scotland, Iceland, and Norway (Smith *et al.*, 2004; Bondevik *et al.*, 2005).

4.17.4 Tsunami Observations

4.17.4.1 Instrumental Measurements

4.17.4.1.1 Coastal tsunami measurements

Coastal sea level has been measured for more than a century by tide gauges at ports and harbors throughout the world. The classic tide gauge is a mechanical type set in a tide or stilling well. Sea level change is detected by a float in the well. Because typical period of ocean tides is several hours or longer, sea level changes inside the well can be assumed to be the same as that outside the well. For tsunamis with shorter period (less than 5 min), however, the system does not always follow the actual sea level change (Satake *et al.*, 1988).

More recent gauges include bubbler, acoustic, and laser gauges. The bubbler gauge is a pneumatic or gas-purged pressure gauge system. Compressed air or nitrogen gas is purged into a tube with an orifice, and its pressure is measured. As the gas slowly escapes from the orifice, the gas pressure is considered to be equal to water pressure, then converted to water level. The acoustic gauge sends ultrasonic waves in a tube that are reflected at the sea level. Sea level changes can be measured from the travel times of the acoustic wave (**Figure 6**). Laser gauge employs the same principle.

The Global Sea Level Observation System (GLOSS), an international program, has been formed to establish high-quality global sea level network for climate change, oceanography, and coastal research (Woodworth and Player, 2003). Many GLOSS data are available in real time through the Internet. Because the main purpose of the sea level measurements is to monitor ocean tides or long-term sea level change, the typical sampling interval is not always short enough for tsunami monitoring. For tsunami measurements, high sampling rate, at least 1 min, is required.

4.17.4.1.2 Offshore tsunami measurements

As tsunami becomes slow but large and complex near the coast, it would be ideal to detect its arrival offshore. Offshore wave gauges and GPS tsunami gauges have been developed and installed around Japan. Among about 50 acoustic wave gauges outside harbors at the water depths of tens of meters, about 10 recorded the tsunami from the 2003 Tokachi-oki earthquake (Nagai *et al.*, 2004). A GPS tsunami gauge system has been installed about 10 km offshore Cape Muroto, Southwestern Japan, where the water depth is about 100 m (Kato *et al.*, 2000, 2005). The gauge uses Real-Time Kinematic GPS technique to estimate the location and altitude of water surface. The sampling interval is 1 s with an accuracy of a few cm. Because it measures water surface movements, it is free from seismic waves unlike bottom pressure gauges (see the next section), and can also be used to monitor wave heights during storms.

4.17.4.1.3 Deep ocean measurements

In the deep ocean, tsunami waveforms are expected to be simpler, free from coastal topographic effects, although their amplitudes are smaller. Deep ocean

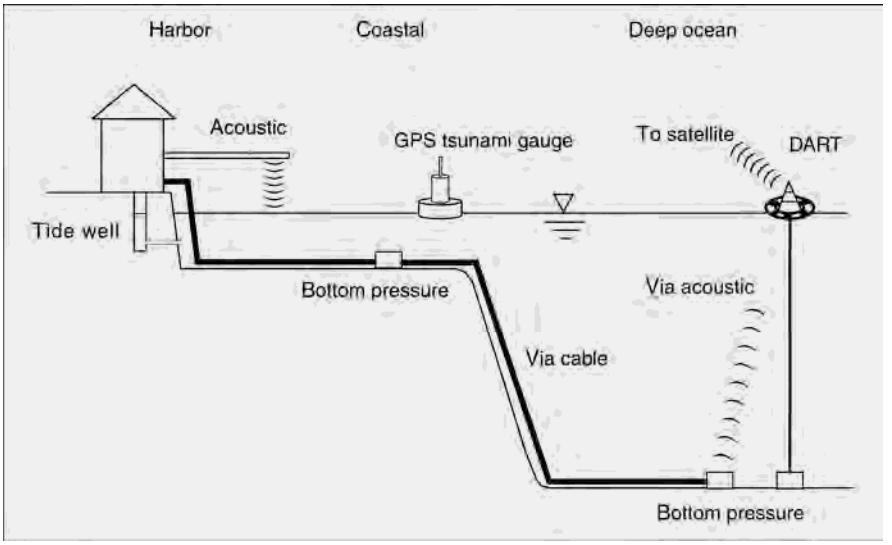


Figure 6 Types of instruments.

measurements of tsunamis have been made by using bottom pressure gauges for early detection and warnings of tsunami. A quartz crystal sensor is typically used as a pressure sensor. The sensor provides accuracy of 1 cm in sea level at a 6000 m depth. Because the pressure transducer is very sensitive to temperature, measurements and correction for temperature are essential. In addition, because the sensor is located at the ocean bottom, it also records ground motion and works as an ocean bottom seismograph. While the frequencies are often different between seismic and tsunami waves, it is necessary to separate them.

Around Japan, more than 10 bottom pressure gauges are installed on six submarine cables (Okada, 1995; Iwasaki *et al.*, 1997; Hirata *et al.*, 2002). Tsunamis from relatively small tsunamigenic earthquakes, the 2004 Off Kii Peninsula earthquakes (M 6.9 and 7.4), were recorded on a GPS tsunami gauge and eight bottom pressure gauges, as well as coastal tide gauges located south of Honshu and Shikoku (Satake *et al.*, 2005a). The US system, called Deep-ocean Assessment and Reporting of Tsunamis (DART) or simply Tsunameter, sends signals to surface buoy via acoustic telemetry in the ocean, then via satellites to land station in real time (Gonzalez *et al.*, 2005).

4.17.4.1.4 Satellite observations

Four satellites captured the tsunami propagation across the Indian Ocean on 26 December 2004 (**Figure 3**): Jason-1 and TOPEX/Poseidon at about 2 h after the earthquake, Envisat at 3 h 15 min, and US Navy's GEOSAT follow-on at 7 h (Gower, 2005). The maximum sea surface height (SSH) of the tsunami was about 0.8 m in the Indian Ocean. The accuracy of satellite altimetry is better than 5 cm and its spatial resolution is approximately 15 arc-min (~27 km). Numerical simulation from the 2004 source model could reproduce the SSHs (Smith *et al.*, 2005; Hirata *et al.*, 2006) (see Section 4.17.2).

4.17.4.1.5 Definition of amplitude and wave heights

The tsunami 'amplitude' is measured from zero to peak, either positive or negative. The 'wave height', also called as range or double amplitude, is measured from trough to peak. Hence, the 'tsunami height' is measured from trough to peak on tide gauges, whereas the 'tsunami amplitude' is measured from estimated tidal level at the time (**Figure 7**). As in seismic waves, 'initial motion', either positive or negative, carries information on the tsunami source. For example, the initial motions of the 2004 tsunami in Indian Ocean were negative to the east of the source (e.g., Thailand) but positive to the west (e.g., Maldives) as shown in **Figure 7**.

4.17.4.2 Field Survey to Measure Run-Up Heights

4.17.4.2.1 Post-tsunami surveys

Post-tsunami field surveys are often carried out after large tsunamis. Usually, experts from different backgrounds, such as coastal engineering, geology, or

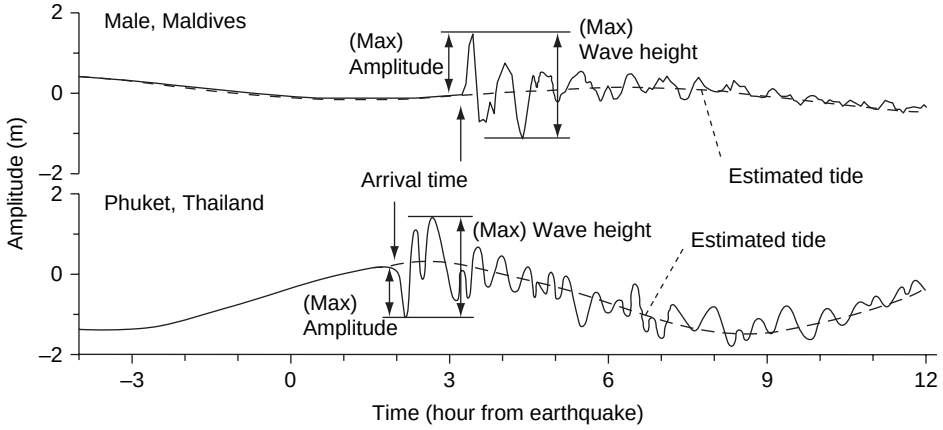


Figure 7 Tsunami waveforms from the 2004 Sumatra–Andaman earthquake recorded at Male (top) and Phuket (bottom).

seismology, and from different countries form international tsunami survey teams that visit affected coasts to document tsunami effects. An essential element of survey is to measure the tsunami heights along the coast (Intergovernmental Oceanographic Commission, 1998). The heights, defined below, are measured from various traces such as damage on buildings or other construction, watermarks left on walls, vegetative markers such as trim lines and saltwater discoloration or die-off, or lines of debris deposited on beaches. When no physical evidence is found, the tsunami heights are estimated based on eyewitness accounts, but considered as less reliable. Locations are usually identified by portable GPS measurements and the heights are measured by traditional leveling or laser surveying. Interviews with local people also include degree of ground shaking, tsunami arrival times, the number of waves with periods, and the duration. Post-tsunami surveys also include description and documentations of building damage and tsunami deposits. Synolakis and Okal (2005) review and summarize 15 post-tsunami surveys made by international tsunami survey teams during the 11-year period between 1992 and 2002.

4.17.4.2.2 Definitions of inundation and run-up heights

Tsunami height on land is called 'flow depth' when measured from the ground level, and 'inundation height' measured relative to sea level at the time of tsunami arrival. The height at maximum inundation is called 'run-up height'. Because the tide levels at the time of measurements and tsunami arrival are usually different, the corrections are necessary for survey data. Unless the sea surface of inundated tsunami is flat, run-up height is not necessarily the same as the inundation height or flow depth at the shore. The horizontal distance of water penetration measured from the coast is called 'inundation distance' (**Figure 8**).

4.17.4.3 Study of Historical Documents

Tsunamis and their effects documented in historical records are also used to study old earthquakes and tsunamis for which no instrumental data are available. Some countries in Europe, South America, or Asia have a long history of written records that provide accurate dates for past earthquakes and

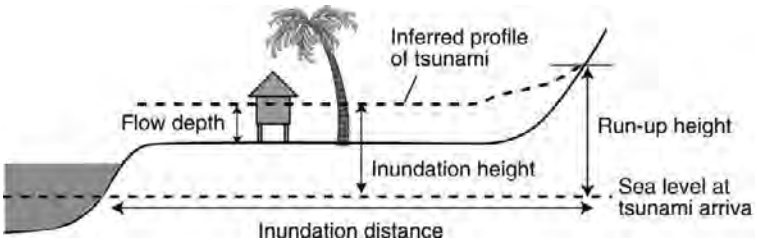


Figure 8 Definitions of tsunami heights on land.

tsunamis. For example, tsunami heights from the 1755 Lisbon earthquake have been estimated from historical documents as 6 m in Lisbon, more than 15 m at Southwest Portugal, and more than 10 m in Spain and Morocco (Baptista *et al.*, 1998).

Estimation of tsunami heights from damage description requires many assumptions. For example, the Japanese descriptions of the January 1700 tsunami, generated by a giant earthquake in Cascadia subduction zone, include flood in farmed fields and a government warehouse and damage to coastal houses (Atwater *et al.*, 2005). To convert these reported effects into estimated heights, Satake *et al.* (2003) considered various factors, such as the relationship between tsunami damage and heights, sea level change in the last 300 years, and similarity with the 1960 Chilean tsunami. Japanese tsunami data indicate that the house damage is partial when tsunami height is less than 2 m, but becomes significant when tsunami height becomes more than 4 m (Hatori, 1984).

4.17.4.4 Geological Methods

Even older tsunamis are studied from their deposits by geological means (Dawson and Shi, 2000). In Europe, inferred tsunami deposits from the Storegga slides in the North Sea have been found and studied in Scotland, Norway, and Iceland (Dawson *et al.*, 1988). In the US, tsunami deposits from prehistoric great earthquakes along the Cascadia subduction zone have been studied (Atwater *et al.*, 2005). In Australia, geological evidence of Late Pleistocene tsunamis has been studied (Bryant *et al.*, 1996). In Japan, tsunami deposits in various environments such as lagoons, beaches, or marshes have been studied for modern, historic, and prehistoric tsunamis (Minoura and Nakaya, 1991; Nanayama *et al.*, 2003).

On some coasts, particularly those remote from tsunami sources, it can be difficult to distinguish between tsunami and storm deposits. The universal sedimentological criteria for identifying tsunami deposits have not been well established (Nanayama *et al.*, 2000; Tuttle *et al.*, 2004). For this context, studies of deposits from modern tsunamis are important and careful descriptions have been done for the recent tsunamis such as the 1993 Hokkaido tsunami (Nishimura and Miyaji, 1995), the 1998 Papua New Guinea tsunami (Gelfenbaum and Jaffe 2003), or the 2004 Indian Ocean tsunami (see Section 4.17.2). These case studies indicate that the tsunami deposits are highly variable and depend on the coastal topography and geology.

4.17.4.5 Tsunami Magnitude Scales

4.17.4.5.1 Imamura-Iida scale, m

The traditional tsunami magnitude scale is the so-called Imamura-Iida scale, m. Although the original definition was descriptive, the value is approximately equal to

$$m = \log_2 h \qquad [1]$$

where h is the maximum run-up height in meters (Iida *et al.*, 1967). This scale is similar to the earthquake intensity scale, and is especially convenient for old tsunamis from which no instrumental records exist.

Hatori (1979) extended the Imamura-Iida's m scale to include far-field tsunami data as follows:

$$m = 3 + \log\left\{(h/0.5)(R/1000)^{1/2}\right\}/\log\sqrt{5} \qquad [2]$$

where h is in meters and R is the distance from source to coast in kilometers. This indicates that the tsunami height decays with distances as $1/\sqrt{R}$, which is theoretically predicted for nondispersive waves traveling a long distance (Comer, 1980).

4.17.4.5.2 Tsunami intensity

Soloviev (1970) pointed out that Imamura-Iida's m scale is more like an earthquake intensity scale rather than a magnitude. He also distinguished the maximum tsunami height h and the mean tsunami height $H_{av.}$ He then defined tsunami intensity I as

$$I = \log_2\left(\sqrt{2}H_{av}\right) \qquad [3]$$

Comparison of [1] and [3] suggests that the mean tsunami height is given as $1/\sqrt{2}$ times the maximum height. The maximum intensity I on the coast nearest to the source is used to quantify the tsunami source.

4.17.4.5.3 Tsunami magnitude M_t

Another magnitude scale M_t, called tsunami magnitude, was defined and assigned for many earthquakes by Abe (1979). The definition of M_t for a trans-Pacific tsunami is (Abe, 1979)

$$M_t = \log H + C + 9.1 \qquad [4]$$

and for a regional ($100\,\text{km} < \Delta < 3500\,\text{km}$) tsunami is (Abe, 1981)

$$M_t = \log H + \log\Delta + 5.8 \qquad [5]$$

where H is a maximum amplitude on tide gauges in meters, C is a distance factor depending on a combination of the source and the observation points, and Δ is the nautical distance in kilometers. The above formulas were calibrated with the moment magnitude scale, M_w, of earthquakes. Equation [5] indicates that for the same M_t, the tsunami amplitude H decays with $1/\Delta$, which is different from the assumption used for Imamura-Iida's m scale, and may be valid only for regional distances.

4.17.4.6 Tsunami Catalogs

The location, size, run-up heights, and damage from past tsunamis have been compiled by several researchers and published as catalogs. For the Pacific tsunamis, the Iida catalog (Iida *et al.*, 1967) contains numerous tsunami height data as far back as AD 173. Russian catalogs (Soloviev and Go, 1974a, 1974b; Soloviev *et al.*, 1992) include Pacific tsunami data (but except for Russia) with some tide gauge records, tabulated tsunami heights, and descriptions for each tsunami. Regional catalogs including tide gauge records were also published for Alaska (Lander, 1996), the west coast of the US (Lander *et al.*, 1993), Japan (Iida, 1984; Watanabe, 1998), and Indian Ocean (Berninghausen, 1966; Murty and Rafiq, 1991).

Digital tsunami databases have been compiled, termed Historical Tsunami Data Base at the Institute of Computational Mathematics and Mathematical Geophysics, Novosibirsk (Gusiakov, 2005; Institute of Computational Mathematics and Mathematical Geophysics, 2005) and at the National Geophysical Data Center, NOAA.

4.17.5 Tsunami Generation

In this section, we review tsunami generation by earthquakes, landslides, and volcanic eruptions.

4.17.5.1 Earthquake Tsunamis

4.17.5.1.1 Ocean bottom and water surface displacements

Analytical solutions for linear gravity waves generated from several types of sources have been obtained by various researchers. Kajiura (1963) generalized the solutions by introducing the tsunami Green's functions. He showed that the initial conditions given at the ocean bottom and at the water surface differ by a factor of $1/\cosh kd$, where k is wave number of displacement. For large wavelength, $\lambda >> 2\pi d$, this factor becomes 1 because $\cosh kd = \cosh(2\pi d/\lambda) \approx 1$. Therefore, if the wavelength of ocean-bottom deformation is much larger than the water depth, it can be assumed that the water surface displacement is the same as the bottom displacement.

Kajiura (1970) examined the energy exchange between the solid Earth and ocean water for tsunamis generated by bottom deformation with a finite rise time. He showed that the efficiency of tsunami generation, defined as a ratio of dynamic tsunami energy to static energy, becomes nearly 1, if the duration is short compared to the time required for a tsunami to travel through the source area. For an earthquake with $M \sim 8$ at the water depth of 5000 m, the source size is about 100 km and the tsunami travel time over that distance is about 8 min. If the duration is much shorter, for example, 1 min or less, the efficiency is more than 0.9. This means that the bottom deformation can be considered as instantaneous. If the duration is too short, shorter than $2\sqrt{d/g}$, a part of the energy flux goes into acoustic wave. For the above case, this limit is 45 s.

4.17.5.1.2 Fault parameters and ocean-bottom deformation

The ocean-bottom deformation due to faulting can be calculated using the elastic theory of dislocation. The displacement in an infinite homogeneous medium due to a dislocation across a fault is given by Volterra's theorem (Steketee, 1958) and the explicit formulas for half-space are given (e.g., Mansinha and Smylie, 1971; Okada, 1985). Usually, only the vertical component of the ocean bottom is considered for tsunami generation. When the tsunami source is on a steep ocean slope and the horizontal displacement is large, the vertical displacement of water due to the horizontal displacement of the slope must also be considered (Tanioka and Satake, 1996a).

The fault parameters needed to compute surface deformation are the fault location, geometry (strike, dip, and rake), size (length L and width W), and average slip $\bar{u}$ (or the strike-slip and dip-slip components). The seismic moment M_0 is given as

$$M_0 = \mu \bar{u} S = \mu \bar{u} L W \qquad [6]$$

where S is the fault area and μ is rigidity. The moment magnitude is defined as (Hanks and Kanamori, 1979)

$$M_w = \frac{2}{3} \log M_0 - 10.7 \qquad [7]$$

where M_0 is given in dyne cm (10^{-7} N m). The effects of fault parameters on tsunami generation were examined by Yamashita and Sato (1974) and more recently by Geist (1998).

4.17.5.1.3 Estimation of earthquake source parameters from tsunamis

Tsunami data can be used to study earthquake source processes in a way similar to that in which seismic waves are used. Abe (1973) analyzed the tsunami arrival times and first-motion data from the 1968 Tokachi-oki earthquake. By drawing inverse refraction diagrams from each station (see Section 4.17.6.2), he estimated the tsunami source area, which agrees well with the aftershock area. From the first motion of tsunami waves, the initial water surface disturbance was estimated as uplift at the southeastern edge and subsidence at the northwestern edge. This pattern is very similar to the vertical bottom deformation due to the faulting, which was independently estimated from seismological analysis.

Numerical simulation of tsunami has been carried out for many fault models around Japan (e.g., Aida, 1978). The best fault models were obtained by trial and error comparisons of observed and computed tsunami waveforms or coastal tsunami heights. Tsunami waveforms recorded on tide gauges were then inverted to estimate the slip distribution on the fault (Satake, 1989). In this method, the fault plane is first divided into several subfaults, and the deformation on the ocean bottom is computed for each subfault with a unit amount of slip. Using this as an initial condition, tsunami waveforms are numerically computed on actual bathymetry. The observed tsunami waveforms are expressed as a superposition of waveforms computed for each subfault as follows:

$$A_{ij}(t) \cdot x_j = b_i(t) \qquad [8]$$

where $A_{ij}(t)$ is the computed waveform as a function of time t, or Green function, at the ith station from the jth subfault; x_j is the amount of slip on the jth subfault; and $b_i(t)$ is the observed tsunami waveform at the ith station. The slip x_j on each subfault can be estimated by a least-squares inversion of the above set of equations. The tsunami waveform inversion has been applied to many large and great earthquakes in the Alaska-Aleutians (Johnson, 1999), northern Japan (Satake, 1989; Tanioka *et al.*, 1995), the Nankai trough (Satake, 1993; Tanioka and Satake, 2001a, 2001b), and the 2004 Sumatra–Andaman earthquake (Fujii and Satake, 2007). These results show that the slip distributions are not uniform but are concentrated in small regions, which correspond to asperities (**Figure 3**).

4.17.5.1.4 Tsunamigenic and tsunami earthquakes

We call an earthquake that generates tsunami as 'tsunamigenic earthquake'. A 'tsunami earthquake', on the contrary, is defined as an earthquake which excites much larger tsunamis than expected from its seismic waves (Kanamori, 1972). The discrepancy can be quantified in terms of the surface wave magnitude (M_s) and tsunami magnitude (M_t) scales; when M_t is larger than M_s by more than 0.5 units, it is called a tsunami earthquake (Abe, 1989). Newman and Okal (1998) introduced a new parameter, called slowness parameter $\Theta = \log_{10}(E^E/M_0)$, a ratio of radiated seismic energy E^E to seismic moment M_0. They show that Θ is about −5 for regular earthquakes but smaller by 1–1.4 for tsunami earthquakes, and they proposed to use this parameter for discrimination of tsunami earthquakes.

Several proposals have been made for the generating mechanisms of tsunami earthquakes. Slow and long rupture processes (Kanamori, 1972) reconcile the large discrepancy between seismic and tsunami waves. High-angle secondary faulting in the accretionary wedge can explain the relatively large tsunami excitation (Fukao, 1979). For a given slip, the seismic moment is smaller in the shallower part of subduction zones because of the smaller rigidity, and the tsunami becomes larger (Okal, 1988). Tsunami heights may not be anomalously large if compared to M_w rather than M_s (Pelayo and Wiens, 1992); the large discrepancy between the tsunami and M_s is due to a saturation of M_s at around 7.3 for those earthquakes that occurred in the accretionary wedge.

Recent tsunami waveform modeling shows that the fault parameters and locations of tsunami earthquakes share a common feature: narrow and shallow faulting near the trench axis (**Figure 9**). The 1992 Nicaragua earthquake ($M_s = 7.2$, $M_t = 7.9$) was the first tsunami earthquake recorded on modern broadband seismic instruments, and seismological studies showed that the duration was very long for its size, about 100 s (Kanamori and Kikuchi, 1993). Comparison of numerically computed tsunami waveforms with tide gauge records (Satake, 1994) showed that a narrow (~40 km) and shallow (extending only to the upper 10 km of the ocean bottom) fault is responsible for the tsunami generation.

Among the tsunami earthquakes, the 1896 Sanriku earthquake ($M_s = 7.2$, $M_t = 8.6$) and the 1946 Aleutian

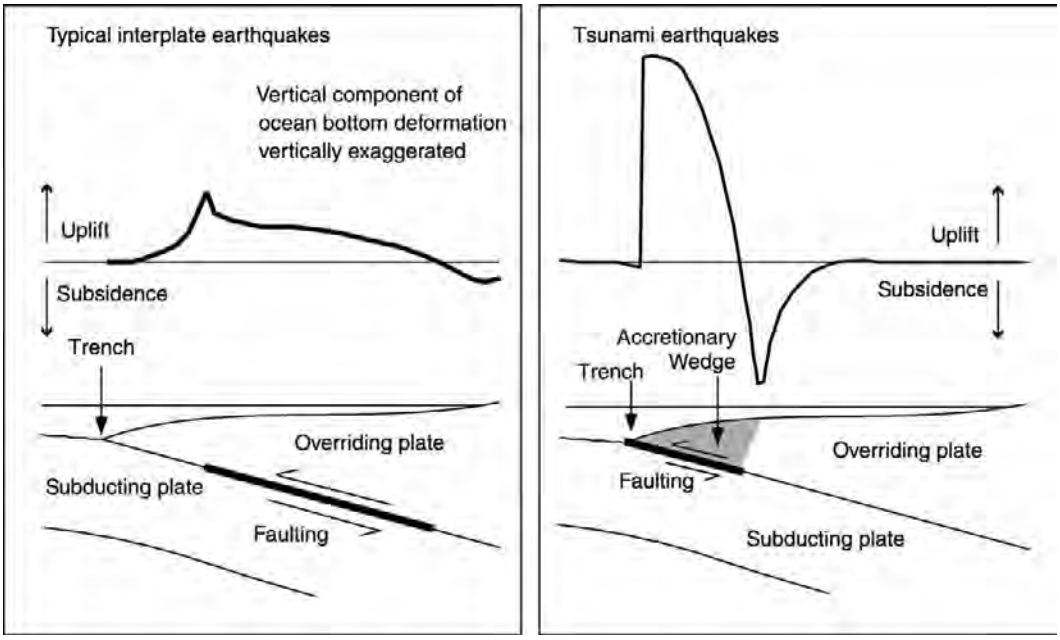


Figure 9 Vertical deformation of seafloor (top) due to faulting (bottom) for typical interplate earthquake (left) and tsunami earthquakes (right). Modified from Satake K and Tanioka Y (1999) Sources of tsunami and tsunamigenic earthquakes in subduction zones. *Pure and Applied Geophysics* 154: 467–483.

earthquake ($M_s = 7.3$, $M_t = 9.3$) were very anomalous (Kanamori, 1972; Kanamori and Kikuchi, 1993). While the tsunami modeling showed a similar character in fault parameters to the Nicaragua earthquake, that is, narrow fault near trench axis (Tanioka and Satake, 1996b; Johnson and Satake, 1997; Tanioka and Seno, 2001a, 2001b), Kanamori and Kikuchi (1993) suggested that submarine slumping was involved in the unusually large tsunami generation. These tsunami earthquakes occurred in trenches with large amounts of sediments and an accretionary prism where slumping occasional occur. Other tsunami earthquakes such as the 1992 Nicaragua earthquake occurred in subduction zones with little sediment, where a slow rupture is responsible for large tsunamis (Kanamori and Kikuchi, 1993).

Tanioka *et al.* (1997) showed that the 1896 tsunami earthquake occurred in a region where the ocean-bottom topography is rough, characterized by a well-developed horst and graben structure. Polet and Kanamori (2000) extended this model to global subduction zones, based on the examination of source spectra of large ($M > 7$) earthquakes in the 1990s. Okal and Newman (2001), however, failed to find any correlation between the slowness parameter Θ and either depth, focal mechanism, or seismic moment for central America, Java, and Peru, suggesting an absence of regional signal for the occurrence of tsunami earthquakes. Bilek and Lay (2002), from analysis of source time functions of tsunamigenic and tsunami earthquakes, found that shallower earthquakes have longer duration and claimed that tsunami earthquakes are simply larger earthquakes at very shallow depth. Seno (2002) proposed that increase in fluid pressure causes temporal change in frictional property near trench axis and responsible for slow rupture; hence, the tsunami earthquakes are transient phenomena.

Seno and Hirata (2007) showed that the 2004 earthquake had some component of 'tsunami earthquake', slow duration, and updip rupture of seismogenic zone. They inferred that the seismogenic rupture propagated toward north at the rupture velocity of 2.5 km s^{-1}, but was followed by slower rupture ($\sim$0.7 km s^{-1}) at shallower part, which has about a third of the total seismic moment.

4.17.5.2 Tsunamis from Landslides

Tsunamis from submarine landslides have been a hot research topic since the 1990s. Abnormally large and concentrated tsunami heights from the 1992 Flores Island, Indonesia, and 1998 Sissano, Papua New Guinea, earthquakes and the nonseismic 1994

Skagway tsunami, Alaska, provided case studies. Marine surveys revealed detailed bathymetric features of landslides or slumps. Physical experiments in tanks and numerical computations of landslide-generated tsunamis have been developed. Such multidisciplinary studies have been summarized in several special issues (Keating *et al.*, 2000; Bardet *et al.*, 2003; Tappin, 2004).

4.17.5.2.1 Examples

Post-tsunami field surveys after large tsunamigenic earthquakes sometimes show abnormally large tsunami heights concentrated in a small portion of coasts. While offshore bathymetry or coastal topography (such as V-shaped bays) can cause such focusing of incoming tsunamis, locally large tsunami may be caused by a local landslide, either subaerial or submarine, triggered by ground shaking. Large water waves caused by a rockslide in Lituya Bay at the time of 1958 earthquake are described in Section 4.17.3.1.4. Concurrent with the 1964 Alaska earthquake (M_w 9.2), in addition to Pacific-wide tsunamis, local tsunamis from landslides, were reported at more than 20 locations, with the largest run-up height of 50 m (Plafker *et al.*, 1969; Lander, 1996). More recently, the tsunami heights from the 1992 Flore Island earthquake were mostly 4–5 m on most of the northern coast of the island, but 26 m at Riangkroko (Yeh *et al.*, 1993). The tsunami heights from the 1998 Papua New Guinea tsunami, also described in Section 4.17.3.1.1, show local peak of 15 m (**Figure 10**).

Larger-scale submarine slumping and turbidity currents were triggered by the 1929 Grand Banks earthquake (Heezen and Ewing, 1952; Fine *et al.*, 2005a). The maximum tsunami run-up height of 13 m and 28 casualties made it the worst tsunami disaster in Canada. The turbidity current also caused damage on the trans-Atlantic submarine cables at 12 locations over a range of 600 km.

An example of nonseismic landslide tsunami is the 1994 Skagway tsunami (Kulikov *et al.*, 1996). A submarine landslide, without seismic trigger, occurred in Skagway Harbor in Alaska and generated local tsunami of 5–10 m high. The tide gauge records show a dispersive character, because of small size of the source (see the next section).

4.17.5.2.2 Characteristics and discrimination of landslide tsunamis

Scale length of submarine mass failure is very different from that of submarine faulting. While vertical movement can be hundreds of meters, the horizontal scale is mostly less than 10 km for submarine mass failure. Seafloor deformation due to faulting, on the contrary, has at most several meters of vertical movement but the spatial extent can be several hundreds of kilometers. As will be discussed in next section, an earthquake source almost always satisfies the long-wave (shallow water) condition, while the submarine landslide may not. Hence, the water waves from submarine landslides show a dispersive character and attenuate quickly without reaching a long distance.

Okal and Synolakis (2004) fit the run-up height distribution $\zeta(y)$ along a beach (in the y-direction) by the following function:

$$\zeta(y) = \frac{b}{[(y-c)/a]^2 + 1} \qquad [9]$$

where parameters a, b, and c are estimated from the observed data (**Figure 10**). They show that $I_2 = b/a$, the aspect ratio of maximum run-up heights to extent of run-up along beach, can be used as a discriminant of the nature of tsunami source, either dislocation (earthquake fault) or landslide source. Seismic dislocation sources have I_2 smaller than 10^{-4}, whereas a landslide source has a larger value.

4.17.5.2.3 Modeling landslide tsunamis

Submarine landslide sources of tsunamis have been also modeled in many ways. These can be grouped into three categories: dynamic fluid flow, kinematic rigid-body motion, and static initial profile.

In the dynamic fluid flow approach, landslide is treated as fluid and its evolution is calculated simultaneously with tsunami generation. Jiang and Leblond (1994) treated a landslide as incompressible viscous flow and seawater as inviscid flow, and formulated the multiphase fluid dynamics. Fine *et al.* (2005a) applied the multiphase flow method to the 1929 Grand Banks landslide tsunami; Imamura *et al.* (2001), Heinrich *et al.* (2001), and Titov and Gonzalez (2001) applied a similar method to the submarine landslide of the 1998 Papua New Guinea tsunami.

In the rigid-body approach, temporal change of water height due to progress of submarine landslide is prescribed and used to compute tsunami propagation. Harbitz (1992) assumed a box-shaped slide for the Storegga slide, Ward (2001) used a chain of rigid boxed to model the Storegga and other submarine slides. Tinti *et al.* (1997) introduced a block model of landslide and simulated the landslide motion, then

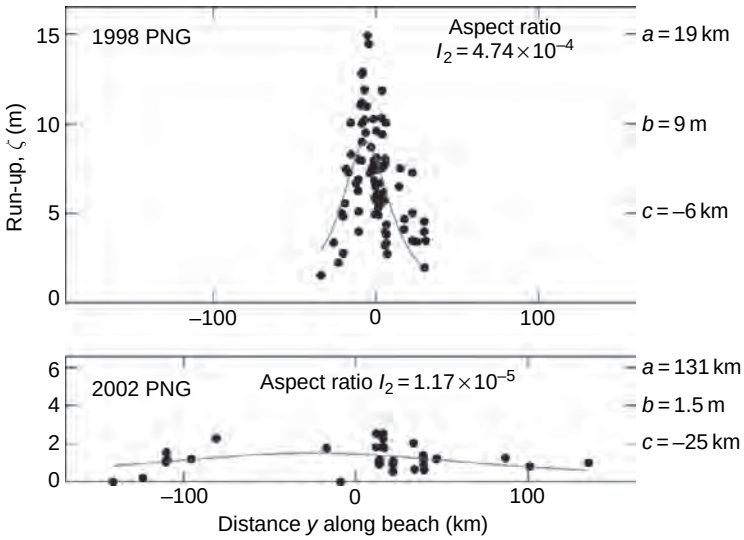


Figure 10 Comparison of tsunami run-up heights for the 1998 and 2002 Papua New Guinea tsunamis. Estimated parameters of [9] are also shown. Modified from Okal EA and Synolakis CE (2004) Source discriminants for near-field tsunamis. *Geophysical Journal International* 158: 899–912.

used the results for the tsunami computations. Satake *et al.* (2002) introduced a simple kinematic landslide model on the basis of bathymetry change.

In the static approach, an initial profile due to landslide is assumed. Watts *et al.* (2003), based on the results of physical and numerical experiments, introduced a parametric initial profile for submarine landslide, and applied such initial profiles for several landslide-generated tsunamis including the 1946 Aleutian, 1994 Skagway, and 1998 Papua New Guinea tsunamis. Synolakis *et al.* (2002) and Satake and Tanioka (2003) used the same approach to model the 1998 Papua New Guinea tsunami.

For water waves generated from subaerial landslides, various studies such as theoretical, experimental, numerical and field approaches are summarized in Slingerland and Voight (1979).

4.17.5.3 Volcanic Tsunamis

Tsunamis are one of the most important hazards associated with volcanic activity. About a quarter of human casualties directly associated with volcanic eruption is due to tsunamis (Tilling, 1989). Volcanic tsunamis can be roughly categorized into five groups according to their origin: (1) earthquakes; (2) landslides and avalanches of cold rock and lahars; (3) flow of volcanic ejecta such as pyroclastic flow, base surge, hot rock avalanches, or lava; (4) submarine explosion; and (5) caldera collapse or subsidence. From case studies of 106 volcanic tsunamis, Latter (1981) classified 69 of them into the 10 groups. In addition, he also classified tsunamis generated from air waves of explosions (4.5%) with possible examples in Bezymianny, Kamchatka in 1955–1956, as well as some far-field tsunamis from the 1883 Krakatoa eruption.

4.17.5.3.1 Volcanic tsunamis due to earthquakes

According to Latter (1981), some 22% of volcanic tsunamis are caused by earthquakes accompanying eruptions. Examples include the 1914 Sakurajima, Japan, 1930 Stromboli, Italy, and 1933 Severgin, Kuril Islands, tsunamis.

4.17.5.3.2 Volcanic tsunamis due to landslides

Latter (1981) separated the tsunamis generated from landslides and avalanches of cold rock (7%) and lahars (5%). A typical example of landslide is from Shimabara, Japan. In 1792, Mayuyama, a small mountain near Unzen volcano, collapsed and the landslide entered Ariake-kai, a shallow bay with typical depth of 10–20 m. Because the horizontal extent of the slide is about 3000 m, much longer than the water depth, this satisfied the long-wave approximation, hence causing large (10–20 m) tsunami height and

significant damage (~5000 casualties) on the other side of the bay. The total casualties was ~15 000.

4.17.5.3.3 Volcanic tsunamis from flow of volcanic ejecta

About a third of volcanic tsunamis can be classified in this group. Latter (1981) further separated them into pyroclastic flows (ignimbrites and nuees ardentes; 20%), base surges (7%), avalanches of hot rock (8%), or lava (1%).

Latter (1981) considered the 1883 Krakatoa tsunami as due to pyroclastic flow, although this is controversial (see Section 4.17.5.3.5). Other examples of pyroclastic flow include St. Vincent and Mt. Pelee in lesser Antilles in 1902 and St. Augustine (Alaska) in 1883. Other types of ejecta are base surge in Taal volcano in the Philippines, hot rock avalanche at Stromboli, Italy, and lava avalanche at Matavanu volcano, Samoa in 1906–1907.

4.17.5.3.4 Volcanic tsunamis from submarine explosion

Submarine explosion occupies about 19% of volcanic tsunamis according to Latter (1981). Submarine volcano, Bayonnaize Rocks (along Izu-Bonin Islands about 300 km south of Honshu), erupted in 1952. Tsunamis from this eruption, recorded at Hachijo Island about 130 km north, show very short period (~1 min) wave with a dispersive character. Unoki and Nakano (1953) applied a Cauchy–Poisson wave (deep water waves from initial water surface disturbance) to the observed tsunami record and estimate the source size (~2 km) and water surface uplift or impulse force.

4.17.5.3.5 Volcanic tsunamis from caldera collapse

About 9% of volcanic tsunamis were from caldera collapse or subsidence. Large-scale caldera collapse was recorded at the times of the 1883 Krakatoa and the 1883 Ritter Island (Papua New Guinea) eruptions. Ward and Day (2003) modeled the Ritter Island tsunami by using the landslide model.

The generation mechanism of the 1883 Krakatoa tsunami had been long debated. A numerical computation of this tsunami and comparison with tide gauge record and run-up heights indicated that the tsunami was generated by submarine eruption, rather than caldera collapse, pyroclastic flow, or falling ejecta (Nomanbhoy and Satake, 1995).

4.17.6 Tsunami Propagation

4.17.6.1 Weak Coupling between Elastic Earth and Ocean

In the Earth-ocean-coupled system, an earthquake source would generate waves both in the elastic Earth and in the ocean, and the propagation of a tsunami in the ocean layer would excite waves in the elastic Earth. In contrast, the traditional decoupled approach first calculates the elastic deformation on the ocean bottom generated by an earthquake, then uses this as an initial condition for tsunami wave propagation such that the tsunami wave in the ocean would have no influence on the elastic Earth. Theoretical treatment of tsunamis viewed as coupled system is similar to that of Rayleigh waves or spheroidal oscillations. Ward (1980) and Okal (1988) extended the free oscillation approach to solve the generation and propagation of tsunamis, while Comer (1984a) extended the Rayleigh wave approach to solve the tsunami problem. The calculated dispersion curve shows that, except at very low frequencies, the phase velocity is very similar to that of the gravity wave. Both approaches give almost identical solutions, because of the small compressibility of water (Comer 1984b). In other words, the coupling between the elastic Earth and the ocean is extremely weak, which explains why very large earthquakes are necessary to generate tsunamis.

4.17.6.2 Ray-Theoretical Approach

If the tsunami wavelength is much smaller than the scale of heterogeneity in propagation velocity, or the change in water depth, then we can apply the geometrical ray theory of optics. The wavefronts of propagating tsunami can be drawn on the basis of Huygens' principle. Such a diagram is called a 'refraction diagram'. Refraction diagrams can be prepared for major tsunami sources and used for tsunami warning; as soon as the epicenter is known, the tsunami arrival times can be readily estimated. **Figure 1** shows the refraction diagram from the 2004 Andaman–Sumatra earthquake with wavefronts at each hour. It is shown that the tsunami was expected to arrive at Thai and Indian coasts in 2 h and at Africa in 7 hs.

The refraction diagrams can also be drawn backwards from coasts. Such a diagram is called an 'inverse refraction diagram' and used to estimate the tsunami source area. In this case, the wavefronts or

rays are traced backwards for the corresponding travel times (Miyabe, 1934). The traced wavefronts from each observed station bound the tsunami source area.

4.17.6.3 Coastal Amplification and Green's Law

From the conservation of energy along the rays, it can be shown that

$$b_0 d_0^{1/2} h_0^2 = b_1 d_1^{1/2} h_1^2 \qquad [10]$$

where d is the water depth, b is the distance between the rays, h is the tsunami amplitude, and the subscripts 0 and 1 indicate two different locations. If the tsunami amplitude at location 0 (e.g., deep water) is known, the tsunami amplitude after propagation can be estimated as

$$h_1 = \left(\frac{b_0}{b_1}\right)^{1/2} \left(\frac{d_0}{d_1}\right)^{1/4} h_0 \qquad [11]$$

This is known as Green's law. The ratio b_1/b_0 represents the spreading of rays, which can be graphically obtained from refraction diagrams. Note that this applies only to direct waves and does not account for of the reflected waves or edge waves.

4.17.6.4 Tsunami as a Gravity Wave

In this section, we take a two-dimensional Cartesian coordinate system with z-axis vertical upward (the origin is on the undisturbed water level) and consider propagation of a wave (**Figure 11**). When gravity is the restoring force, it is called a gravity wave. Euler's equation of motion can be written as

$$\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V} = -\mathbf{g} - \frac{1}{\rho} \nabla p \qquad [12]$$

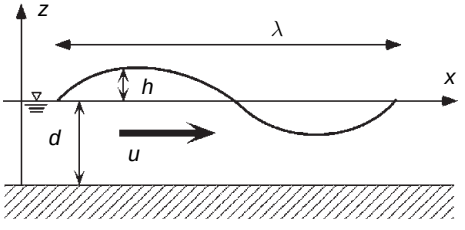


Figure 11 Definition of coordinate system and the parameters for gravity waves.

where $\mathbf{V}$ is the velocity vector (whose x- and z-components are u and w, respectively), $\mathbf{g}$ is the gravitational acceleration, ρ and p are water density and pressure, respectively.

The phase velocity c of gravity waves is given as

$$c = \left(\frac{g}{k} \tanh kd\right)^{1/2} = \left(\frac{g\lambda}{2\pi} \tanh \frac{2\pi d}{\lambda}\right)^{1/2} \qquad [13]$$

where k is the wave number, λ is the wave length, and d is the water depth.

When the horizontal scale of motion, or the wavelength λ, is much larger than the water depth ($\lambda >> d$), the vertical acceleration of water is negligible compared to the gravity. This means that the horizontal motion of water mass is almost uniform from bottom to surface. Such a wave is called a shallow-water wave or long wave. For tsunamis, typical ocean depth is about 4–5 km and a large earthquake has a source area of several tens to hundreds of kilometers. Hence, the long-wave approximation is appropriate for most cases.

When the amplitude is small compared to the water depth ($h << d$), the equation of motion, or conservation of momentum, can be written as

$$\frac{\partial u}{\partial t} = -g \frac{\partial h}{\partial x} \qquad [14]$$

and the equation of continuity, or conservation of mass, as

$$\frac{\partial h}{\partial t} = -\frac{\partial}{\partial x}(\mathrm{d}u) \qquad [15]$$

Such a wave is called a small-amplitude, or linear long wave. The assumptions are valid for most tsunami propagation paths except for near-shore regions.

From [14] and [15], by assuming that water depth d is constant, we obtain the wave equation

$$\frac{\partial^2 h}{\partial t^2} = c^2 \frac{\partial^2 h}{\partial x^2}, \quad c = \sqrt{gd} \qquad [16]$$

in which the phase velocity is determined by the water depth only. Note that the above phase velocity can be also obtained from [13] if $\lambda >> d$.

4.17.6.5 Basic Equations for Tsunami Propagation

Including the bottom friction and the Coriolis force, the momentum equation for shallow-water (long)

waves can be written for a three-dimensional case as follows:

$$\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial x}+V\frac{\partial U}{\partial y}=-f\,V-g\frac{\partial h}{\partial x}-C_{\mathrm{f}}\frac{U\sqrt{U^2+V^2}}{d+h} \quad [17]$$

$$\frac{\partial V}{\partial t}+U\frac{\partial V}{\partial x}+V\frac{\partial V}{\partial y}=f\,U-g\frac{\partial h}{\partial y}-C_{\mathrm{f}}\frac{V\sqrt{U^2+V^2}}{d+h}$$

and the mass equation is

$$\frac{\partial h}{\partial t}+\frac{\partial}{\partial x}\{U(h+d)\}+\frac{\partial}{\partial y}\{V(h+d)\}=0 \quad [18]$$

where the coordinate system is $x=$ east and $y=$ south, f is the Coriolis parameter, C_{f} is a nondimensional bottom frictional coefficient, and U and V are the depth-averaged velocities in x- and y-directions, respectively. The velocities are averaged from the bottom to the surface:

$$U=\frac{1}{h+d}\int_{-d}^{h}u\,\mathrm{d}z=\frac{Q_x}{h+d}$$
$$V=\frac{1}{h+d}\int_{-d}^{h}v\,\mathrm{d}z=\frac{Q_y}{h+d} \quad [19]$$

where Q_x and Q_y are flow rates in x- and y-directions. A typical value of a non-dimensional frictional coefficient C_{f} is 10^{-3} for coastal water and 10^{-2} for run-up on land (Satake, 1995).

4.17.6.6 Dispersion

Here we will briefly examine general gravity waves deviating from the small-amplitude, linear shallow-water (long) wave approximations (Peregrine, 1972; Mei, 1989). We have already seen that the phase velocity of a general gravity wave is given by [13]. If we expand the hyperbolic tangent function into its Taylor series and include the second term, [13] becomes

$$c=\sqrt{gd}\left\{1-\frac{1}{3}(kd)^2\right\}^{1/2}\approx\sqrt{gd}\left\{1-\frac{1}{6}(kd)^2\right\} \quad [20]$$

The corresponding equation of motion is

$$\frac{\partial u}{\partial t}=-g\frac{\partial h}{\partial x}+\frac{1}{3}d^2\frac{\partial^3 u}{\partial x^2\partial t} \quad [21]$$

and the equation of continuity is

$$\frac{\partial h}{\partial t}=-d\frac{\partial u}{\partial x} \quad [22]$$

These equations, including frequency dispersion, are known as the linear Boussinesq equation.

If we relax the small-amplitude ($h << d$) assumption, the nonlinear, nondispersive equations of motion and continuity are

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=-g\frac{\partial h}{\partial x} \quad [23]$$

$$\frac{\partial h}{\partial t}=-\frac{\partial}{\partial x}\{u(h+d)\} \quad [24]$$

These equations are for the finite-amplitude shallow-water waves and account for amplitude dispersion. For the linear case (ignoring the advection term), the phase velocity is given by $c=\sqrt{g(d+h)}$. This means that the larger the amplitude, the faster the wave speed. As a consequence, peaks of a wave catch up with troughs in front of them, and the forward facing portion of the wave continues to steepen and it will eventually break.

4.17.6.7 Numerical Computations

The tsunami propagation depends on the water depth, which is not uniform but the variation is relatively well known compared to velocity distribution of seismic waves. Hence, numerical computations on realistic or actual bathymetry have been popularly made. The historical review of tsunami numerical simulation is presented by Shuto (1991).

Let us consider a simple one-dimensional propagation (on constant depth) of linear long waves. Denoting the average velocity (x-component only) as U, the equations of motion and continuity are given as follows:

$$\frac{\partial U}{\partial t}=-g\frac{\partial h}{\partial x} \quad [25]$$

$$\frac{\partial h}{\partial t}=-d\frac{\partial U}{\partial x} \quad [26]$$

Using the staggered (leap-frog) grid system for U and h, and denoting time $t=l\,\Delta t$ as superscripts and space $x=m\,\Delta x$ as subscripts where l and m are integers, these become

$$\frac{\left[U_m^{\,l+(1/2)}-U_m^{\,l-(1/2)}\right]}{\Delta t}=-g\frac{\left[h^l_{m+(1/2)}-h^l_{m-(1/2)}\right]}{\Delta x} \quad [27]$$

$$\frac{\left[h^{l+1}_{m+(1/2)}-h^l_{m+(1/2)}\right]}{\Delta t}=-d\frac{\left[U_{m+1}^{\,l+(1/2)}-U_m^{\,l+(1/2)}\right]}{\Delta x} \quad [28]$$

It can be shown that the error will grow with time unless the following stability condition is met:

$$\sqrt{gd}\left(\frac{\Delta t}{\Delta x}\right) \le 1 \text{ or } \Delta t \le \frac{\Delta x}{\sqrt{gd}} \quad [29]$$

which is known as Courant–Friedrichs–Lewy (CFL) condition. The physical meaning of the stability condition is that the time step Δt must be equal to or smaller than the time required for the disturbance to travel the spatial grid size Δx. If Δt is larger, the values other than $(l, m-1)$, $(l-1, m)$ and $(l, m+1)$ will influence the value at $(l+1, m)$.

For the two-dimensional case, the CFL condition becomes

$$\Delta t \le \frac{\Delta x}{\sqrt{2gd}} \quad [30]$$

There are two types of boundaries involved in the tsunami computation. The first one is land–ocean boundary. The simplest assumption is a total reflection of energy on the coast. For the computation of tsunami inundation on land, however, a moving land–ocean boundary is needed; the boundary moves depending on the relative heights of water height and land. The other kind of boundary is the open ocean boundary along the outside of the computational region. A typical assumption is the radiation condition, which means that waves go out of the computational region keeping their slopes.

The computer program for one-dimensional finite-difference computation is very simple. Since there is no overlap between the two (height and velocity) grids (**Figure 12**) in the linear case, in the program the spatial grid points m and $m + (1/2)$ can be treated by the same index (e.g., `i`) and the time $l - 1/2$ and l can be treated as the same time (e.g., `it`). Writing U and h using the two-dimensional array as U(`i, it`) and h(`i, it`), [27] and [28] become

$$U(\mathtt{i},\mathtt{it}+1) = U(\mathtt{i},\mathtt{it}) - g\frac{\Delta t}{\Delta x}[h(\mathtt{i},\mathtt{it}) - h(\mathtt{i}-1,\mathtt{it})] \quad [31]$$

$$h(\mathtt{i},\mathtt{it}+1) = h(\mathtt{i},\mathtt{it}) - d\frac{\Delta t}{\Delta x}[U(\mathtt{i}+1,\mathtt{it}+1) - U(\mathtt{i},\mathtt{it}+1)] \quad [32]$$

The algorithm thus is as follows:

set up parameters
read bathymetry data
read initial condition h(`i, 1`)
while (`it`<`it`$_{end}$)

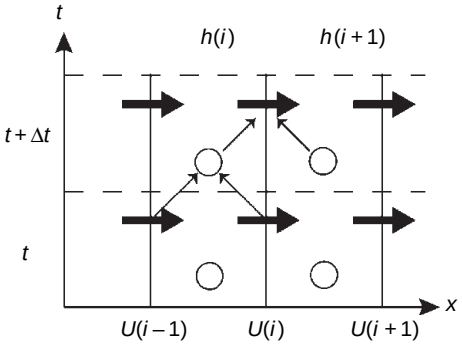


Figure 12 Staggered leap-frog system.

compute U(`i, it+1`) from U(`i, it`) and h(`i, it`)
set U(`i, it+1`) = 0 for land boundary
compute h(`i, it+1`) from U(`i, it+1`) and h(`i, it`)
adjust h(`i, it+1`) for open boundary
exchange variables
h(`i, it+1`)→h(`i, it`)
U(`i, it+1`)→U(`i, it`)
`it`+=1
end

Figure 13 shows the snapshots of tsunami propagation from the 2004 Sumatra–Andaman earthquake.

4.17.7 Tsunami Warning and Hazard Reduction Systems

The December 2004 tsunami in the Indian Ocean caused the worst tsunami disaster in the history. In addition to the gigantic size of the earthquake, lack of preparation for tsunamis on the coast of Indian Ocean was also responsible for the tragedy. Preparation for tsunami includes a tsunami warning system based on seismic and sea level monitoring, hazard assessment such as making hazard maps, and education and awareness to the coastal residents. Such systems on tsunami hazard reduction rely on the scientific knowledge of earthquakes and tsunamis.

The US National Tsunami Hazard Mitigation Program (Bernard, 2005) is a comprehensive program based on three components: warning guidance, hazard assessment, and mitigation (**Figure 14**). Tsunami warning systems can be grouped into a basin-wide (such as Pacific or Indian Ocean), regional, and local systems. Hazard assessment includes making inundation maps or probabilistic maps. Mitigation includes

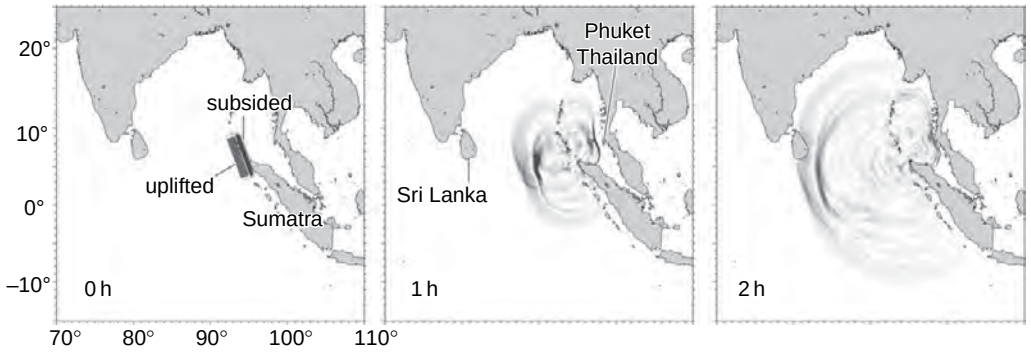


Figure 13 Snapshots of simulation for the 2004 tsunami in Indian Ocean.

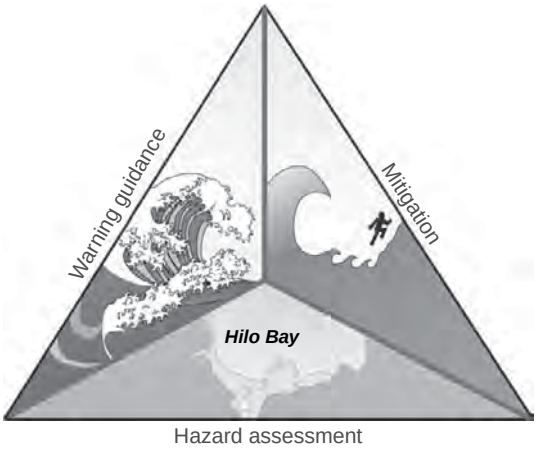


Figure 14 Three key elements for tsunami hazard reduction: warning guidance, hazard assessment, and mitigation (Bernard, 2005).

public awareness such as implementing tsunami signs or educating coastal residents.

4.17.7.1 Far-Field Tsunami Warning Systems

Tsunamis travel very fast as an ocean wave, at about 710 km h^{-1}, or 0.2 km s^{-1}, for a water depth of 4000 m, yet much slower than seismic waves, which travel at 5–10 km s^{-1}. This velocity difference between the seismic and tsunami waves makes it possible to issue a tsunami warning by detecting seismic waves before the actual tsunami arrival. For distant tsunamis that propagate across the Pacific or Indian Ocean, there are a few hours for accurate and reliable warning. The Richard H. Hagemeyer Pacific Tsunami Warning Center (PTWC), located in Hawaii, is the operational center for the international tsunami warning system, and issues warnings for regional and distant tsunamis in the Pacific Basin to almost all the countries and states around the Pacific Ocean (McCreery, 2005). The West Coast/Alaska Tsunami Warning Center, located in Alaska, is responsible for tsunami warning along the US and Canadian coasts except for Hawaii. At these centers, seismic waveform data from several tens of stations are continuously monitored, and once a large earthquake is detected, the location, depth, and magnitudes are assessed using a combination of automatic and interactive procedures by watchstanders (**Figure 15**). Recent deployment of advanced seismological parameters such as M_m (Okal and Talandier, 1989), M_{wp}(Tsuboi *et al.*, 1995), or Θ (Newman and Okal, 1998) makes it possible to quickly assess the earthquake size and tsunami potential, and to issue the tsunami warnings in less than 25 min. To avoid false alarm, it is very important to confirm that a tsunami was actually generated; hence, sea levels are monitored with both costal tide gauges and tsunameters (DART buoys). If a tsunami is observed, then the tsunami warning is updated (Titov *et al.*, 2005b). If the tsunami cannot be confirmed, then the warning is canceled.

At the time of 26 December 2004 tsunami, PTWC issued the first information bulletin at 1:14 GMT, only 15 min after the earthquake. An earthquake was located off the west coast of Northern Sumatra, but the magnitude (M_{wp}) was estimated as 8.0. The second bulletin was issued at 2:08 GMT, 69 min after the earthquake yet before the tsunami arrivals at Thai, Sri Lankan, or Indian coast. The earthquake size (M_m) was updated to 8.5 and a possibility of local tsunami was mentioned in the bulletin.

The French Polynesia Tsunami Warning Center developed and adopted TREMORS (Tsunami Risk Evaluation through seismic Moment in a Real time System) in 1987 (Reymond *et al.*, 1991). TREMORS

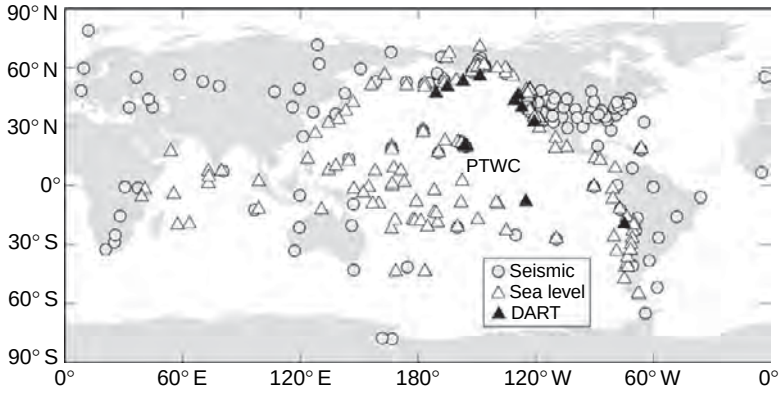


Figure 15 Distribution of seismic stations and sea level monitoring, coastal gauges, and tsunameter (DART) stations that send the data to PTWC in real time.

features automatic detection and location of seismic events using a single three-component long-period seismic station, and estimation of seismic moment through variable-period mantle magnitude M_m (Okal and Talandier, 1989). The advantage of using long-period seismic waves for tsunami warning is the ability to detect 'tsunami earthquakes'. The disadvantage is that longer time is needed before the seismic waves are recorded. TREMORS is a hybrid system; it updates the moment estimate as more seismic data are available.

The Pacific-wide tsunami warning system was established after the 1946 Aleutian tsunami that affected the Hawaiian Islands (Section 4.17.3.1.5). After the 1960 Chilean tsunami (Section 4.17.3.1.3), the Pacific tsunami warning system became international, and an International Coordination Group for The Tsunami Warning system in The Pacific (ICG/ITSU) was formed under the auspices of UNESCO Intergovernmental Oceanographic Commission. After the 2004 tsunami in Indian Ocean, it was renamed as ICG/PTWS (ICG for the Pacific Tsunami Warning and Mitigation System), and similar groups were formed for the other ocean basins: ICG/IOTWS for Indian Ocean, ICG/NEAMTWS for the Northeastern Atlantic and the Mediterranean Sea, and ICG/CARIBE-EWS for the Caribbean Sea.

4.17.7.2 Local Tsunami Warning Systems

For local tsunamis, time is more critical. Japan Meteorological Agency (JMA) is responsible for issuing tsunami warnings in Japan. JMA uses seismic data from a few hundreds of stations to detect tsunamigenic earthquakes, and sea level data to confirm the tsunami generation. There are six regional tsunami warning centers and each of them is staffed with multiple seismologists on duty for 24 h a day and 7 days a week. When an earthquake is detected on the seismic network, the location and size (magnitude) are automatically estimated within a few minutes. Likelihood of tsunami generation is judged from the location and magnitude, and the staff seismologists issue tsunami warning or advisory. The warning or advisory, as well as other related information such as the earthquake location, size, and the expected tsunami arrival times, is disseminated by using the Geostationary Meteorological Satellite.

In the 1950s when JMA first introduced the tsunami warning system, it took about 20 min to issue tsunami warnings (**Figure 16**). In 1983, when a large earthquake occurred in Japan Sea, the warning was issued in 12 min after the earthquake, but the tsunami arrived in 7 min after the earthquake and 100 people died. JMA had improved the warning system, and in 1993 the tsunami warning was issued in 5 min after the South West Hokkaido earthquake. Even though, the tsunami arrived in less than 5 min and the 230 people died (Section 4.17.3.1.2). JMA further improved the system and now issues tsunami warning within 3–5 min after a large earthquake.

It is also very important to improve the accuracy of tsunami forecast and to reduce false alarm. JMA introduced a numerical simulation technique in April 1999 (Tatehata, 1997); tsunami generation and propagation for 100 000 different cases were calculated in advance, and the results have been stored as a database. When a large earthquake occurs, the most appropriate case for an actual location and a magnitude of the earthquake is retrieved from the database.

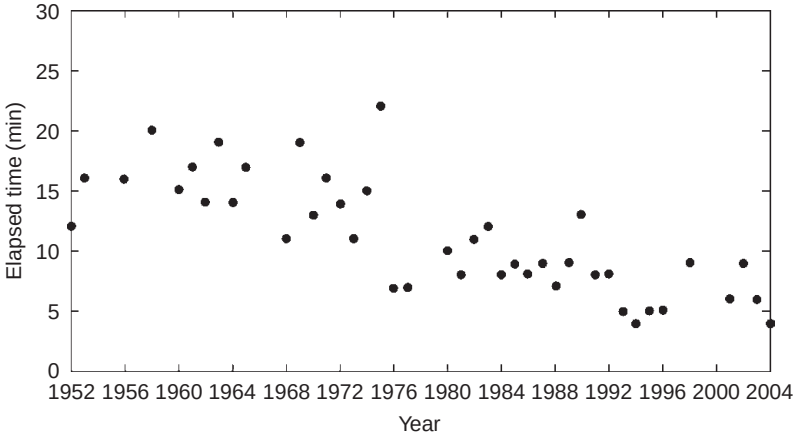


Figure 16 Improvements in elapsed time between earthquakes and tsunami warning issued by Japan Meteorological Agency.

A similar system based on database has been developed for the Pacific Ocean (Titov *et al.*, 2005b).

After a tsunami warning is issued, the information must be disseminated to coastal residents. In Japan, coastal residents receive warning through media, TV and radio, emergency broadcast systems, or patrol cars from police or fire stations. Periodic practice and drills are very important to keep the system functional. In Japan, annual national drills are conducted on 1 September, national disaster prevention day. In May 2006, The ICG/PTWS planned and conducted the first Pacific-wide exercise on issuing, disseminating, and responding to tsunami warning messages.

4.17.7.3 Long-Term and Probabilistic Forecast

The tsunami hazard probabilities can be estimated from historical data. If we know that a particular coast has experienced N tsunamis over a period of time T, assuming a Poisson process, the probability P of having at least one tsunami for future time period τ is given as

$$P = 1 - \exp(-N\tau/T) \qquad [33]$$

For example, if two tsunamis were recorded in the last 100 years at a particular location, the probability of having another tsunami in the next 10 years is 18%. The 10-year probability is 63% if 10 tsunamis occurred in the last 100 years. If the water heights from past tsunamis are known, the probabilities for each range of tsunami heights can be similarly computed. This method is useful for assessing tsunamis from various origins such as trans-Pacific tsunamis.

Great earthquakes at the plate boundaries may recur more regularly than Poisson process, although it is controversial. When the cyclic nature of past earthquakes is known, it can be used to calculate the probability of earthquake occurrence. The time interval between earthquakes can be modeled using a probability density function such as Weibull or log–normal distribution (Rikitake, 1976; Utsu, 1984). The Brownian passage-time (BPT) distribution has been recently used, because it models steady tectonic loading plus Brownian perturbation (Matthews *et al.*, 2002). Then the probability of next earthquake in the specific time window can be computed if the average recurrence period and the date of the most recent event, as well as aperiodicity parameter or variation coefficient, are known.

Probabilistic earthquake forecast had some success in Japan. In March 2003, the Japanese government made a long-term forecast for great ($M \sim 8$) earthquakes along the Kuril trench. The committee estimated that the probability in the next 30 years (starting March 2003) is 60% in Tokachi-oki and 20–30% in Nemuro-oki. This forecast was based on the mean inter-event time of great earthquakes for the southern Kuril region in the nineteenth and twentieth centuries (Earthquake Research Committee, 2004; Satake *et al.*, 2005b). Six months later, on 26 September 2003, a great Tokachi-oki earthquake (M 8.0) actually occurred.

When the fault parameters of the characteristic tsunamigenic earthquake are known, tsunami heights can be estimated by numerical computations, and the probabilistic tsunami heights can be calculated. Such a probabilistic estimate of tsunami heights was made in the 1980s for the Japanese coasts (Rikitake and

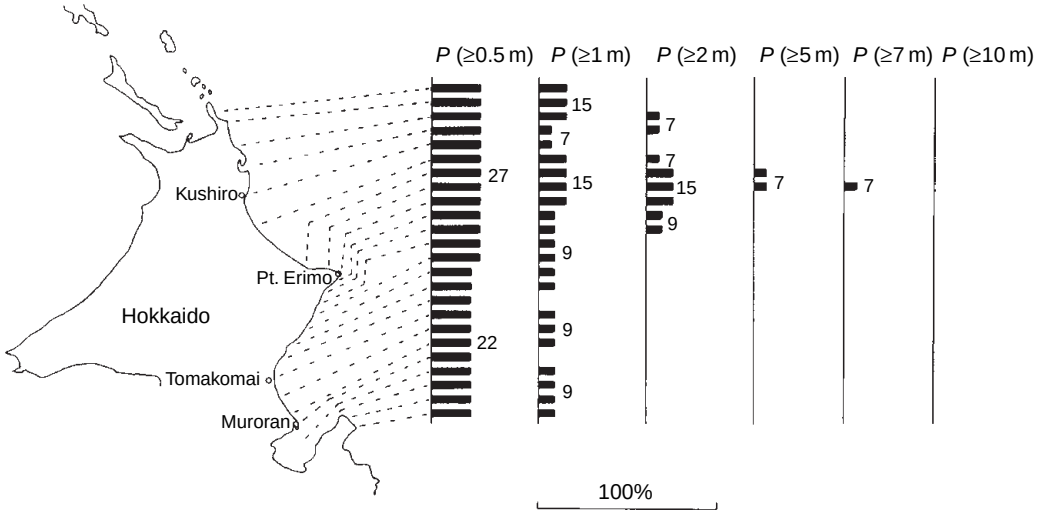


Figure 17 Probabilities for the Pacific coast of Hokkaido hit by a tsunami with wave height exceeding 0.5, 1, 2, 5, 7, and 10 m during a period from 2000 to 2010. The upper half of the coast was hit by a tsunami with 2–4 m in 2003. From Rikitake T and Aida I (1988) Tsunami hazard probability in Japan. *Bulletin of Seismological Society of America* 78: 1268–1278.

Aida, 1988). For the Pacific coast of central Japan where a large earthquake (Tokai earthquake) is expected, the probability of having 5 m tsunami for the period between 2000 and 2010 is as high as 41%. For the Pacific coast of Hokkaido along the Kuril Trench, where the 2003 Tokachi-oki earthquake actually generated the tsunami, the probability of 2 m tsunami was estimated as 7–15% and that of 5 m tsunami was 7% (**Figure 17**). The actual tsunami heights in 2003 were up to 4 m (Tanioka *et al.*, 2004), but the coastal run-up and the tsunami source were different from those of the previous Tokachi-oki earthquake in 1952 (Satake *et al.*, 2006b). Similar probabilistic estimates have also been made for the US coast (Geist and Parsons, 2006).

4.17.7.4 Tsunami Hazard Maps and Public Awareness

Once the coastal residents receive a tsunami warning message, they need to know what it means and where to evacuate. An effective tool for the public is a map showing area to evacuate for a future tsunami. Possible inundation zones are estimated either from historical data of past tsunamis or by tsunami numerical simulation based on maximum credible earthquakes. Safe places such as tsunami shelters to evacuate can also be shown in the tsunami hazard maps. Tsunami hazard map help coastal communities prepare for tsunami hazards.

On the basis of tsunami hazard maps, evacuation plans can be established by responsible officials such as civil defense officials (Gonzalez *et al.*, 2005). One of the lessons from the 2004 Asian tsunami is that not only coastal residents but also foreign tourists need to be informed on tsunami hazards. For the Hawaiian Islands, tsunami hazard maps are prepared for most of the occupied coasts and published in the local phone book that are available at every hotel room (**Figure 18**). The maps show, in addition to the inundation area and places to evacuate, evacuation instructions. For example, those in high-rise buildings are instructed to move vertically to the third or higher floors, rather than moving out of the area.

Tsunami hazard assessment has been carried out on the Japanese coasts. Numerical simulations were used to estimate the time and heights of tsunami and possible damage including casualties. For example, for the future Tonankai-Nankai earthquake off the western Japan, inundation of 40 000 houses and 2200–8600 casualties are estimated (Central Disaster Management Council, 2006).

For public awareness, booklets and video on tsunamis have been prepared and used to educate coastal residents. In Japan, a famous story of a village chief who evacuated the entire residents, by burning his harvest at the time of the 1854 Nankai earthquake tsunami (known as "Inamura-no-Hi"), has been popularly used for hazard education.

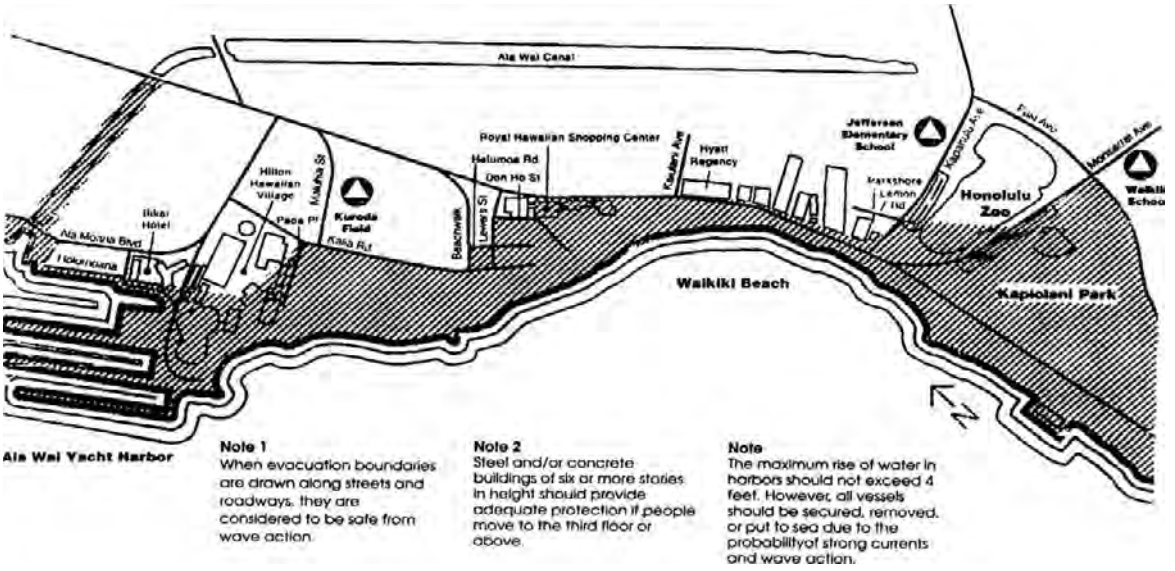


Figure 18 Tsunami evacuation maps for Waikiki beach of Honolulu, Hawaii. The maps were prepared by University of Hawaii in cooperation with Hawaii State Civil Defense Agency and published in local phone books.

4.17.8 Summary

Study of tsunami has interdisciplinary natures as summarized in the chapter. The tsunami sources include giant earthquake such as the 2004 Sumatra–Andaman earthquake or unusual 'tsunami earthquakes' for which more seismological studies are needed. Compared to seismic sources, generation mechanism of tsunamis from landslides or volcanic eruptions is less understood. While the tsunami propagation in deep ocean has been treated by shallow-water theory and numerical computations have been made popularly, the actual propagation in deep ocean was captured by the satellite for the first time at the 2004 Indian Ocean tsunami. Similar deep-sea observations in the future would provide chance to examine the propagation characteristics such as dispersion effects. The coastal behavior of tsunamis will need more studies on theory, field-data, and physical experiments. Geological studies of past tsunamis are also important, as tsunami is one of infrequent phenomena.

The 2004 tsunami in the Indian Ocean demonstrated that the effect of tsunami can be global. The tsunami damage was extended through the Indian Ocean, and the tsunami was recorded throughout the world. In addition, the victims of the 2004 tsunami were not only limited to the coastal residents in Asian countries but also included tourists from Europe. For the reduction of future tsunami damage, it is essential to educate the global citizens about tsunamis. After the 1960 Chilean tsunami, our scientific knowledge on tsunami has significantly improved and we can now detect the occurrence of earthquake and assess the possibility of tsunamis in real time. Such scientific and technological development can be and should be used to save human lives in the world.

References

Abe K (1973) Tsunami and mechanism of great earthquakes. *Physics of the Earth and Planetary Interiors* 7: 143–153.

Abe K (1979) Size of great earthquakes of 1873–1974 inferred from tsunami data. *Journal of Geophysical Research* 84: 1561–1568.

Abe K (1981) Physical size of tsunamigenic earthquakes of the northwestern Pacific. *Physics of the Earth and Planetary Interiors* 27: 194–205.

Abe K (1989) Quantification of tsunamigenic earthquakes by the Mt scale. *Tectonophysics* 166: 27–34.

Abe K (2000) Predominance of long periods in large Pacific tsunamis. *Science of Tsunami Hazards* 18: 15–34.

Aida I (1978) Reliability of a tsunami source model derived from fault parameters. *Journal of Physics of the Earth* 26: 57–73.

Ammon CJ, Ji C, Thio HK, *et al.* (2005) Rupture process of the 2004 Sumatra–Andaman earthquake. *Science* 308: 1133–1139.

Atwater BF, Musumi-Rokkaku S, Satake K, Tsuji Y, Ueda K, and Yamaguchi DK (2005) The Orphan Tsunami of 1700. *USGS Professional Paper* 1707: 133 pp.

Baptista MA, Heitor S, Miranda JM, Miranda P, and Victor LM (1998) The 1755 Lisbon tsunami; Evaluation of the tsunami parameters. *Journal of Geodynamics* 25: 143–157.

Bardet JP, Synolakis CE, Davies HL, Imamura F, and Okal EA (2003) Landslide tsunamis: Recent findings and research directions. *Pure and Applied Geophysics* 160: 1793–1809.

Bernard EN (2005) The US National Tsunami Hazard Mitigation Program: A successful State–Federal partnership. *Natural Hazards* 35: 5–24.

Berninghausen WH (1966) Tsunamis and seismic seiches reported from regions adjacent to the Indian Oceans. *Bulletin of the Seismological Society of America* 56: 69–74.

Bilek SL and Lay T (2002) Tsunami earthquakes possibly widespread manifestations of frictional conditional stability. *Geophysical Research Letters* 29: doi:10.1029/2002GL015215.

Bilham R (2005) A flying start, then a slow slip. *Science* 308: 1126–1127.

Bondevik S, Lovholt F, Harbitz C, Mangerud J, Dawson A, and Svendsen JI (2005) The Storegga Slide tsunami – comparing field observations with numerical simulations. *Marine and Petroleum Geology* 22: 195–208.

Borrero JC (2005a) Field survey of Northern Sumatra and Banda Aceh, Indonesia after the tsunami and earthquake of 26 December 2004. *Seismological Research Letters* 76: 312–320.

Borrero JC (2005b) Field data and satellite imagery of tsunami effects in Banda Aceh. *Science* 308: 1596.

Bryant EA, Young RW, and Price DM (1996) Tsunami as a major control on coastal evolution, southeastern Australia. *Journal of Coastal Research* 12: 831–840.

Bugge T, Belderson RH, and Kenyon NH (1988) The Storegga slide. *Philosophical Transactions of the Royal Society of London Series A* 325: 357–388.

Central Disaster Management Council (2006) *Hazard Estimation for Earthquakes Along Japan and Kuril Trenches (in Japanese)*. Tokyo: Cabinets Office.

Comer RP (1980) Tsunami height and earthquake magnitude: theoretical basis of an empirical relation. *Geophysical Research Letters* 7: 445–448.

Comer RP (1984a) The tsunami mode of a flat earth and its excitation by earthquake sources. *Geophysical Journal of the Royal Astronomical Society* 77: 1–27.

Comer RP (1984b) Tsunami generation: a comparison of traditional and normal mode approaches. *Geophysical Journal of the Royal Astronomical Society* 77: 29–41.

Cummins P (2004) Small threat, but warning sounded for tsunami research. *AusGeo News* 75: 4–7.

Dawson AG, Long D, and Smith DE (1988) The Storegga Slides – Evidence from eastern Scotland for a possible tsunami. *Marine Geology* 82: 271–276.

Dawson AG and Shi SZ (2000) Tsunami deposits. *Pure and Applied Geophysics* 157: 875–897.

Earthquake Research Committee (2004) Long-term evaluation of seismicity along the Kuril Trench (in Japanese). *Publications of Earthquake Research Committee,* pp. 1–74 . Tokyo: Earthquake Research Committee.

Fine, IV, Rabinovich AB, Bornhold BD, Thomson RE, and Kulikov EA (2005a) The Grand Banks landslide-generated tsunami of November 18, 1929: preliminary analysis and numerical modeling. *Marine Geology* 215: 45–57.

Fine, IV, Rabinovich AB, and Thomson RE (2005b) The dual source region for the 2004 Sumatra tsunami. *Geophysical Research Letters* 32: L16602, doi: 10.1029/2005GL023521.

Fujii Y and Satake K (2007) Tsunami source model of the 2004 Sumatra–Andaman earthquake inferred from tide gauge and satellite data. *Bulletin of the Seismological Society of America* 97: S192–S207.

Fukao Y (1979) Tsunami earthquakes and subduction processes near deep-sea trenches. *Journal of Geophysical Research* 84: 2303–2314.

Geist E (1998) Local tsunamis and earthquake source parameters. *Advances in Geophysics* 39: 117–209.

Geist E (2000) Origin of the 17 July 1998 Papua New Guinea tsunami; earthquake or landslide? *Seismological Research Letters* 71: 344–351.

Geist E and Parsons T (2006) Probabilistic analysis of tsunami hazards. *Natural Hazards* 37: 277–314.

Geist E, Titov VV, Arcas D, Pollitz FF, and Bilek SL (2007) Implications of the December 26, 2004 Sumatra–Andaman Earthquake on Tsunami Forecast and Assessment Models for Great Subduction Zone Earthquakes. *Bulletin of the Seismological Society of America* 97: S249–S270.

Gelfenbaum G and Jaffe B (2003) Erosion and sedimentation from the 17 July, 1998 Papua New Guinea tsunami. *Pure and Applied Geophysics* 160: 1969–1999.

Gonzalez FI, Bernard EN, Meinig C, Eble MC, Mofjeld HO, and Stalin S (2005) The NTHMP tsunameter network. *Natural Hazards* 35: 25–39.

Gower J (2005) Jason 1 detects the 26 December 2004 tsunami. *Eos, Transactions, American Geophysical Union* 86: 37–38.

Gusiakov VK (2005) Tsunami generation potential of different tsunamigenic regions in the Pacific. *Marine Geology* 215: 3–9.

Hanks T and Kanamori H (1979) A moment magnitude scale. *Journal of Geophysical Research* 84: 2348–2350.

Hanson JF, Reasoner C, and Bowman JR (2007) High frequency tsunami signals of the great Indonesian earthquakes of 26 December and 28 March 2005. *Bulletin of the Seismological Society of America* 97: S232–S248.

Harbitz CB (1992) Model simulations of tsunamis generated by the Storegga slides. *Marine Geology* 105: 1–21.

Hatori T (1979) Relation between tsunami magnitude and wave energy. *Bulletin of the Earthquake Research Institute (University of Tokyo)* 54: 531–541.

Hatori T (1984) On the damage to houses due to tsunamis. *Bulletin of the Earthquake Research Institute (University of Tokyo)* 59: 433–439.

Heezen BC and Ewing M (1952) Turbidity currents and submarine slumps, and the 1929 Grand Banks earthquake. *American Journal of Science* 250: 849–873.

Heinrich P, Piatanesi A, and Hebert H (2001) Numerical modelling of tsunami generation and propagation from submarine slumps: the 1998 Papua New Guinea event. *Geophysical Journal International* 145: 97–111.

Hirata K, Aoyagi M, Mikada H, *et al.* (2002) Real-time geophysical measurements on the deep seafloor using submarine cable in the southern Kurile subduction zone. *IEEE Journal of Oceanic Engineering* 27: 170–181.

Hirata K, Satake K, Tanioka Y, *et al.* (2006) The 2004 Indian Ocean tsunami: Tsunami source model from satellite altimetry. *Earth, Planets and Space* 58: 195–201.

Iida K (1984) *Catalog of Tsunamis in Japan and Its Neighboring Countries*, 52pp. Toyato: Aichi Institute of Technology.

Iida K, Cox DC, and Pararas-Carayannis G (1967) *Preliminary catalog of tsunamis occurring in the Pacific Ocean*. Honolulu: University of Hawaii, 274pp.

Imamura F, Hashi K, and Imteaz MMA (2001) Modeling for tsunamis generated by landsliding and debris flow. In: Hebenstreit GT (ed.) *Tsunami Research at the End of a Critical Decade,*, pp. 209–228. Dordrecht: Kluwer Academic.

Institute of Computational Mathematics and Mathematical Geophysics (2005) Integrated Tsunami Database for the World Ocean, Novosibirsk, CD-ROM.

Intergovernmental Oceanographic Commission (1998) *Manuals and Guides 37: Post-tsunami Survey Field Guide*. Paris: UNESCO.

International Federation of Red Cross and Red Crescent Societies (2005) *World Disasters Report*, 251pp. Geneva: International Federation of Red Cross and Red Crescent Societies.

Ishii M, Shearer PM, Houston H, and Vidale JE (2005) Extent, duration and speed of the 2004 Sumatra–Andaman

earthquake imaged by the Hi-Net array. *Nature* 435: 933–936.
Iwasaki SI, Eguchi T, Fujinawa Y, *et al.* (1997) Precise tsunami observation system in the deep ocean by an ocean bottom cable network for the prediction of earthquakes and tsunamis. In: Hebenstreit G (ed.) *Perspectives on Tsunami Hazard Reduction*, pp. 47–66. Dordrecht: Kluwer Academic.
Jaffe BE, Borrero JC, Prasetya GS, *et al.* (2006) Northwest Sumatra and offshore islands field survey after the December 2004 Indian Ocean tsunami. *Earthquake Spectra* 22: S105–S135.
Jiang L and Leblond PH (1994) 3-Dimensional modeling of tsunami generation due to a submarine mudslide. *Journal of Physical Oceanography* 24: 559–572.
Johnson JM (1999) Heterogeneous coupling along Alaska-Aleutians as inferred from tsunami, seismic and geodetic inversions. *Advances in Geophysics* 39: 1–116.
Johnson JM and Satake K (1997) Estimation of seismic moment and slip distribution of the April 1, 1946, Aleutian tsunami earthquake. *Journal of Geophysial Research* 102: 11765–11774.
Kajiura K (1963) The leading wave of a tsunami. *Bulletin of the Earthquake Research Institute (University of Tokyo)* 41: 535–571.
Kajiura K (1970) Tsunami source, energy and the directivity of wave radiation. *Bulletin of the Earthquake Research Institute (University of Tokyo)* 48: 835–869.
Kanamori H (1972) Mechanism of tsunami earthquakes. *Physics of the Earth and Planetary Interiors* 6: 246–259.
Kanamori H and Kikuchi M (1993) The 1992 Nicaragua earthquake: A slow tsunami earthquake associated with subducted sediments. *Nature* 361: 714–716.
Kato T, Terada Y, Ito K, *et al.* (2005) Tsunami due to the 2004 September 5th off Kii peninsula earthquake, Japan, recorded by a new GPS buoy. *Earth, Planets and Space* 57: 297–301.
Kato T, Terada Y, Kinoshita M, *et al.* (2000) Real-time observation of tsunami by RTK-GPS. *Earth, Planets and Space* 52: 841–845.
Kawata Y, Benson B, Borrero JC, *et al.* (1999) Tsunami in Papua New Guinea was as intense as first thought. *Eos, Transactions, American Geophysical Union* 80: 101–105.
Kayanne H, Ikeda Y, Echigo T, *et al.* (2007) Coseismic and postseismic creep in Andaman Islands associated with the 2004 Sumatra–Andaman earthquake. *Geophysical Research Letters* 34: L01310, doi: 10.129/2006GL028200.
Keating BH, Waythomas CF, and Dawson AG (2000) *Special issue: Landslides and Tsunamis–Introduction. Pure and Applied Geophysics* 157: 871–873.
Kulikov EA, Rabinovich AB, Thomson RE, and Bornhold BD (1996) The landslide tsunami of November 3, 1994, Skagway Harbor, Alaska. *Journal of Geophysical Research* 101: 6609–6615.
Lander JF (1996) *Tsunamis Affecting Alaska: 1737–1996.* NGDC Key to Geophysical Research Documentation No. 29, Boulder, CO.
Lander JF and Lockridge PA (1989) *United States Tsunamis*. Boulder, CO: National Geophysical Data Center.
Lander JF, Lockridge PA, and Kozuch MJ (1993) *Tsunamis Affecting the West Coast of the United States 1806–1992*. Boulder, CO: National Geophysical Data Center.
Latter JH (1981) Tsunamis of volcanic origin: summary of causes, with particular reference to Krakatoa, 1883. *Bulletin of Volcanology* 44: 467–490.
Lay T, Kanamori H, Ammon CJ, *et al.* (2005) The great Sumatra-Andaman earthquake of 26 December 2004. *Science* 308: 1127–1133.
Liu PLF, Lynett P, Fernando H, *et al.* (2005) Observations by the International Tsunami Survey Team in Sri Lanka. *Science* 308: 1595–1595.
Malik JN and Murty CVR (2005) Landscape changes in Andaman and Nicobar Islands (India) due to Mw 9.3 tsunamigenic Sumatra earthquake of 26 December 2004. *Current Science* 88: 1384–1386.
Mansinha L and Smylie DE (1971) The displacement fields of inclined faults. *Bulletin of the Seismological Society of America* 61: 1433–1440.
Matsuyama M, Walsh JP, and Yeh H (1999) The effect of bathymetry on tsunami characteristics at Sisano Lagoon, Papua New Guinea. *Geophysical Research Letters* 26: 3513–3516.
Matthews MV, Ellsworth WL, and Reasenberg PA (2002) A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America* 92: 2233–2250.
McCreery CS (2005) Impact of the National Tsunami Hazard Mitigation Program on operations of the Richard H. Hagemeyer Pacific Tsunami Warning Center. *Natural Hazards* 35: 73–88.
Mei CC (1989) *The Applied Dynamics of Ocean Surface Waves.* Advances Series on Ocean Engineering, volume 1. Singapore: World Scientific.
Meltzner AJ, Sieh K, Abrams M, *et al.* (2006) Uplift and subsidence associated with the great Aceh–Andaman earthquake of 2004. *Journal of Geophysical Research* 111: B02407 (doi: 10.1029/2005JB003891).
Merrifield MA, Firing YL, Aarup T, *et al.* (2005) Tide gauge observations of the Indian Ocean tsunami, December 26, 2004. *Geophysical Research Letters* 32, doi: 10.1029/2005GL022610.
Miller DJ (1960) Giant waves in Lituya Bay, Alaska. *USGS Professional Paper* 354-C: 51–86.
Minoura K and Nakaya S (1991) Traces of tsunami preserved in inter-tidal lacustrine and marsh deposits - Some examples from northeast Japan. *The Journal of Geology* 99: 265–287.
Miyabe N (1934) An investigation of the Sanriku tunami based on mareogram data. *Bulletin of the Earthquake Research Institute (University of Tokyo)* 1: 112–126.
Moore A, Nishimura Y, Gelfenbaum G, Kamataki T, and Triyono R (2006) Sedimentary deposits of the 26 December 2004 tsunami on the northwest. *Earth, Planets and Space* 58: 253–258.
Murty TS and Rafiq M (1991) A tentative list of tsnamis in the marginal seas of the North Indian Ocean. *Natural Hazards* 4: 81–83.
Nagai T, Ogawa H, Nukada K, and Kudaka M (2004) Characteristics of the observed 2003 Tokachi-oki earthquake tsunami profile. *Coastal Engineering Journal* 46: 315–327.
Nagarajan B, Suresh I, Sundar D, *et al.* (2006) The Great Tsunami of 26 December 2004: A description based on tide-gauge data from the Indian subcontinent and surrounding areas. *Earth, Planets and Space* 58: 211–215.
Nanayama F, Satake K, Furukawa R, *et al.* (2003) Unusually large earthquakes inferred from tsunami deposits along the Kuril trench. *Nature* 424: 660–663.
Nanayama F, Shigeno K, Satake K, *et al.* (2000) Sedimentary differences between the 1993 Hokkaido-nansei-oki tsunami and the 1959 Miyakojima typhoon at Taisei, southwestern Hokkaido, northern Japan. *Sedimentary Geology* 135: 255–264.
Neetu S, Suresh I, Shankar R, *et al.* (2005) Comment on ''The Great Sumatra-Andaman Earthquake of 26 December 2004''. *Science* 310: 1431.
Newman AV and Okal EA (1998) Teleseismic estimates of radiated seismic energy: The E/M-0 discriminant for tsunami earthquakes. *Journal of Geophysical Research* 103: 26885–26898.
Nishimura Y and Miyaji N (1995) Tsunami deposits from the 1993 Southwest Hokkaido earthquake and the 1640

Hokkaido-Komagatake eruption, northern Japan. *Pure and Applied Geophysics* 144: 719–733.
Nomanbhoy N and Satake K (1995) Generation mechanism of tsunamis from the 1883 Krakatau eruption. *Geophysical Research Letters* 22: 509–512.
Okada M (1995) Tsunami observation by ocean bottom pressure gauge. In: Tsuchiya Y, Shuto N (eds.) *Tsunami: Progress in Prediction, Disaster Prevention and Warning*, pp. 287–303. Dordrecht: Kluwer Academic.
Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America* 75: 1135–1154.
Okal EA (1988) Seismic parameters controlling far-field tsunami amplitudes: A review. *Natural Hazards* 1: 67–96.
Okal EA and Newman AV (2001) Tsunami earthquakes: The quest for a regional signal. *Physics of the Earth and Planetary Interiors* 124: 45–70.
Okal EA, Plafker G, Synolakis CE, and Borrero JC (2003) Near-field survey of the 1946 Aleutian Tsunami on Unimak and Sanak Islands. *Bulletin of the Seismological Society of America* 93: 1226–1234.
Okal EA and Synolakis CE (2004) Source discriminants for near-field tsunamis. *Geophysical Journal International* 158: 899–912.
Okal EA, Synolakis CE, Fryer GJ, *et al.* (2002) A field survey of the 1946 Aleutian tsunami in the far field. *Seismological Research Letters* 73: 490–503.
Okal EA and Talandier J (1989) M_m: A variable-period mantle magnitude. *Journal of Geophysical Research* 94: 4169–4193.
Ortiz M and Bilham R (2003) Source area and rupture parameters of the 31 December 1881 M-w = 7.9 Car Nicobar earthquake estimated from tsunamis recorded in the Bay of Bengal. *Journal of Geophysical Research* 108: (doi: 10.129/2002JB001941).
Pelayo AM and Wiens DA (1992) Tsunami earthquakes: slow thrust-faulting events in the accretionary wedge. *Journal of Geophysical Research* 97: 15321–15337.
Pelinovsky E, Choi BH, Stromkov A, Didenkulova I, and Kim H-S (2005) Analysis of tide-gauge records of the 1883 Krakatau tsunami. In: Satake K (ed.) *Tsunamis: Case Studies and Recent Developments,*, pp. 57–78. Dordrecht: Springer.
Peregrine H (1972) Equations for water waves and the approximations behind them. In: Meyer RE (ed.) *Waves on Beaches and Resulting Sediment Transport,* pp. 95–121. Academic Press.
Piatanesi A and Lorito S (2007) Rupture process of the 2004 Sumatra–Andaman earthquake from tsunami waveform inversion. *Bulletin of the Seismological Society of America* 97: S223–S231.
Plafker G, Kachadoorian R, Eckel EB, and Mayo LR (1969) Effects of the earthquake of March 27, 1964 on various communities. *USGS Professional Paper* 542-G: 1–50.
Polet J and Kanamori H (2000) Shallow subduction zone earthquakes and their tsunamigenic potential. *Geophysical Journal International* 142: 684–702.
Press F and Harkrider D (1966) Air-sea waves from the explosion of Krakatoa. *Science* 154: 1325–1327.
Reymond D, Hyvernaud O, and Talandier J (1991) Automatic detection, location and quantification of earthquakes: application to tsunami warning. *Pure and Applied Geophysics* 135: 361–382.
Rikitake T (1976) Recurrence of great earthquakes at subduction zones. *Tectonophysics* 35: 335–362.
Rikitake T and Aida I (1988) Tsunami hazard probability in Japan. *Bulletin of the Seismological Society of America* 78: 1268–1278.
Satake K (1988) Effects of Bathymetry on Tsunami Propagation - Application of Ray Tracing to Tsunamis. *Pure and Applied Geophysics* 126: 27–36.
Satake K (1989) Inversion of tsunami waveforms for the estimation of heterogeneous fault motion of large submarine earthquakes – the 1968 Tokachi-Oki and 1983 Japan Sea earthquakes. *Journal of Geophysical Research* 94: 5627–5636.
Satake K (1993) Depth distribution of coseismic slip along the Nankai Trough, Japan, from joint inversion of geodetic and tsunami data. *Journal of Geophysical Research* 98: 4553–4565.
Satake K (1994) Mechanism of the 1992 Nicaragua tsunami earthquake. *Geophysical Research Letters* 21: 2519–2522.
Satake K (1995) Linear and nonlinear computations of the 1992 Nicaragua earthquake tsunami. *Pure and Applied Geophysics* 144: 455–470.
Satake K, Aung TT, Sawai Y, *et al.* (2006a) Tsunami heights and damage along the Myanmar coast from the December 2004 Sumatra-Andaman earthquake. *Earth, Planets and Space* 58: 243–252.
Satake K, Baba T, Hirata K, *et al.* (2005a) Tsunami source of the 2004 off the Kii peninsula earthquakes inferred from offshore tsunami and coastal tide gauges. *Earth, Planets and Space* 57: 173–178.
Satake K, Hirata K, Yamaki S, and Tanioka Y (2006b) Re-estimation of tsunami source of the 1952 Tokachi-oki earthquake. *Earth, Planets and Space* 58: 535–542.
Satake K, Nanayama F, Yamaki S, Tanioka Y, and Hirata K (2005b) Variability among tsunami sources in the 17th–21st centuries along the southern Kuril trench. In: Satake K (ed.) *Tsunamis: Case Studies and Recent Developments*, pp. 157–170. Dordrecht: Springer.
Satake K, Okada M, and Abe K (1988) Tide gauge response to tsunamis: Measurements at 40 tide gauge stations in Japan. *Journal of Marine Research* 46: 557–571.
Satake K, Shimazaki K, Tsuji Y, and Ueda K (1996) Time and site of a giant earthquake in Cascadia inferred from Japanese tsunami records of January 1700. *Nature* 379: 246–249.
Satake K, Smith JR, and Shinozaki K (2002) Three-dimensional reconstruction and tsunami model of the Nuuanu and Wailau landslides, Hawaii. In: Takahashi E, Lipman P, Garcia M, Naka J, and Aramaki S (eds.) *Hawaiian Volcanoes: Deep Underwater Perspectives.* 128, pp. 333–346. Washington D.C: American Geophysical Union.
Satake K and Tanioka Y (1999) Sources of tsunami and tsunamigenic earthquakes in subduction zones. *Pure and Applied Geophysics* 154: 467–483.
Satake K and Tanioka Y (2003) The July 1998 Papua New Guinea earthquake: Mechanism and quantification of unusual tsunami generation. *Pure and Applied Geophysics* 160: 2087–2118.
Satake K, Wang KL, and Atwater BF (2003) Fault slip and seismic moment of the 1700 Cascadia earthquake inferred from Japanese tsunami descriptions. *Journal of Geophysical Research* 108: (doi: 10.1029/2003JB002521).
Seno T (2002) Tsunami earthquake as transient phenomena. *Geophysical Research Letters* 29: (doi: 10.029/2002GL014868).
Seno T and Hirata K (2007) Did the 2004 Sumatra–Andaman earthquake involve a component of tsunami earthquakes? *Bulletin of the Seismological Society of America* 97: S296–S306.
Shuto N (1991) Numerical simulation of tsunamis – Its present and near future. *Natural Hazards* 4: 171–191.
Shuto N and Matsutomi H (1995) Field survey of the (1993) Hokkaido-Nansei-Oki earthquake tsunami. *Pure and Applied Geophysics* 144: 649–663.
Simkin T and Fiske R (1983) *Krakatau, 1883 – the volcanic eruption and its effects*. Washington, D.C: Smithsonian Institution Press.
Singh SK, Ortiz M, Gupta HK, and Ramadass DGA (2006) Slow slip below Port Blair, Andaman, during the great Sumatra–

Andaman earthquake of 26 December 2004. *Geophysical Research Letters* 33: (doi: 10.1029/2005GL025025).
Slingerland R and Voight B (1979) Occurrence, properties, and predictive models of landslide-generated water waves. In: Voight B (ed.) *Rockslides and Avalanches*, 1, pp. 317–397. Amsterdam: Elsevier.
Smith DE, Shi S, Cullingford RA, *et al.* (2004) The Holocene Storegga slide tsunami in the United Kingdom. *Quaternary Science Reviews* 23: 2291–2321.
Smith WHF, Scharroo R, Titov VV, Arcas D, and Arbic BK (2005) Satellite altimeters measure tsunami, early model estimates confirmed. *Oceanography* 18: 10–12.
Soloviev SL (1970) Recurrence of tsunamis in the Pacific. In: Adams W M (ed.) *Tsunamis in the Pacific Ocean*, pp. 149–164. Honolulu: East-West Center Press.
Soloviev SL and Go CN (1974a) *Catalogue of Tsunamis on the Western Shore of the Pacific Ocean (173–1968)*. Moscow: Nauka Publishing House.
Soloviev SL and Go CN (1974b) *Catalogue of Tsunamis on the Eastern shore of the Pacific Ocean (1513–1968)*. Moscow: Nauka Publishing House.
Soloviev SL, Go CN, and Kim KS (1992) *Catalog of Tsunamis in the Pacific 1969–1982*. Moscow: Academy of Science of USSR.
Song YT, Ji C, Fu L-L, *et al.* (2005) The 26 December 2004 tsunami source estimated from satellite radar altimetry and seismic waves. *Geophysical Research Letters* 32, doi: 10.1029/2005GL023683.
Stein S and Okal EA (2005) Speed and size of the Sumatra earthquake. *Nature* 434: 581–582.
Steketee JA (1958) On Volterra's dislocations in a semi-infinite elastic medium. *Canadian Journal of Physics* 36: 192–205.
Synolakis CE, Bardet JP, Borrero JC, *et al.* (2002) The slump origin of the 1998 Papua New Guinea tsunami. *Proceedings of the Royal society of London Series A* 458: 763–789.
Synolakis CE and Okal EA (2005) 1992–2002: Perspective on a decade of post-tsunami surveys. In: Satake K (ed.) *Tsunamis: Case Studies and Recent Developments*, pp. 1–29. Dordrecht: Springer.
Takahashi T, Shuto N, Imamura F, and Ortiz M (1995) Source models for the 1993 Hokkaido-Nansei-Oki earthquake tsunami. *Pure and Applied Geophysics* 144: 747–767.
Tanioka Y, Nishimura Y, Hirakawa K, *et al.* (2004) Tsunami run-up heights of the 2003 Tokachi-oki earthquake. *Earth, Planets and Space* 56: 359–365.
Tanioka Y, Ruff L, and Satake K (1997) What controls the lateral variation of large earthquake occurrence along the Japan trench? *Island Arc* 6: 261–266.
Tanioka Y and Satake K (1996a) Tsunami generation by horizontal displacement of ocean bottom. *Geophysical Research Letters* 23: 861–864.
Tanioka Y and Satake K (1996b) Fault parameters of the 1896 Sanriku tsunami earthquake estimated from tsunami numerical modeling. *Geophysical Research Letters* 23: 1549–1552.
Tanioka Y and Sataka K (2001a) Coseismic slip distribution of the 1946 Nankai earthquake and aseismic slips caused by the earthquake. *Earth, Planets and Space* 53: 235–241.
Tanioka Y and Satake K (2001b) Detailed coseismic slip distribution of the 1944 Tonankai earthquake estimated from tsunami waveforms. *Geophysical Research Letters* 28: 1075–1078.
Tanioka Y, Satake K, and Ruff L (1995) Total analysis of the 1993 Hokkaido-Nansei-Oki earthquake using seismic-wave, tsunami, and geodetic data. *Geophysical Research Letters* 22: 9–12.
Tanioka Y and Seno T (2001a) Sediment effect on tsunami generation of the 1896 Sanriku tsunami earthquake. *Geophysical Research Letters* 28: 3389–3392.
Tanioka Y and Seno T (2001b) Detailed analysis of tsunami waveforms generated by the 1946 Aleutian tsunami earthquake. *Natural Hazards and Earth System Sciences* 1: 171–175.
Tanioka Y, Yudhicara, Kusunose T, *et al.* (2006) Rupture process of the 2004 great Sumatra-Andaman earthquake estimated from tsunami waveforms. *Earth, Planets and Space* 58: 203–209.
Tappin DR (2004) Special issue: Submarine-slump-generated tsunamis – Preface. *Marine Geology* 203: 199–200.
Tappin DR, Matsumoto P, Watts P, *et al.* (1999) Sediment slump likely caused 1998 Papua New Guinea tsunami. *Eos, Transactions, American Geophysical Union* 80: 329, 334, 340.
Tatehata H (1997) The new tsunami warning system of the Japan Meteorological Agency. In: Hebenstreit G (ed.) *Perspectives on Tsunami Hazards Reduction*, pp. 175–188. Dordrecht: Kluwer.
Tilling RI (1989) Volcanic Hazards and Their Mitigation - Progress and Problems. *Reviews of Geophysics* 27: 237–269.
Tinti S, Bortolucci E, and Vannini C (1997) A block-based theoretical model suited to gravitational sliding. *Natural Hazards* 16: 1–28.
Titov V, Rabinovich AB, Mofjeld HO, Thomson RE, and Gonzalez FI (2005a) The global reach of the 26 December 2004 Sumatra tsunami. *Science* 309: 2045–2048.
Titov VV and Gonzalez FI (2001) Numerical study of the source of the July 17, 1998 PNG tsunami. In: Hebensreit GT (ed.) *Tsunami Research at the End of a Critical Decade*, pp. 197–207. Dordrecht: Kluwer Academic.
Titov VV, Gonzalez FI, Bernard EN, *et al.* (2005b) Real-time tsunami forecasting: Challenges and solutions. *Natural Hazards* 35: 41–58.
Titov VV and Synolakis CE (1997) Extreme inundation flows during the Hokkaido-Nansei-Oki tsunami. *Geophysical Research Letters* 24: 1315–1318.
Tobita M, Suito H, Imakiire T, Kato M, Fujiwara S, and Murakami M (2006) Outline of vertical displacement of the 2004 and 2005 Sumatra earthquakes revealed by satellite radar imagery. *Earth, Planets and Space* 58: e1–e4.
Tsai VC, Nettles M, Ekstrom G, and Dziewonski AM (2005) Multiple CMT source analysis of the 2004 Sumatra earthqauke. *Geophysical Research Letters* 32: (doi: 10.1029/2005GL023813).
Tsuboi S, Abe K, Takano K, Yamanaka Y (1995) Rapid-Determination of Mw from Broad-Band P-Wave-Forms. *Bulletin of the Seismological Society of America* 85: 606–613.
Tsuji Y, Namegaya Y, Matsumoto H, *et al.* (2006) The 2004 Indian tsunami in Thailand: Surveyed runup heights and tide gauge records. *Earth, Planets and Space* 58: 223–232.
Tuttle MP, Ruffman A, Anderson T, and Jeter H (2004) Distinguishing tsunami from storm deposits in eastern North America: The 1929 grand banks tsunami versus the 1991 Halloween storm. *Seismological Research Letters* 75: 117–131.
Unoki S and Nakano M (1953) On the Cauchy-Poisson waves caused by the eruption of a submarine volcano (III). *Oceanographical Magazine* 4: 139–150.
Utsu T (1984) Estimation of parameters for recurrence models of earthquakes. *Bulletin of the Earthquake Research Institute (University of Tokyo)* 59: 53–66.
Ward S (1980) Relationships of tsunami generation and an earthquake source. *Journal of Physics of the Earth* 28: 441–474.
Ward SN (2001) Landslide tsunami. *Journal of Geophysical Research* 106: 11201–11215.
Ward SN and Day S (2003) Ritter Island Volcano – lateral collapse and the tsunami of 1888. *Geophysical Journal International* 154: 891–902.

Watanabe H (1998) *Comprehensive List of Tsunamis to Hit the Japanese Islands*. Tokyo: University of Tokyo Press.
Watts P, Grilli ST, Kirby JT, Fryer GJ, and Tappin DR (2003) Landslide tsunami case studies using a Boussinesq model and a fully nonlinear tsunami generation model. *Natural Hazards Earth System Sciences* 3: 391–402.
Woodworth PL and Player R (2003) The permanent service for mean sea level: An update to the 21st century. *Journal of Coastal Research* 19: 287–295.
Yamashita T and Sato R (1974) Generation of tsunami by a fault model. *Journal of Physics of the Earth* 22: 415–440.
Yeh H, Imamura F, Synolakis C, Tsuji Y, Liu P, and Shi S (1993) The Flores Island tsunamis. *Eos, Transactions, American Geophysical Union* 74: 369, 371–373.
Yuan XH, Kind R, and Pedersen HA (2005) Seismic monitoring of the Indian Ocean tsunami. *Geophysical Research Letters* 32: L15308, doi: 10.1029/2005GL023464.

Relevant Websites

http://tsun.sscc.ru – Institute of Computational Mathematics and Mathematical Geophysics, Novosibirsk.
http://www.ngdc.noaa.gov – National Geophysical Data Center, NOAA.

4.18 Physical Processes That Control Strong Ground Motion

J. G. Anderson, University of Nevada, Reno, NV, USA

4.18.1 Introduction

Strong-motion seismology is the branch of seismology specializing in understanding ground motions near large earthquakes. It is important first for the practical expectation of our civilization that structures will be built to withstand earthquakes, and this cannot be done without knowing the character of those ground motions. The second reason it is important is that the ground motions recorded near the earthquake are generally least filtered and modified by transmission through the Earth, and thus give the best data for scientific study of the earthquake source.

The goal of strong-motion seismology is to be able to predict motions near an earthquake sufficiently well that the predictions can be used for engineering applications, both for scientific and practical reasons. The scientific motivation is that the ability to predict the motions is evidence that the phenomena are understood. The practical reasons are the need to know the character of ground motions at a site of engineering interest, given that the sources and sites, and the geological features between, are unique to every situation. In most cases the character of strong ground motions is approximately understandable, and the amplitudes of the motions critical for engineering designs are predictable within modest uncertainties.

Some recent relatively extensive reviews of aspects of research in the field of strong-motion seismology include Takemura *et al.* (1995) and Anderson (2003). The volume compiled by Lee *et al.* (2003) includes several articles that give detailed reviews of aspects of the field. While there is overlap of topics considered here and in those other recent reviews, the emphasis in this article is on the physical processes that affect strong ground motions.

4.18.2 What Is Strong Ground Motion?

The field of strong-motion seismology could initially be identified with a type of instrument, designed to remain on scale and record the ground motion with fidelity under the conditions of the strongest ground motions. The history of instrumentation in the United States is reviewed by Trifunac and Todorovska (2001). The first US instruments were deployed in 1932 by the US Coast and Geodetic Survey. The history of the Japanese program, reviewed by Takemura *et al.* (1995), began in 1951 and the first significant record was obtained in 1956 (Takahashi, 1956). Early instruments recorded analog traces proportional to the acceleration of the ground, and thus were called strong-motion accelerographs. They were typically designed so that ground motions up to the acceleration of gravity ($1g$) would be on scale. Going off-scale meant that a light beam moved off the edge of the recording film, or that a pen scratching a waxed paper swung off the edge of the paper. In one widely used compact photographic accelerograph, while the instrument was at rest, the light beam would be at least 2 cm from the edge of the film, so the instrument was designed such that a 2 cm deflection occurred under $1g$ acceleration. The thickness of the light beam defined the minimum ground motion that could be recorded. Supposing that a well-focused light beam exposes the film in a line 0.5 mm wide, and that the center of the beam can be estimated to be 0.2 mm, the minimum acceleration resolved in this case was about $0.01g$. This roughly defined the lower limit of ground motion considered by the early strong-motion seismology studies. Conveniently, $0.01g$ is only a little stronger than the lower limit of ground motions that a person at rest is able to feel. Thus the early range of strong-motion seismology approximately coincided with ground motions that humans are able to feel.

The main interest in strong-motion seismology was driven by the engineering need to design structures safely. Normally, ground motions that cause damage exceed about $0.1g$. The first significant strong-motion accelerogram, recorded in Long Beach, California, during the 10 March 1933 Long Beach earthquake ($M_W = 6.4$) had peak horizontal and vertical accelerations of $0.19g$ and $0.28g$, respectively.

Modern strong-motion instruments use digital recording with at least 16 bit and typically 24 bit analog-to-digital conversion (e.g., Kinoshita, 2003; Shin *et al.*, 2003). With a 24 bit instrument, assuming conservatively that there are at least 19 bits above noise sources, the least significant bit in a $1g$ accelerograph is reduced to under $4 \times 10^{-6} g$, or $<0.004\ \mathrm{cm\ s^{-2}}$. Where the dynamic range of the analog instrument was about 40 db ($=20 \log_{10}(1g/0.01g)$), the dynamic range of the current generation of digital accelerographs is between 110 and 135 db (Trifunac and Todorovska, 2001). Since much smaller ground motions can be recorded on these instruments than on the earlier generations of analog instruments, the distinction between strong-motion and traditional seismology is blurred.

Processing of strong-motion accelerograms is reviewed by Takemura *et al.* (1995) and Shakal *et al.* (2003). Low-frequency pulses that carry a permanent displacement of the ground associated with slip on the fault are present in some cases of strong ground motion, but never present in weak-motion seismology. The static displacement has been recovered in a few cases by double integration of digital strong-motion accelerograms with sufficient resolution (e.g., Anderson *et al.*, 1986; Anderson *et al.*, 2000; Ma *et al.*, 2001; Boore *et al.*, 2002).

4.18.2.1 Parameters to Characterize Strong Motion

Strong ground motion is complex, so it cannot be characterized with one or even a small number of parameters. Since it is recorded with an accelerometer, the most common parameter used to characterize a record is the peak acceleration. With baseline correction (e.g., Shakal *et al.*, 2003), that usually involves filtering to remove low-frequency noise, acceleration can be integrated to obtain velocity and displacement. Thus the peak velocity and, less often, peak displacement are used to characterize a record. Duration of strong shaking can be measured in two ways: one definition is the time during which most of the energy (e.g., 90%) arrives (e.g., Trifunac and Brady, 1975), and the other is the time between the first and last peak exceeding some threshold (e.g., 5%g, Bolt, 1969; Kramer, 1996). As distance from the source increases, amplitudes decrease and waves progressively disperse, so that the duration by the first definition increases while the duration by the second definition decreases to zero (no strong shaking if the selected threshold is not exceeded). Two classes of spectra are also important. Fourier spectra, first, are used to decompose the shaking into contributions from waves as a function of frequency and to make inferences about the source. The response spectra, second, measure the peak response (displacement, velocity, or acceleration) of the mass of an idealized, damped (usually at 5% of critical), single-degree-of-freedom oscillator when the base of the oscillator is subjected to the accelerogram (e.g., Hudson, 1979). Response spectra are formed by synthesizing the response of a suite of damped oscillators with different natural periods. The response spectrum is perhaps the most basic tool and the most used measure of strong motion for engineering applications (Jennings, 2003). It forms the basis for a design spectrum, which specifies the required strength of structures in building codes and for major projects.

4.18.2.2 Amplitude of Strong Motion

It is useful to consider the amplitudes of the largest strong ground motions that have been instrumentally recorded. **Figures 1** and **2**, from Anderson *et al.* (2005a), illustrate this.

Figure 1 shows the ranking of peak accelerations from the 70 largest acceleration records that are included in the COSMOS Virtual Data Center (VDC) for strong-motion accelerograms. The COSMOS VDC (Archuleta *et al.*, 2004) is an important resource allowing for the identification and retrieval of strong-motion data fitting a large number of criteria from most countries of the world. **Figure 1**, which may be incomplete, shows that there are about 25 records with peak accelerations greater than $1g = 980\,\mathrm{cm\,s^{-2}}$. These represent perhaps 0.1% of the set of records that are identified as 'strong-motion' records as defined by the COSMOS VDC (earthquakes with magnitude $M \geq 5$ in seismically active areas, $M \geq 4.5$ in less-active areas, or peak acceleration $\geq 0.1\%g$). Only records from 'free-field' or small structure conditions are included in **Figure 1**. Records from large structures may be deamplified if recorded in the basement, due to the inertial mass of the structure, or amplified by the structure if recorded on the roof, due to the flexibility of the structure as it responds to the ground shaking.

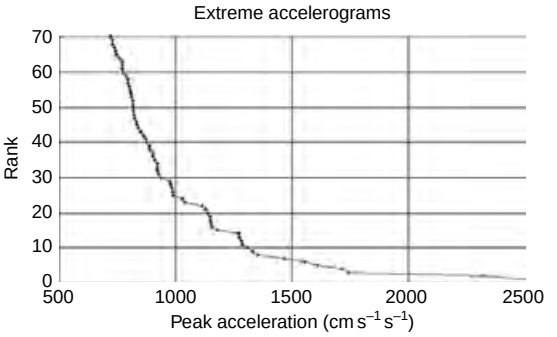


Figure 1 Extreme accelerations recorded in the 'free-field' or in small structures. The table with extreme records has been primarily drawn from the COSMOS Virtual Data Center. Note that these points are all peaks on individual components of strong-motion records. The peak vector acceleration is not necessarily recorded, and statistically its median is 25% greater than the largest horizontal component of a randomly oriented accelerograph. Expanded from Anderson JG, Purvance M, Brune JN, and Anooshehpoor A (2005a) Developing Constraints on Extreme Ground Motions Based on Seismic and Geological Observations (extended abstract), 2nd International Workshop, Strong Ground Motion Prediction and Earthquake Tectonics in Urban Areas, 25–27 October, 2005, Earthquake Research Institute, University of Tokyo, Japan, 63–66.

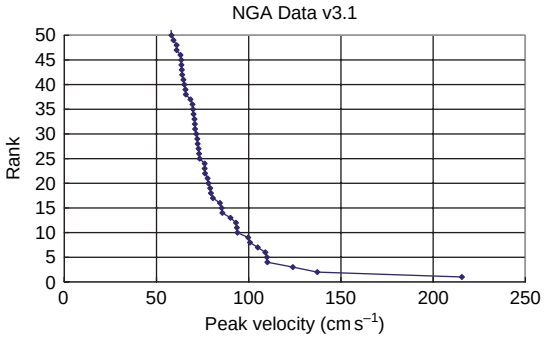


Figure 2 Extreme velocities recorded in free-field conditions. This set of observations with extreme velocity records has been drawn from the PEER NGA database. The peak velocities in this database are the geometric mean of the two horizontal components. On average, these are about 75% of the peak vector velocity. From Anderson JG, Purvance M, Brune JN, and Anooshehpoor A (2005a) Developing Constraints on Extreme Ground Motions Based on Seismic and Geological Observations (extended abstract), 2nd International Workshop, Strong Ground Motion Prediction and Earthquake Tectonics in Urban Areas, 25–27 October, 2005, Earthquake Research Institute, University of Tokyo, Japan, 63–66.

Anderson *et al.* (2005a) report that about two-thirds of the records on **Figure 1** come from thrust faulting earthquakes, and that about half of the records were on the hanging wall of a thrust fault. There was no obvious tendency for the extreme peak accelerations to occur more frequently with higher magnitude events. Most points on **Figure 1** are from earthquakes in the magnitude 6.5–7.0 range. Hanks and Johnson (1976) suggested that extremes in peak acceleration should be independent of magnitude at 10 km distance from the fault, where the distance is generalizable to any distance close to the fault. Their hypothesis is that this should be true if the dynamic shear stresses associated with localized failure on the fault are independent of magnitude. In three earthquakes with $M \geq 7.4$ from 1999 to 2002, mean values of peak accelerations were substantially smaller than expected based on ground motion prediction equations (GMPEs) (which are discussed later) that model peak acceleration as a function of magnitude, distance, site conditions, and geometrical factors (e.g., Anderson *et al.*, 2000; Boore, 2001, Campbell and Bozorgnia, 2003; Frankel, 2004). Although the sample of such large earthquakes is still small, these observations have generated speculation that such low accelerations are a characteristic of the largest earthquakes, at least on well-developed faults.

Figure 2, also from Anderson *et al.* (2005a), shows the ranking of extreme recorded velocities. Because many accelerograms have not been integrated to velocity, **Figure 2** draws upon a much smaller database, namely the Next Generation Attenuation (NGA) database prepared by the Pacific Earthquake Engineering Research (PEER) center. This database includes only about 2000 accelerograms which have been integrated to obtain velocity in a uniform, well-documented manner. **Figure 2** shows nine records for which the geometric mean of the two horizontal components exceeded 100 cm s^{-1} and 35 records in excess of 70 cm s^{-1}. A preliminary conclusion, suggested by Anderson *et al.* (2005a), is that most of the extreme velocities are associated with near-fault pulses for earthquakes with $M > 6$, as will be discussed later in this article. The largest peak velocities do come from the largest earthquakes in the data, and Anderson *et al.* (2005a) suggest that there may be a magnitude-dependent upper bound for peak velocity recorded at the surface of the Earth.

The largest amplitudes of acceleration and velocity observed up to this time will undoubtedly be exceeded on occasions by future events. Is there evidence for substantially more severe motions? Much higher accelerations at very high frequencies have been observed in mines: McGarr *et al.* (1981) observed 12*g* at a distance of 71 m from an earthquake with $M_L = 0.72$ in a deep mine in South Africa. The frequencies associated with this acceleration were between 20 and 400 Hz, and so much higher than the frequency range of 1–10 Hz that is more typical of strong motions in earthquake.

There are occasional reports of objects thrown into the air by earthquakes, requiring peak accelerations greater than 1*g*; in such cases the distance thrown sets a lower bound on initial velocity. A famous example is documented by Oldham (1899). Oldham reported displaced boulders or other stone objects in the 1897 earthquake in Assam, India ($M \sim 8.5$), had been thrown from their original location by distances of up to 2.5 m. Other examples, including eight cases where objects have been thrown from 2 to 4 m, are summarized by Midorikawa (1994). Ohmachi and Midorikawa (1992) suggest that observations of thrown objects in the 1984 Western Nagano Prefecture, Japan, earthquake (up to 3.0 m) do not require ground motions substantially greater than the amplitudes represented on **Figures 1** and **2**.

Records with extreme accelerations are not necessarily the same as the records with extreme velocities. Considering **Figures 1** and **2**, we consider that accelerations greater than 1*g*, and velocities greater than 100 cm s^{-1} are exceptionally strong levels of ground shaking.

4.18.3 Representation Theorem

Earthquakes are caused when the rocks on opposite sides of a fault slip suddenly. The characteristics of strong ground motions that result from this instability are strongly affected by the geometrical and dynamic characteristics of the faulting. The geometrical characteristics include the size, shape, depth, and orientation of the fault area that slipped during the earthquake and the amount, direction, and complexity of slip. The orientation of the fault and the amount and direction of offset may be variable over the fault surface. The important dynamic parameters include where on the fault the rupture initiated (hypocenter), how rapidly it spread over the fault (rupture velocity), how quickly (rise time) and smoothly the slip took place at each point on the fault, and how coherently adjacent points on the fault moved. These dynamic parameters are functions of the stresses acting on the fault, the physical properties of the rock surrounding the fault, and the strength of the fault itself, and vary over the surface of a large fault during major earthquakes.

The resulting ground motion is generally expressed mathematically using a representation theorem. The origin and applications of this approach are discussed thoroughly in Chapter 4.02. Using the notation of Aki and Richards (1980), one form for the representation theorem is

$$u_n(\mathbf{x}, t) = \int_{-\infty}^{\infty} \mathrm{d}\tau \int_{\Sigma}\int [u_i(\boldsymbol{\xi}, \tau)] \times c_{ijpq} v_j \frac{\partial G_{np}(\mathbf{x}, t-\tau; \boldsymbol{\xi}, 0)}{\partial \xi_q} \mathrm{d}\Sigma \qquad [1]$$

In this equation, $u_n(\mathbf{x}, t)$ gives the nth component of the displacement of the ground at an arbitrary location $\mathbf{x}$ and at time t. The vector $\mathbf{v}$ is normal to the fault, and the positive direction of the normal defines the positive side of the fault for defining the slip discontinuity. The ith component of the discontinuity in the slip across the fault is given by $[u_i(\boldsymbol{\xi}, \tau)] = u_i^+(\boldsymbol{\xi}, \tau) - u_i^-(\boldsymbol{\xi}, \tau)$, where $\boldsymbol{\xi}$ represents a location on the fault surface Σ and τ is the time that this discontinuity occurs. The Green's function is given by $G_{np}(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$. This gives the motion in the n direction at location $\mathbf{x}$ and time t caused by an impulsive point force acting in the p direction at location $\boldsymbol{\xi}$ and time τ. Finally, c_{ijpq} gives the elastic constants of the medium. For an isotropic medium, $c_{ijpq} = \lambda\delta_{ij}\delta_{pq} + \mu(\delta_{ip}\delta_{jq} + \delta_{iq}\delta_{jp})$, where λ and μ are the Lamé constants and δ_{ij} is the Kronecker delta (equal to 1 when $i = j$ and zero otherwise). The Lamé parameter μ is the shear modulus, the ratio of an increment in shear stress to the corresponding increment in shear strain. The bulk modulus, k, gives the ratio of an increment in hydrostatic pressure to the subsequent increment in volumetric strain. The Lamé parameter λ is related to k and μ as $\lambda = k - (2/3)\mu$. By convention, in eqn [1] summation takes place over repeated indices.

Equation [1] represents the ground motion at the site as the linear combination, through the integral over space, of the contributions from each point on the fault surface. The convolution over time incorporates the effect of the rupture at each point taking a finite amount of time to reach its final value. Through the representation theorem, the problem of predicting ground motions requires specification of the offset on the fault as a function of location and time, which incorporates earthquake source physics, and a specification of the Green's function, which incorporates wave propagation.

Equation [1] has certain limitations. For large amplitudes of seismic waves, the stress–strain relationship in the Earth, especially near the surface, becomes nonlinear. In this case, the assumed linear superposition of waves from different parts of the fault does not apply. A common approximation for this case is to predict the motion in 'rock' at some depth beneath the surface, where linearity is assumed to apply, and then treat the wave propagation from that depth to the surface as a vertical propagation problem through the nonlinear medium. On the earthquake source side, writing the time dependence in the Green's function as $(t - \tau)$ assumes that the Green's function is independent of time. This assumption can break down if the faulting process affects the propagation of the seismic waves from the source to the station. For instance, if seismic shear waves are not transmitted through the fault where it is slipping, but they are transmitted through the fault where it is not slipping, then eqn [1] would not hold exactly. In spite of these limitations, eqn [1] forms the mathematical basis for nearly all model-based ground motion predictions and has enabled the inversion of strong ground motion records to obtain models of the slip function for numerous earthquakes.

Isochrone theory is useful for understanding the relationship of ground motions in the near-source region (where the Green's function is not well approximated by a plane wave leaving the source). The basic theory is given in Bernard and Madariaga (1984) and Spudich and Frazer (1984, 1987). An

isochrone is the locus of points on a fault from which the seismic waves all arrive at a selected station at the same time. Thus, in an earthquake, isochrones are different for every station. The isochrone velocity is the velocity at which isochrones move over the surface of the fault. With the implicit assumption that the slip velocity at a point on the fault is discontinuous when the rupture starts and stops, Spudich and Fraser show that the high-frequency radiation reaching a station is large when the isochrone velocity for waves reaching that station is large or when the slip velocity of the dislocation on the fault is large.

The Fourier spectrum is defined as

$$a(\omega) = \int_{-\infty}^{\infty} a(t)\exp(-\mathrm{i}\omega t)\mathrm{d}t \qquad [2]$$

where $a(t)$ is a general time series. The real Fourier amplitude spectrum, $|a(\omega)|$, is the magnitude of the complex spectrum. In simplifying to the Fourier amplitude spectrum, the phase from the original signal, and the ability to recover the original signal through the inverse Fourier transform, is lost. The spectrum is usually plotted as a function of the natural frequency, f, rather than the angular frequency $\omega = 2\pi f$.

The Fourier spectrum of eqn [1] is

$$u_n(\mathbf{x}, \omega) = \int \int_{\Sigma} \left[u_i\left(\boldsymbol{\xi}, \omega\right)\right] \times c_{ijpq} v_j \frac{\partial G_{np}\left(\mathbf{x}, \omega; \boldsymbol{\xi}, 0\right)}{\partial \xi_q} \mathrm{d}\Sigma \qquad [3]$$

Equation [3] is simplified because convolution in the time domain transforms to a simple multiplication in the frequency domain. Assuming the origin of the spatial coordinates is someplace on the source, a common scalar representation of eqn [3] is

$$|u_n(r, f)| = S_{\mathrm{F}}(f) S_{\mathrm{GF}}(r, f) \qquad [4]$$

In eqn [4], $S_{\mathrm{F}}(f)$ gives the strength of the excitation of the seismic waves at the fault and $S_{\mathrm{GF}}(r, f)$ describes the effect of wave propagation on the signal once it leaves the source. Both $S_{\mathrm{F}}(f)$ and $S_{\mathrm{GF}}(r, f)$ are Fourier amplitude spectra. Equation [4] is most nearly valid when $|\boldsymbol{\xi}| \ll r$, where $r = |\mathbf{x}|$ is the distance from the origin to the observation point, so that $S_{\mathrm{F}}(f)$ represents the sum of all the energy leaving the fault. Equation [4], and extensions of eqn [4] in which physical processes that affect the Green's function are isolated, provides a convenient and widely used way to understand the physical phenomena that influence the observed ground motion.

Based on eqns [1] and [4], strong motion can be understood through an understanding of the contribution of the wave propagation from a finite fault to the site of interest (the Green's function) and the contribution from the source, that is, the slip on the fault. These are considered in further sections.

4.18.4 Green's Functions

Equation [1] shows that the Green's function and the source model are necessary to understand strong motion. In general the wave propagation is through rather complex media, in which direct waves, reflected and multipathed waves, and surface waves can be important depending on the situation. The important wave propagation phenomena will be discussed in the next section. This section focuses narrowly on approaches to obtain the Green's function, while the physics that affect the Green's function is left to the next section.

4.18.4.1 Empirical Green's Functions

As pointed out by Hartzell (1978), seismograms from a small earthquake solve exactly the wave propagation between the source and the station. Such a seismogram is not strictly a Green's function, since the focal mechanism is included and the rise time of the small event may have some effect, but it is called an empirical Green's function (EGF). The relationship between the seismogram from a small earthquake and a Green's function is explained in the context of eqn [1]. We begin by visualizing that the fault to be modeled is discretized into K elements ($k = 1, 2, \ldots, K$) of equal area, designated as $\delta\Sigma_k$, as shown in **Figure 3**. A small earthquake in element k, with moment M_{0k}^{EGF}, is to be used as an EGF. While the actual fault dimension of this small earthquake might be estimated from its corner frequency, we define the elemental slip of this event as

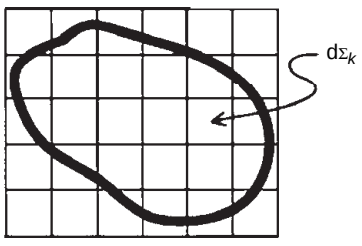


Figure 3 Sketch of a fault divided into subfaults.

$|\delta u_k| = \left(M_{0k}^{\mathrm{EGF}}/\mu\delta\Sigma\right)$. In the context of eqn [1], the slip discontinuity is a vector with components $[\delta u_{ki}]$ for $i = 1, 2, 3$. The magnitude of this vector is $|\delta u_k| = \left(\sum_{i=1}^{3} \delta u_{ki}\delta u_{ki}\right)^{1/2}$. The approximation used for EGF analysis assumes that the small event is a good approximation for a point source, at location $\boldsymbol{\xi}_k$ and at time τ_k, with an impulsive time function, so we substitute $[u_i(\boldsymbol{\xi}, \tau)] = [\delta u_{ki}]\delta(\tau - \tau_k)\delta(\xi - \xi_k)$ into eqn [1], to obtain

$$u_n^{\mathrm{EGF}-k}(\mathbf{x}, t) = [\delta u_{ki}]c_{ijpq}\nu_j \frac{\partial G_{np}(\mathbf{x}, t - \tau_k; \boldsymbol{\xi}_k, 0)}{\partial \xi_q}\delta\Sigma = u_n^{\mathrm{EGF}-k}(\mathbf{x}, t - \tau_k; \boldsymbol{\xi}_k) \quad [5]$$

To derive the proper way to use the EGF, one simply substitutes eqn [5] back into eqn [1]. The substitution is valid with certain assumptions. The first is that the EGF has the same focal mechanism as the earthquake to be modeled. Under this assumption, the normal vector $\mathbf{v}$ is the same and the direction of the dislocation vector $[\delta \mathbf{u}_k]$ is the same, so all the terms in the sums in eqn [1] and [5] contribute in the same proportions. The second is that the EGF has a corner frequency higher than any frequency of interest. A third, much safer, assumption is that the medium and hence the Green's function does not change with time so that τ_k can be replaced with a general time τ. Then

$$u_n(\mathbf{x}, t) = \int_{-\infty}^{\infty} \mathrm{d}\tau \sum_{k=1}^{K} \frac{\left[u_k\left(\boldsymbol{\xi}, \tau\right)\right]}{\delta u_k} \times u_n^{\mathrm{EGF}-k}\left(\mathbf{x}, t - \tau, \boldsymbol{\xi}\right) \quad [6]$$

The integration in eqn [1] has been approximated by a sum over all the fault elements. At this point it is convenient to switch the order of operations, and write the equation so that the convolution is carried out first. Equation [6] becomes

$$u_n(\mathbf{x}, t) = \sum_{k=1}^{K} \frac{\int_{-\infty}^{\infty} [u_k(\boldsymbol{\xi}, \tau)]\, u_n^{\mathrm{EGF}-k}(\mathbf{x}, t - \tau; \boldsymbol{\xi})\mathrm{d}\tau}{\delta u_k} \quad [7]$$

Multiplying both the numerator and denominator in eqn [7] by $\mu\ \delta\Sigma$, eqn [7] becomes

$$u_n(\mathbf{x}, t) = \sum_{k=1}^{K} \frac{\int_{-\infty}^{\infty} M_{0k}(\boldsymbol{\xi}, \tau)\, u_n^{EGF-k}(\mathbf{x}, t - \tau; \boldsymbol{\xi})\, \mathrm{d}\tau}{M_{0k}^{\mathrm{EGF}}} \quad [8]$$

where $M_{0k}(\boldsymbol{\xi}, \tau)$ is the total seismic moment that has been released in element k of the fault at time τ during the rupture of the large event. The convolution, in eqn [8], of the EGF with a function of finite duration for each element of the fault is an essential step, as the simple summation of EGFs cannot match large earthquake spectra over a broad bandwidth. This problem is discussed more fully in a later section.

4.18.4.1.1 Effect of signal-to-noise for EGFs

Figure 4 shows a series of accelerograms for a range of moment magnitudes from 2 to 8 and their corresponding Fourier amplitude spectra. The Fourier spectrum is defined in eqn [2]. At low frequencies, the Fourier amplitude spectrum of acceleration is much smaller at low magnitudes than at high magnitudes. In large-magnitude earthquakes, site and basin effects at low frequencies may become major factors in controlling the distribution of damage to significant structures. Thus predicting the low frequencies well is extremely important. However, Figure 4 shows that the spectral amplitude of a small earthquake at the low frequencies might drop below a useable level, controlled by instrumental or environmental noise, thus preventing that small event from providing useful information in a frequency band of critical interest. Thus the EGF needs to be selected carefully. Broadband seismographs of the types typically used for high-gain seismic stations may record lower frequencies with a useful signal-to-noise ratio, and thus could be installed at sites where EGFs will eventually be needed. Synthetic seismograms generated using eqn [8] are only valid at frequencies that are well above the noise level in the EGF.

4.18.4.2 Synthetic Green's Functions

Synthetic Green's functions can be computed for any source–station pair, for any focal mechanism, and for low frequencies. Thus the limitations imposed by the sparsity of ground motion recordings is eliminated. The disadvantages are that Earth structure is not known well enough to compute the synthetics precisely, and that as the frequency increases, the numerical demands quickly exceed the capabilities of present computers.

The analytical, closed form solution is available for motion in an infinite, homogeneous medium (e.g., Aki and Richards, 1980):

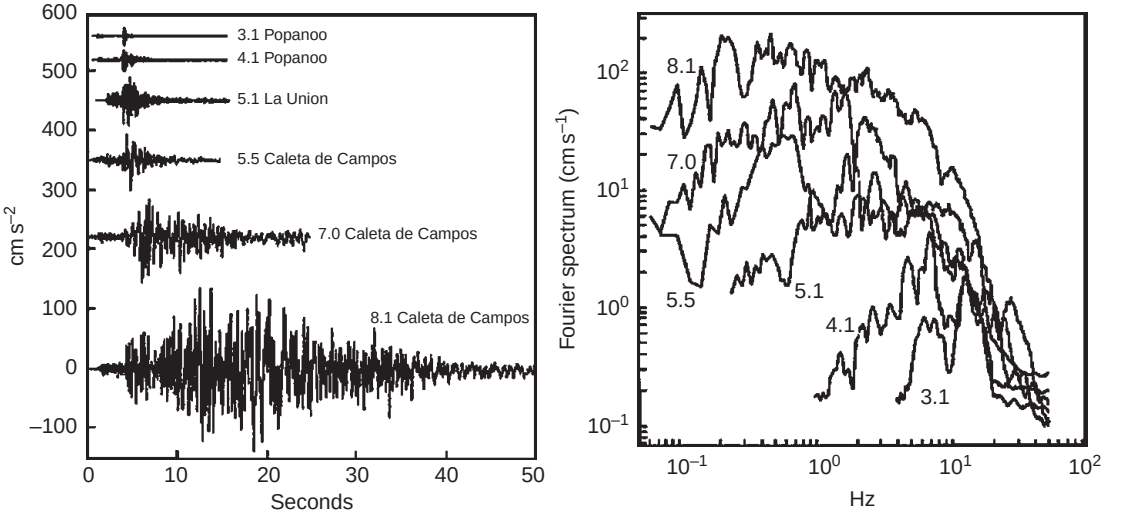


Figure 4 Accelerograms (*left*) and Fourier amplitude spectra, for a range of event sizes. The earthquakes were chosen to have a hypocentral distance of approximately 25 km, and to sample, as nearly as possible, earthquakes separated by one magnitude unit in size. From Anderson JG and Quaas R (1988) Effect of magnitude on the character of strong ground motion: An example from the Guerrero, Mexico, strong motion network. *Earthquake Spectra* 4: 635–646.

$$\mathbf{u}\vec{\mathbf{u}}(\mathbf{x},\,t) = \mathbf{u}^{\mathrm{N}}(\mathbf{x},\,t) + \mathbf{u}^{\mathrm{IP}}(\mathbf{x},\,t) + \mathbf{u}^{\mathrm{IS}}(\mathbf{x},\,t) + \mathbf{u}^{\mathrm{FP}}(\mathbf{x},\,t) + \mathbf{u}^{\mathrm{FS}}(\mathbf{x},\,t) \quad [9]$$

The individual terms each depend on the time function describing the seismic moment release at the source, $M_0(t)$. The near-field term is

$$\mathbf{u}^{\mathrm{N}}(\mathbf{x},\,t) = \frac{\mathbf{R}^{\mathrm{N}}(\theta,\,\varphi)}{4\pi\rho\, r^4}\int_{r/\alpha}^{r/\beta} \tau\, M_0(t-\tau)\mathrm{d}\tau \quad [10]$$

where $\vec{\mathbf{R}}^{\mathrm{N}}(\theta,\,\varphi) = 9\sin 2\theta\cos\varphi\ \hat{r} - 6(\cos 2\theta\cos\varphi\ \hat{\theta} - \cos\theta\sin\varphi\ \hat{\varphi})$. The intermediate-field P-wave term is

$$\vec{\mathbf{u}}^{\mathrm{IP}}(\mathbf{x},\,t) = \frac{\vec{\mathbf{R}}^{\mathrm{IP}}(\theta,\,\varphi)}{4\pi\rho\alpha^2 r^2} M_0\left(t-\frac{r}{\alpha}\right) \quad [11]$$

where $\vec{\mathbf{R}}^{\mathrm{IP}}(\theta,\,\varphi) = 4\sin 2\theta\cos\varphi\ \hat{r} - 2(\cos 2\theta\cos\varphi\ \hat{\theta} - \cos\theta\sin\varphi\ \hat{\varphi})$. The intermediate-field S-wave term is

$$\vec{\mathbf{u}}^{\mathrm{IS}}(\mathbf{x},\,t) = \frac{\vec{\mathbf{R}}^{\mathrm{IS}}(\theta,\,\varphi)}{4\pi\rho\beta^2 r^2} \dot{M}_0\left(t-\frac{r}{\beta}\right) \quad [12]$$

where $\vec{\mathbf{R}}^{\mathrm{IS}}(\theta,\,\varphi) = -3\sin 2\theta\cos\varphi\ \hat{r} + 3(\cos 2\theta\cos\varphi\ \hat{\theta} - \cos\theta\sin\varphi\ \hat{\varphi})$. The far-field P-wave term is

$$\vec{\mathbf{u}}^{\mathrm{FP}}(\mathbf{x},\,t) = \frac{\vec{\mathbf{R}}^{\mathrm{FP}}(\theta,\,\varphi)}{4\pi\rho\alpha^3 r} \dot{M}_0\left(t-\frac{r}{\alpha}\right) \quad [13]$$

where $\vec{\mathbf{R}}^{\mathrm{FP}} = \sin 2\theta\cos\varphi\ \hat{\mathbf{r}}$ and finally the far-field S-wave term is

$$\vec{\mathbf{u}}^{\mathrm{FS}}(\mathbf{x},\,t) = \frac{\vec{\mathbf{R}}^{\mathrm{FS}}(\theta,\,\varphi)}{4\pi\rho\beta^3 r} \dot{M}_0\left(t-\frac{r}{\beta}\right) \quad [14]$$

where $\vec{\mathbf{R}}^{\mathrm{FS}} = (\cos 2\theta\cos\varphi\ \hat{\theta} - \cos\theta\sin\varphi\ \hat{\varphi})$. In these equations, α, β, and ρ are the P-wave velocity, S-wave velocity, and density, respectively. The source is supposed to be at the origin, and in spherical coordinates, $\mathbf{x} = (r, \theta, \varphi)$ so r is the distance from the source to the station. The angle θ is the polar angle, measured from the normal to the fault. The angle φ is measured in the fault plane, with $\varphi = 0$ representing the direction of fault slip on the side of the fault where $\theta < 90°$.

The moment release on the fault, $M_0(t)$, is a monotonic function, reaching the static moment when slip on the fault ends. Thus the near-field and intermediate-field terms (eqns [10]–[12]), which are often referred to collectively as 'near-field terms' carry a static offset. The distance-dependence of these terms is r^{-2} or faster. Ground motions associated with the sum of these terms are sometimes referred to as the 'fling step' (e.g., Bolt and Abrahamson, 2003). Because $\alpha \approx \sqrt{3}\beta$, the S-wave term in eqn [12] is likely to dominate in the fling step. Based on the records that are now available, the verb 'lurch' would better describe these motions. The time derivative, $\dot{M}_0(t)$, returns to zero after the earthquake, so the far-field terms do not contribute to a static offset. The far-field terms (eqns [13] and [14]) have an amplitude dependence of r^{-1}, so at large

distances these terms dominate over the near-field and intermediate-field terms.

The Green's function has also been solved for a source in a flat-layered medium, where each layer is homogeneous and isotropic (e.g., Luco and Apsel (1983) present an approach that minimizes numerical instability). The solution in a layered medium requires numerical evaluation of certain integrals to obtain the Fourier transform of the Green's function (e.g., Apsel and Luco, 1983). These integrals are within the capability of desktop computers.

A realistic Earth model differs significantly from the flat-layered approximation. Most major cities in tectonically active areas are built on sedimentary basins. As will be discussed in the next section, basins have a first-order effect on the ground motions within the basin, and the flat-layered approximation significantly underestimates both the amplitude and the duration of the low-frequency ground motions. The flat-layered approach may be useful as an approximation for the high frequencies in the seismogram. However, for the low frequencies the preferred approach is to develop a three-dimensional (3-D) model for seismic velocities within and outside of the basin, and then to solve the equations of motion for an elastic medium directly by a finite-difference or finite-element numerical method (e.g., Frankel and Vidale, 1992; Graves, 1996; Bao *et al.*, 1998; Komatitsch and Tromp, 1999; Hartzell *et al.*, 1999; Hisada and Bielak, 2003; Furumura and Chen, 2005). Komatitsch *et al.* (2004) show a good example of how the finite-difference model solution is significantly better than a layer structure at low frequency.

4.18.5 Source Models

Strong motion cannot be understood without an understanding of the effect of a finite source. In a large earthquake, the fault is tens to hundreds of kilometers long. To evaluate eqn [1], a source model $[u_i(\boldsymbol{\xi}, \tau)]$ that describes this slip is necessary. Direct observations are impossible, since this slip takes place at depths from the surface to about 15 km or more for shallow crustal earthquakes. For that reason, much of the available information about the space-time behavior of the source comes from inverting eqn [1] for the slip as a function of time and location on the fault. Earthquake source physics is reviewed in Chapter 4.02, and inversions using eqn [1] are discussed in Chapter 4.07. Therefore, this discussion is brief. Before considering observed results, we describe the Haskell model that captures some of the lowest-frequency characteristics of the slip and sets up the terminology for later models.

4.18.5.1 Haskell's Model

Haskell (1964) proposed that a propagating ramp is a simple but appropriate model for slip as a function of time on a fault. The Haskell model is applied to a rectangular fault with length L in the x-direction (slip is zero except for $0 < x < \mathrm{L}$) and width W in the y-direction (slip is zero except for $0 < y < W$). On the part of the fault that slips, the time function is

$$[u(x, y, t)] = \begin{cases} 0 & t < \dfrac{x}{v_r} \\ \dfrac{D_0}{t_r}\left(t - \dfrac{x}{v_r}\right) & \dfrac{x}{v_r} \le t \le \dfrac{x}{v_r} + t_r \\ D_0 & \dfrac{x}{v_r} + t_r < t \end{cases} \quad [15]$$

There are five parameters in this model. Fault length (L) and width (W) are defined above. The fault slip is D_0. The kinematic part of the rupture is specified by the rupture velocity v_r and rise time t_r. The rupture starts at $x = 0$ and progresses across the fault at the rupture velocity. After the rupture front reaches a point x, the fault begins to slip. The slip increases linearly with time to the final offset D_0, and the rise time is the duration during which the slip velocity is nonzero. The slip velocity on the fault, $\dot{D} = D_0/t_r$, is controlled by the strength of the fault and by the effective stress that is acting across the fault, σ. In the Brune (1970, 1971) model, $\dot{D} = (\sigma/\mu)\,\beta$, where β is the velocity of shear waves in the medium. If $\sigma = 10\,\mathrm{MPa} = 100\,\mathrm{bars}$, $\mu = 3 \times 10^4\,\mathrm{MPa}$ and $\beta = 3000\,\mathrm{m\,s^{-1}}$, then $\dot{D} = 1\,\mathrm{m\,s^{-1}}$. Thus by this model, peak velocities of 1–2 $\mathrm{m\,s^{-1}}$ suggest that the effective shear stress is on the order of 10–20 MPa for shallow crustal faulting.

4.18.5.2 Inversions for Slip Models

Strong-motion data for many strong-motion records have now been inverted to estimate the slip function on the fault as a function of time, using techniques described in Chapter 4.07. As of October 2005, inversions were available for at least 56 earthquakes, in a database of slip functions compiled by Martin Mai. **Figure 5** shows the displacement and fault slip velocity for one model (Sekiguchi and Iwata, 2001) of the 1999 Chi Chi, Taiwan earthquake ($M_W = 7.6$). This

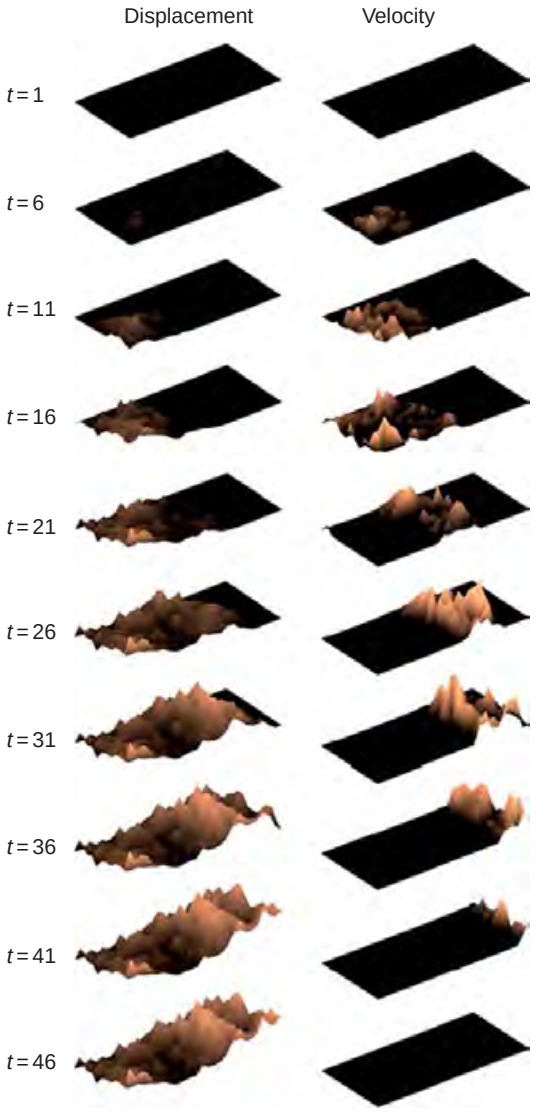


Figure 5 Slip model for the 16 September 1999 Chi Chi, Taiwan, earthquake ($M_W = 7.6$). This slip model is as determined by Sekiguchi and Iwata (2001) using the strong-motion data from this earthquake. Taiwan is among the best-instrumented countries in the world, and this earthquake is by far the best-recorded continental thrust earthquake of its size. The fault is 78 km long and 39 km wide. It is oriented approximately north–south, and in this view the long back edge is near the surface with the north end to the right. The fault dips 30° to the east, and in this view we are looking at the fault towards the northwest. The column on the left shows total slip on the fault, where the peak (achieved at $t = 31$ s and later) is 9.1 m. The column on the right shows the corresponding velocity. The peak velocity is 2.5 m s^{-1}, in the frame for $t = 31$ s.

is only one of several models for the Chi Chi earthquake (e.g., Chi *et al.*, 2001; Ma *et al.*, 2001; Wu *et al.*, 2001; Zeng and Chen, 2001). These models generally differ in detail, but the broad features are similar. The column on the left, showing displacement, shows that compared with the Haskell model, the final slip is variable over the fault surface. The sequence of inverted slips at different times shows the rupture starting at the hypocenter and spreading initially more or less circularly outwards. Later in the event, the slip more closely resembles the propagating ramp model. This feature shows up perhaps more clearly in the velocity column (right). The slip appears somewhat slow to get organized near the hypocenter, but later in the record it is spatially correlated and behaves more like a propagating ramp moving along the fault, as proposed by Haskell. However, **Figure 5** suggests that the velocity also is rather variable for different points within the slipping portion of the fault. Areas on the fault with high slip are called 'asperities'.

Heaton (1990) summarized several early inversions for fault slip. A key observation is that slip has a relatively short rise time everyplace on the fault, thus confirming that a propagating ramp is a reasonable first-order model.

4.18.5.3 Fourier Amplitude Spectrum of the Seismic Source

The Fourier spectrum of the P- and S-wave radiation from a point source can be obtained by taking a Fourier transform of eqns [13] and [14]. Since the moment rate function, $\dot{M}_0(t)$, is a monotonic function, its first derivative is a one-sided pulse. The Fourier transform of any one-sided pulse is flat at low frequencies, and then decreases with increasing frequency above some corner frequency. Using eqn [2], it is easy to show that a boxcar-shaped pulse $p(t) = P_0$ for $0 \le t \le T_P$ and $p(t) = 0$ otherwise has the Fourier amplitude spectrum $p(f) = P_0 T_P \sin(\pi f T_P)/(\pi f T_P)$. At low frequencies, this function is asymptotic to the area under the pulse. It has its first zero at $f_Z = 1/T_P$, and when plotted on logarithmic axes the high-frequency and low-frequency assymptotes intersect at a corner frequency $f_c = 1/(2T_P)$. For any general one-sided pulse, the corner frequency is inversely related to the pulse width. In accord with this general mathematical principle, the Fourier transforms of the far-field P- and S-wave pulses given by eqns [13] and [14] will have low-frequency assymptotes proportional to $\dot{M}_0$, and a corner frequency related to the size of the fault.

The model by Brune (1970) is used frequently as a starting point for a model of the source function. By

his complete stress drop model, the source term in eqn [4] is

$$S_F(f) = \frac{M_0}{\left(1 + (f/f_c)^2\right)} \quad [16]$$

where $f_c = (2.34\beta/2\pi a_R)$ and a_R is the radius of circular fault. The Brune stress drop is

$$\Delta\sigma_B = \frac{7}{16}\frac{M_0}{a_R^3} \quad [17]$$

For a constant stress drop, this equation predicts that $M_0 f_c^3 = \text{const}$. By eqn [16], the Fourier amplitude spectrum of the displacement will fall off as f^{-2} at frequencies higher than f_c. Brune (1970) also recognized the possibility of partial stress drop earthquakes, which by his model would have a less rapid falloff, proportional to f^{-1}, for some limited range of frequencies at $f > f_c$.

The shape for $S_F(f)$ is important for the development of stochastic synthetic seismograms, as described in Section 4.18.5.3 To predict the acceleration spectral amplitudes, eqn [4] is multiplied by $(2\pi f)^2$. That factor may be carried in the source term in the papers describing this approach (e.g., Boore, 1983; Atkinson, 1993b; Chen and Atkinson, 2002). Spectral shapes with multiple slopes are described by Brune (1970, 1976), Savage (1972), Joyner (1984), Atkinson (1993b), Atkinson and Silva (2000), Chen and Atkinson (2002), and Garcia *et al.* (2004) among many others.

4.18.5.4 Scaling of Source Parameters

Source models of future earthquakes are, of course, unknown. Thus successful prediction of plausible motions through the use of eqn [1] requires reasonable estimates of the source characteristics, motivating efforts to find general properties from the suite of available source inversions. Scaling relations are thus developed to describe the dependence of earthquake source parameters with moment magnitude. Source parameters of concern are fault length L, fault width W, average slip D, and stress drop $\Delta\sigma$. The seismic moment, M_0, is

$$M_0 = \mu LWD \quad [18]$$

and the moment magnitude, M_W, is

$$M_W = \frac{2}{3}(\log M_0 - \log M_0(0)) \quad [19]$$

The seismic moment of an earthquake with $M_W = 0$ is designated as $M_0(0)$ in eqn [19]. As pointed out by Utsu (2002), there is confusion in the literature over the size of $M_0(0)$. Kanamori (1977) introduced the notation M_W with $M_0(0) = 10^{16.100}$ dyne cm, while Hanks and Kanamori (1979) introduced the notation **M** with $M_0(0) = 10^{16.050}$ dyne cm. Anderson (2003) noted that one observatory practice uses $M_0(0) = 10^{16.095}$ dyne cm, but did not check others. A review of online catalogs shows that different catalogs use different formulas. For instance, the Harvard/Lamont CMT catalog uses $M_0(0) = 10^{16.100}$, while the USGS online catalog uses $M_0(0) = 10^{16.050}$. This makes it difficult for strong-motion seismologists to maintain consistency with the agencies that are determining seismic moment from teleseismic records. Caution is advised in working with moment magnitudes.

Based on surface rupture and the dimensions of aftershock zones, Wells and Coppersmith (1994) developed an extensive set of empirical relations between fault area, rupture length, slip, and magnitude. Based on previously recognized differences in earthquake scaling depending on repeat time and tectonic environment (Kanamori and Allen, 1986; Scholz *et al.*, 1986; Wesnousky, 1986), Anderson *et al.* (1996) found equations that used the slip rate in addition to fault length to estimate the magnitude. Hanks and Bakun (2002) suggest a bilinear relationship to predict magnitude from fault length in California. As in **Figure 5**, slip models invariably show very heterogeneous slip, so the definition of length, width, and mean slip can be problematic. Mai and Beroza (2000) defined fault length and width from the autocorrelation distance in the two directions, and derived a different set of relations using results of inversion of the ground motions for 18 earthquakes.

The static stress drop is

$$\Delta\sigma = b_{SD}\mu\frac{D}{L_c} \quad [20]$$

where L_c is the characteristic length of the smaller rupture dimension. Thus L_c would be the fault radius, a_R, for a circular fault or the fault width, W, if $W < L$, for a rectangular fault. The constant b_{SD} depends on the fault geometry, but is of order unity in all cases (Kanamori and Anderson, 1975). To be specific, $b_{SD} = 7\pi/16$ for a circular fault, $b_{SD} = 2/\pi$ for a long strike-slip fault, and $b_{SD} = 4(\lambda + \mu)/\pi(\lambda + 2\mu)$ for a long dip-slip fault.

A fault dimension can be estimated from the corner frequency, as described in the previous section, or pulse duration (e.g., Jones and Helmberger, 1998;

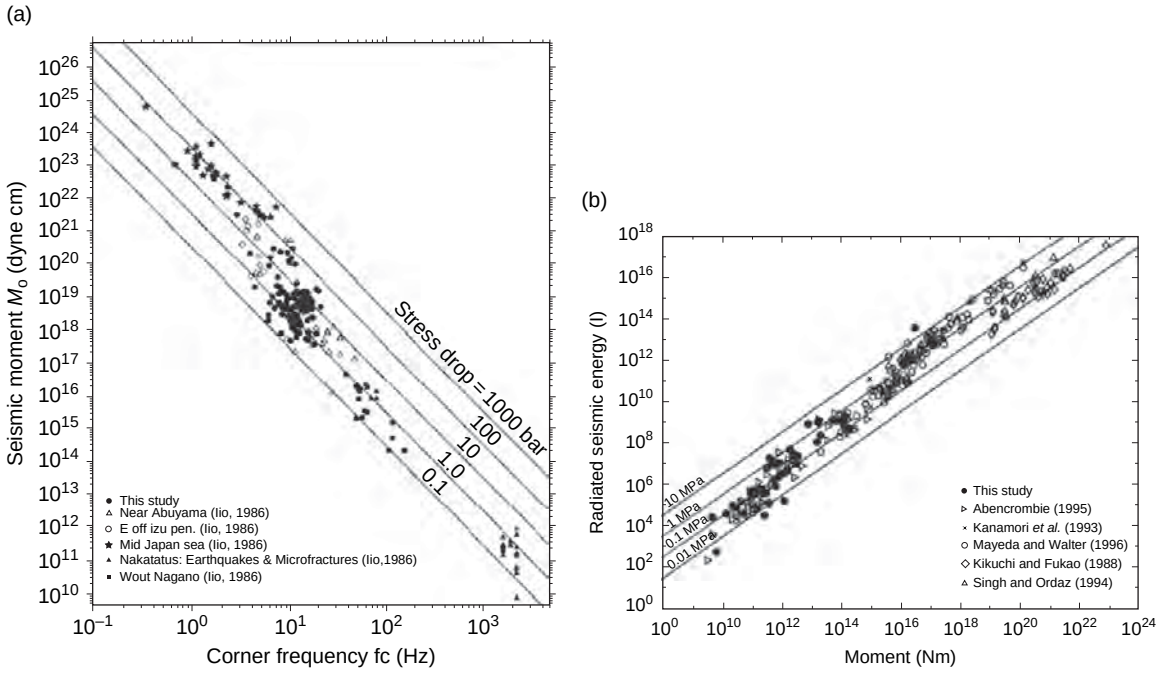


Figure 6 (a) Corner frequency and seismic moment found for a large range of seismic moments from earthquakes in Japan, by Jin *et al.* (2000). (b) Radiated energy and seismic moment for a large range of earthquakes compiled by Prejean and Ellsworth (2001).

Kanamori and Rivera, 2004), allowing the determination of the static stress drop by eqn [20] (**Figure 6(a)**). Since the fault width is not easily obtained from seismograms, this is only applied on the assumption of a circular fault, so the approach of Brune (1970), and eqn [17], is applied. As a broad generalization, measured stress drop tends to be in the range of 0.1–1000 bars (0.01–100 MPa) for all sizes of earthquakes. For small magnitudes, when this approach is applied, a falloff of the spectrum due to attenuation of the seismic waves can mask, and be confused with the corner frequency (e.g., Anderson, 1986). When this occurs, as the magnitude decreases the corner frequency tends to become more constant and the calculated stress drop decreases. Some studies suggest that there is a magnitude dependence in the stress drop (e.g., Hardebeck and Hauksson, 1997; Shi *et al.*, 1998; Jin *et al.*, 2000). Other recent studies have concluded that there is no magnitude dependence or that it is indeterminate (e.g., Atkinson, 1995, 2004; Jones and Helmberger, 1998; Bindi *et al.*, 2001; Margaris and Hatzidimitriou, 2002; Garcia *et al.*, 2004). Mai and Beroza (2000) found that $\Delta\sigma$ is about constant, and $L \propto W \propto D$ for large dip-slip faults, meaning that in the magnitude range that they studied ($M = 5.6$–8.1), these events are self-similar. On the other hand, for strike-slip faults, they suggested that $\Delta\sigma$ increased with magnitude in the range from $M = 5.6$ to $M = 7.3$.

A different stress parameter that is more related to the dynamics of faulting is the apparent stress, σ_A, defined by Wyss (1970) as

$$\sigma_A = \mu \frac{E_R}{M_0} \qquad [21]$$

In eqn [21], E_R is the radiated seismic energy. An equivalent parameter is $\tilde{\epsilon} = (E_R/M_0)$ (e.g., Kanamori and Rivera, 2004). Some recent papers that have estimated σ_A include Prejean and Ellsworth (2001), Hwang *et al.* (2001), Kinoshita and Ohike (2002), Beeler *et al.*, 2003; McGarr and Fletcher (2003), Stork and Ito (2004), and Takahashi *et al.* (2004). Like $\Delta\sigma_B$, estimates of the apparent stress σ_A can be biased towards low values at small magnitudes due to attenuation. Kanamori and Rivera (2004) show how $\Delta\sigma_B$ and σ_A are related to each other and to the rupture velocity on the fault.

4.18.6 Effect of Surface Geology

Much attention in strong-motion seismology is devoted to understanding the effects of near-surface geology on the ground motions. Kawase (2003) has recently also reviewed this topic. It is of course understood that in most places, the seismic compressional and shear velocity increase with depth. The 'near-surface' is probably best taken to mean the geology above 'seismic bedrock', where, somewhat arbitrarily, 'seismic bedrock' has shear velocity $\beta \geq 3.0\,\mathrm{km\,s^{-1}}$ (e.g., Takemura *et al.*, 1995). The main release of seismic energy is within the seismic bedrock.

In uniform rock, joints and microcracks that are open near the surface are closed by increasing pressure with depth. Velocity in laboratory rock samples increases by 10–30% as the pressure is increased from zero to 70–100 MPa (Jaeger and Cook, 1976; Christensen, 1989). Taking a density of $2800\,\mathrm{kg\,m^{-3}}$, the lithostatic pressure reaches 70–100 MPa at depths of about 2.5–3.5 km. Thus the influence of the surface extends to at least that depth. Surface outcrops of rock show site effects, and thus may not always be suitable for reference ground motions (e.g., Castro *et al.*, 1990; Humphrey and Anderson, 1992; Steidl *et al.*, 1996).

In unconsolidated sediments, compression due to the gravitational load of the overburden will cause both density and the elastic moduli to increase with depth. The shear velocity, $\beta = \sqrt{\mu/\rho}$ also increases because the shear modulus increases more rapidly than the density. Laboratory measurements typically find $\mu \propto (\sigma_\mathrm{m}')^{1/2}$, where σ_m' is the mean principal effective stress (e.g., Kramer, 1996). In the simple case where $\sigma_\mathrm{m}' \propto (\text{depth})$, the predicted depth dependence of the shear modulus is $\mu \propto (\text{depth})^{1/2}$, and since the density increases relatively slowly with depth, $\beta \propto (\text{depth})^{1/4}$.

Phenomena associated with propagation of seismic waves between seismic bedrock and the surface include at least all of the following effects: (1) amplification of waves due to the decrease in seismic velocities and density, (2) refraction of body waves towards vertical incidence, (3) reflection and transmission across interfaces, (4) resonance in flat-layered low-velocity layers at the surface, (5) propagation of surface waves, (6) conversion of body waves to surface waves at dipping boundaries, particularly on the edges of basins, (7) trapping and resonance of waves in basins, (8) long linear valleys acting as waveguides, (9) multipathing of body waves causing focusing and high-amplitude ground motions, and similar multipathing of surface waves, (10) scattering causing an increased duration of ground motions, (11) low coherence of seismic waves (especially higher-frequency body waves) also caused by scattering, (12) 'linear' absorbtion of energy approximated as low quality factor Q, (13) scattering by the surface topography, and (14) nonlinear stress–strain relationships in near-surface materials.

4.18.6.1 NEHRP Site Classifications

The United States program in earthquake hazards, called the National Earthquake Hazard Reduction Program (NEHRP) has developed a site classification scheme for the near-surface geology. It is based on the average shear velocity in the upper 30 m, V_{S30}. Setting h_i to be the thickness of a layer and β_i to the corresponding shear velocity, then in general the average shear velocity of n layers is

$$\bar{V}_\mathrm{S} = \sum_{i=1}^{n} h_{i-1} \Big/ \sum_{i=1}^{n} \frac{h_{i-1}}{\beta_{i-1}} \qquad [22]$$

where the numerator is the thickness of the stack of layers, and the denominator is the time for a direct wave to travel through the stack. In eqn [16] the layer starting at zero depth is designated with a zero subscript, thickness h_0 and shear velocity β_0. Then by definition $V_{S30} = \bar{V}_\mathrm{S}$ for the special case where $\sum_{i=1}^{n} h_{i-1}$ is 30 m. The significance of V_{S30} is that 30 m corresponds to 100 ft (actually 100 ft = 30.48 m), where 100 ft is a standard depth for geotechnical investigation in the United States, which still uses the English system of units for common measurements. In research developed under the United States NEHRP, the Building Seismic Safety Council (BSSC, 1997) grouped sites into six 'NEHRP site classes' by V_{S30}, and this classification was subsequently adopted as a standard terminology. This classification of near-surface geology is given in **Table 1**.

4.18.6.2 Amplification due to the Decrease in Shear Velocity

The site classes in **Table 1** are organized by average shear velocity in the upper 30 m. In part this distinction is based on a theoretical model for how the amplitudes of seismic waves change as a result of propagation into a different kind of material. Setting

Table 1 Definition of NEHRP site classes by velocity V_{100} (average shear velocity in the top 100 feet) V_{S30}, and blow count N_{100}

Site class	Velocity range	Blow count range
A: Hard rock	$5000\,\mathrm{ft\,s^{-1}} \le V_{100}$ $1500\,\mathrm{m\,s^{-1}} \le V_{S30}$	Not applicable
B: Rock	$2500 \le V_{100} < 5000\,\mathrm{ft\,s^{-1}}$ $760 \le V_{S30} < 1500\,\mathrm{m\,s^{-1}}$	$N_{100} \ge 100$
C: Very dense soil and soft rock	$1200 \le V_{100} < 2500\,\mathrm{ft\,s^{-1}}$ $360\,\mathrm{m\,s^{-1}} \le V_{S30} < 760\,\mathrm{m\,s^{-1}}$	$50 \le N_{100} < 100$
D: Stiff soil	$600 \le V_{100} < 1200\,\mathrm{ft\,s^{-1}}$ $180\,\mathrm{m\,s^{-1}} \le V_{S30} < 360\,\mathrm{m\,s^{-1}}$	$15 \le N_{100} < 50$
E: Soil	$V_{100} < 600\,\mathrm{ft\,s^{-1}}$ $V_{S30} < 180\,\mathrm{m\,s^{-1}}$	$N_{100} < 15$
F: Soils requiring site-specific studies	Includes liquefiable soils, Peat or clay with high organic content High plasticity clays Very thick soft/medium stiff clay	

After Building Seismic Safety Council (1997).

ρ_0 and β_0 to be the density and shear velocity at the surface, and setting ρ_S and β_S to be the density and shear velocity at the earthquake source, assuming density and velocity to vary smoothly, and neglecting the free-surface amplification, the predicted gain in amplitude at the surface is given by (e.g., Aki and Richards, 1980)

$$G_A = \sqrt{\rho_S \beta_S / \rho_0 \beta_0} \qquad [23]$$

Under the NEHRP recommendations for the national building code (Building Seismic Safety Council, 1997), the national zoning map aims to predict ground motion relative to a default site with $V_{S30} = 760\,\mathrm{m\,s^{-1}}$, that is, on the boundary between NEHRP class B and C. The amplification factors, $\boldsymbol{F_a}$ for short-period amplitudes (peak acceleration) and $\boldsymbol{F_v}$ for response spectral amplitude at 1 s period (which is strongly correlated with peak velocity), depend on the predicted response spectral amplitudes $\boldsymbol{S_s}$ and $\boldsymbol{S_1}$ at those periods, respectively. The amplifications are based at least partly on empirical evidence, such as that developed by Borcherdt and Glassmoyer (1992, 1994). **Tables 2** and **3** are included here to show how the code instructs design spectral levels to be modified for different soil classes. **Table 2** applies at high frequencies, and **Table 3** applies at frequencies centered near 1.0 Hz.

Values of $\boldsymbol{F_a}$ and $\boldsymbol{F_v}$ are plotted in **Figure 7**. **Figure 7** also shows the predictions using eqn [23]. A prominent and interesting feature on **Figure 7** is the presumption of strongly nonlinear soil response for the softer soil classes (Classes D, E). Essentially, nonlinear response is caused by waves that are too strong to be transmitted through the soil, so they lose amplitude rapidly and permanently deform the soil. Rock and the stiffest soils (Classes A, B, and C) are expected to remain linear or very nearly linear. Nonlinear stress–strain relations in soil will be discussed below (see Section 4.18.6.2).

The predictions based on eqn [23] are plotted at the zero-amplitude point on the abscissa because

Table 2 Values of F_a as a function of site class and mapped short-period maximum considered earthquake spectral acceleration

	Mapped maximum considered earthquake spectral response acceleration at short periods				
Site class	$\boldsymbol{S_s \le 0.25}$	$\boldsymbol{S_s = 0.50}$	$\boldsymbol{S_s = 0.75}$	$\boldsymbol{S_s = 1.00}$	$\boldsymbol{S_s \ge 1.25}$
A	0.8	0.8	0.8	0.8	0.8
B	1.0	1.0	1.0	1.0	1.0
C	1.2	1.2	1.1	1.0	1.0
D	1.6	1.4	1.2	1.1	1.0
E	2.5	1.7	1.2	0.9	a
F	a	a	a	a	a

Note: Use straight line interpolation for intermediate values of S_s.
a, Site-specific geotechnical investigation and dynamic site response analyses shall be performed.
After Building Seismic Safety Council (1997).

Table 3 Values of F_v as a function of site class and mapped 1-second period maximum considered earthquake spectral acceleration

	Mapped maximum considered earthquake spectral response acceleration at 1-s periods				
Site class	$S_I \leq 0.1$	$S_I = 0.20$	$S_I = 0.35$	$S_I = 0.4$	$S_I \geq 0.5$
A	0.8	0.8	0.8	0.8	0.8
B	1.0	1.0	1.0	1.0	1.0
C	1.7	1.6	1.5	1.4	1.3
D	2.4	2.0	1.8	1.6	1.5
E	3.5	3.2	2.8	2.4	a
F	a	a	a	a	a

Note: Use straight line interpolation for intermediate values of S_I.
a, Site-specific geotechnical investigation and dynamic site response analyses shall be performed.
After Building Seismic Safety Council (1997).

they are based on the assumption of linear wave propagation. The relative amplitudes of S_1 in the linear range, $S_1 \leq 0.1$, are approximately predicted by eqn [23]. However, the relative amplitudes of S_s in the linear range, $S_s \leq 0.25$, are smaller than predicted for site classes C, D, and E. A reasonable explanation for this is that for the high frequencies, attenuation due to low values of Q in the shallow layers could be responsible for the smaller amplifications (see Section 4.18.6.2). On the other hand, at the intermediate frequencies (1 Hz), attenuation losses will be relatively much smaller, so it is expected that the empirical amplifications will better match theory.

4.18.6.3 Refraction of Body Waves Toward Vertical Incidence

The direction of propagation of seismic waves changes as the wave velocity changes. Snell's law is the fundamental expression that predicts how this will happen. The nature of the change in ray

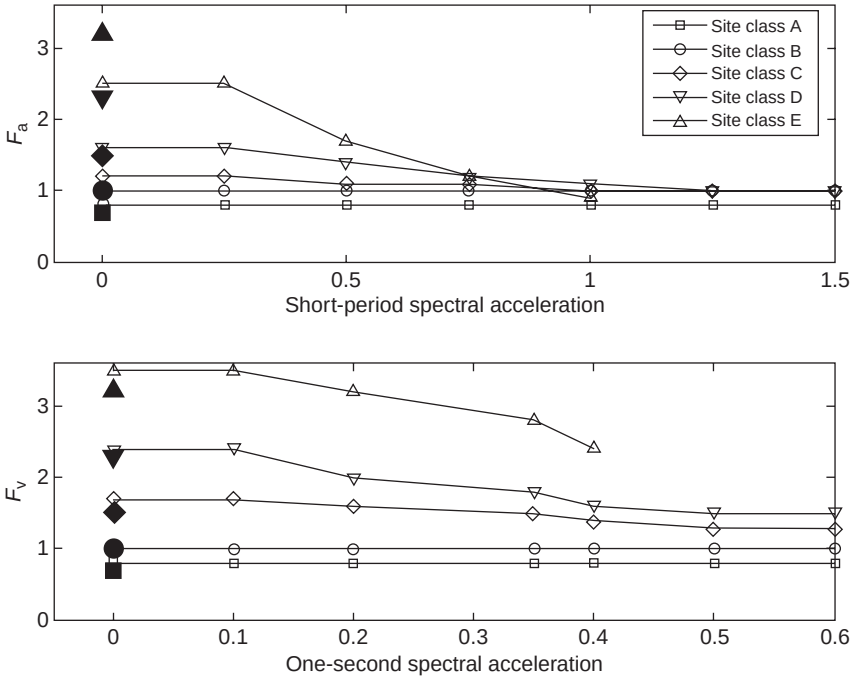


Figure 7 Values of F_a from **Table 2**, values of F_v from **Table 3**. Solid symbols show the predictions, based on eqn [23], of the relative amplifications of surface ground motions relative to amplifications of site class B. The prediction uses the average of the upper- and lower-velocity thresholds for each site class range. For site class A, the prediction uses $V_{S30} = 2250\,\mathrm{m\,s^{-1}}$, and for site class E, it uses $V_{S30} = 140\,\mathrm{m\,s^{-1}}$.

direction on crossing an interface is illustrated in **Figure 8**. The general statement of Snell's law is

$$\frac{\sin j_1}{\beta_1} = \frac{\sin j_2}{\beta_2} \qquad [24]$$

where β_1 is the wave velocity on one side of the interface and β_2 is the wave velocity on the other side of the interface. Angles j_1 and j_2 are measured from the normal to the interface (light vertical line in **Figure 8**).

Figure 9 shows the path that a ray with a large incidence angle will take through a stack of layers with successive properties from engineering bedrock through the site classes A through E. One feature is that in site classes C, D, and E, the energy is turned so that it is propagating very nearly vertical. Because of this feature, it is common in engineering practice to study the effect of surface layers on vertically propagating shear waves. Based on eqns [13] and [14], and confirmed by observations, the S wave is expected to have a larger amplitude than the P wave. Since the S-wave particle motion is perpendicular to the direction of travel, a consequence of this near-vertical propagation is that the motion in the S-wave is expected to be predominantly in the horizontal direction on soft sites. A rule of thumb for strong-motion accelerograms is that the peak vertical acceleration is two-thirds the amplitude of the peak horizontal acceleration. In reality, there is a distribution of values, and while it is true that the median ratio is about 2/3, values in the range from 0.5 to 1.0 are all probable. This ratio is perhaps higher than one might expect on soft soils, but the complete seismic wavefield is much more complex, as will be discussed subsequently.

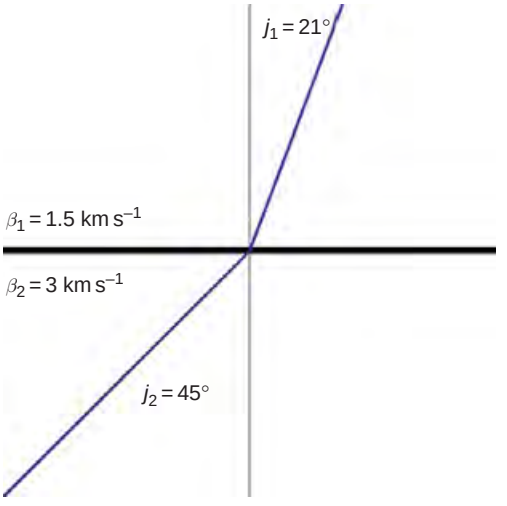


Figure 8 Illustration of change in direction of a seismic ray on crossing an interface (heavy line).

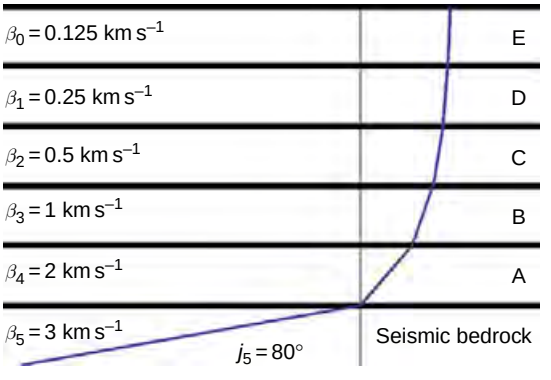


Figure 9 Illustration of the path that an S wave, starting with near-horizontal incidence in seismic bedrock, would take through different site classes to the surface. Site class (**Table 1**) is given in the column on the right. Shear velocities are taken to be representative of each site class.

4.18.6.4 Reflection and Transmission across Interfaces

The physics of what happens at an interface is more complicated than what is shown in **Figures 8** and **9**. For a plane SH wave incident at the interface, besides the path of the upgoing ray, there is a downgoing reflected ray, as indicated in **Figure 10**. For an SH wave, assuming Snell's law allows a transmitted wave, the transmission coefficient (which is the ratio of the amplitude of the transmitted wave, A_T, to the amplitude of the incoming wave, A_0) is

$$\frac{A_T}{A_0} = \frac{2\rho_2\beta_2}{(\rho_2\beta_2 + \rho_1\beta_1)} \qquad [25]$$

and the reflection coefficient (which is the ratio of the amplitude of the reflected wave, A_R, to the amplitude of the incoming wave) is

$$\frac{A_R}{A_0} = \frac{\rho_2\beta_2 - \rho_1\beta_1}{(\rho_2\beta_2 + \rho_1\beta_1)} \qquad [26]$$

The product $\rho\beta$ is called the impedance. In the limit of $\beta_2 >> \beta_1$, the transmission coefficient approaches doubling the amplitude of the incoming wave. The transmitted energy is not doubled, but rather decreased, because the low-velocity material has a much lower impedance. On the other hand, for transmission from a softer to a harder material, when $\beta_1 >> \beta_2$, then the transmission coefficient

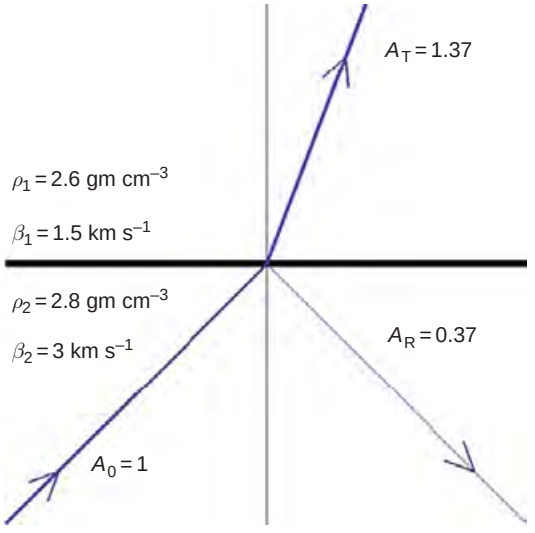


Figure 10 Transmitted and reflected SH wave. Line weights are proportional to the relative amplitudes of the waves.

approaches zero, meaning that the energy is almost totally reflected and is trapped in the low-velocity material. When an SH wave is reflected from a free surface, the 'free surface amplification' is a factor of two. All of the energy is reflected, and the wave is described as a sum of the incoming and reflected wave, each with the amplitude of the incoming wave. Another phenomenon at an interface is the conversion of P-wave energy to SV-wave energy and conversion of SV-wave energy to P-wave energy. The statement of Snell's law for P and SV waves is

$$\frac{\sin i_1}{\alpha_1} = \frac{\sin i_2}{\alpha_2} = \frac{\sin j_1}{\beta_1} = \frac{\sin j_2}{\beta_2} \qquad [27]$$

where α represents a P-wave velocity and angles i_1 and i_2 represent the directions of the incident and transmitted P waves, measured the same way as for S waves. Reflection and transmission coefficients are considerably more complicated in this case, and can be found by referring to a standard seismology textbook. They depend on the angle of incidence in addition to the contrast in velocity and density. The features mentioned above associated with SH-wave transmission carry over to this case: amplitudes are doubled for transmission from a hard to much softer material, and energy is almost totally reflected back into soft material when it encounters a much harder material.

4.18.6.5 Resonance in Flat-Layered Low-Velocity Layers at the Surface

A low-velocity layer near the surface traps seismic energy, and resonance occurs for some frequencies. The phenomenon is usually examined in the context of vertically propagating S waves, which are a good approximation as seen in **Figure 9**. On entering the low-velocity layer, the S-wave amplitude is increased as given by eqn [25]. The wave is totally reflected at the surface. On encountering the bottom of the low-velocity layer, it is mainly reflected as in this case eqn [26] applies for the wave approaching the boundary from the low-velocity side. The negative reflection coefficient indicates a 180° phase shift on this reflection.

If the incoming wave is a single-frequency wave, then for certain frequencies the second peak of the incoming wave is transmitted through the boundary at just the right time to be exactly in phase with the first peak after it has reflected from the surface and the layer boundary. Because of the phase shift in the first peak on the reflection at the base of the layer, the resonance occurs when the wavelength of the wave in the low-velocity layer is four times the layer thickness, that is, when $h_0 = \lambda/4$, where h_0 is the thickness of the top layer and λ is the wavelength (a different meaning from what is represented in eqn [1]). Noting that $\beta = f\lambda$ for a monochromatic shear wave of velocity β and frequency f, the resonant frequency of the layer is

$$f = \frac{\beta}{4h_0} \qquad [28]$$

In this model, resonance occurs at a sequence of frequencies, given by

$$f = \frac{(2k-1)\beta}{4h_0}, \quad \text{for } k = 1, 2, \ldots \qquad [29]$$

However, because realistic geology generally differs from the mathematical idealization of a perfectly homogeneous, flat-layered medium, and because of energy losses, most often the lowest frequency is the only conspicuous resonance in the Fourier amplitude spectrum of a seismogram at the surface.

The response of multiple layers to an incoming plane wave was first solved by Haskell (1953, 1960, 1962). The solution in a linear medium is a fairly standard seismological computer problem (e.g., Aki and Richards, 1980, 2002). The computer program SHAKE (Schnabel *et al.*, 1972) is widely used in engineering practice for the response of nonlinear

soils to vertically incident S waves. At each layer boundary, high-frequency energy is both transmitted forwards and reflected backwards. Waves with oblique incidence can be converted from P waves to SV waves or from SV waves to P waves. Thus, the energy can take many paths to the observer at the surface, resulting in multiple resonance peaks, prolonged duration, and preferential loss of high-frequency energy.

Examples of 1-D interpretations of site response include Hough *et al.* (1992), Steidl *et al.* (1996), Hartzell *et al.* (2000), Satoh *et al.* (2001), and Gil-Zepeda *et al.* (2002).

4.18.6.6 Trapping Energy in the Form of Surface Waves

Figure 9 should also be considered from the viewpoint of energy in shallow layers which is propagating downwards. As depth and shear velocity increase, the ray is bent more toward the horizontal, and if the velocity of a layer is too great for the ray to enter the layer while obeying Snell's law, there is total reflection. Thus seismic energy can be trapped in shallow, relatively low-velocity layers. As the distance increases from the earthquake, this energy organizes into surface waves, which have several features that distinguish themselves from body waves. The theory of surface waves is developed in theoretical seismology textbooks (e.g., Aki and Richards, 1980, 2002).

A few characteristics of surface waves are particularly relevant for strong-motion seismology and are thus worth noting. Because the energy is trapped near the 2-D surface of the Earth and propagating in directions parallel to the surface of the Earth, the geometrical spreading for surface waves is much less than for body waves that spread over three dimensions. Typically geometrical spreading is described as proportional to $r^{-0.5}$, in contrast to the body waves for which geometrical spreading is proportional to r^{-1} (eqns [10] and [11]). Thus as distance from the fault increases, surface waves are increasingly dominant in the seismogram. Surface waves that are organized from trapped SH waves are called Love waves. P and SV waves, coupled, by conversion from P to S or S to P waves at the boundaries, form Rayleigh waves. The particle motion of a Love wave is horizontal and transverse to the direction of wave propagation. The particle motion for a Rayleigh wave is elliptical and in a vertical plane parallel to the direction of wave propagation. The Love wave motion differs significantly from a vertically propagating S wave. For the vertically propagating planar S wave, the motion at the surface is the same everyplace, while a map of Love-wave particle motion shows periodic reversals every half wavelength in the direction of propagation.

Surface-wave propagation requires the right combination of frequency and angle of incidence leading to constructive interference of the contributing body waves. The frequency of this constructive interference depends on the incident angle, and thus surface waves of different frequencies travel with different phase and group velocities. This phenomenon is called surface-wave dispersion. Often, low-frequency surface waves travel significantly faster than higher-frequency surface waves. Because of the dispersion and low geometrical spreading, surface waves can contribute to damage through their relatively long durations and large amplitudes of ground motions at substantial distances from the source. The dispersion characteristics of surface waves are related to the velocity structure. Because they dominate the background seismic noise, the measurement and interpretation of surface-wave dispersion is an important tool for estimating the structure of urban basins and strong-motion accelerograph sites (e.g., Nazarian and Stokoe, 1984; Horike, 1985; Matsushima and Okada, 1990; Sato *et al.*, 1991; Malagnini *et al.*, 1997; Kawase *et al.*, 1998; Louie, 2001).

4.18.6.7 Effect of Basin Edges on Incoming Body Waves

Seismic waves incident on a sloping basin boundary will obey Snell's law at the sloping interface. This can result in a wave with a small incidence angle in the rock below the basin being converted to waves with a large incidence angle in the basin (e.g., Frankel, 1993). This is illustrated in **Figure 11**. The first encounter with the sloping basin boundary turns the ray towards the horizontal, and after the surface reflection the second encounter causes it to travel even much closer to horizontal. In this manner, body waves incident on a basin are converted to surface waves at the basin boundary. The impact on strong motion is that the duration of shaking in the basin is significantly increased, and the amplitude of shaking is also increased (e.g., Frankel *et al.*, 1991; Frankel and Vidale, 1992; Frankel, 1993, 1994). Indeed, Frankel *et al.* (2001) find that the strongest motions at some low frequencies (0.125–0.5 Hz) in

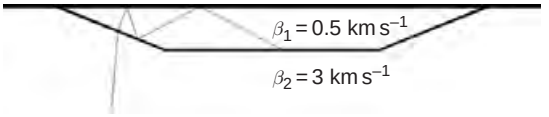


Figure 11 Illustration of the initial path of a ray with nearly vertical incidence, on encountering a low-velocity basin.

the Santa Clara basin originates from surface waves generated at the basin edge, and traveling across the basin at an angle almost orthogonal to the direction to the epicenter.

4.18.6.8 Trapping and Resonance in Basins

Understanding the effect of basins on strong ground motions is important because, especially in geologically active regions, cities are generally located in basins. Perhaps the most well-known example of basin response causing severe and damaging ground motions is the Mexico City basin. In the September 19, 1985 Michoacan, Mexico, earthquake ($M_W = 8.1$), the peak acceleration in Mexico City, over 300 km from the fault, was nearly the same as that observed on rock above the fault (Anderson *et al.*, 1986). The ground motion on one record in Mexico City lasted nearly 300 s with a predominant period of about 2 s. Structures in the 10–20-story range were severely stressed by these motions, and many collapsed, resulting in the loss of 12 000–20 000 lives. Fascinating movies showing the progression of strong motion across Japan, as recorded on K-Net and Kik-Net are described by Furumura *et al.* (2003), including an electronic supplement. An obvious feature of the movies is the prolonged duration of shaking in the major basins of Japan.

The basin effect can be understood as a combination of all the wave propagation phenomena discussed up to this point. Because of the complexity of the geological structure, the problem is more effectively investigated using elastic finite difference, finite element, or other numerical methods to model the basin response (e.g., Bard and Bouchon, 1985; Frankel and Vidale, 1992; Frankel, 1993; Graves, 1993, 1996; Pitarka *et al.*, 1994; Olsen *et al.*, 1995; Pitarka and Irikura, 1996; Stidham *et al.*, 1999; Sato *et al.*, 1999; Bao *et al.*, 1998). These methods are typically applied to model frequencies up to about 0.5–1 Hz.

These methods have been applied to understand the low frequencies of strong ground motion. Proceeding first clockwise around the Pacific rim, some of the areas that have been modeled include Wellington (Benites and Olsen, 2005), Kobe (Kawase, 1996; Pitarka *et al.*, 1998), Tottori (Furumura *et al.*, 2003), Tokyo (Sato *et al*, 1998a, 1998b, 1999). Seattle (Frankel and Stephenson, 2000; Frankel *et al.*, 2002; Pitarka *et al.*, 2004), Salt Lake City (Olsen *et al.*, 1995), San Francisco (Graves, 1993), Santa Clara Valley (Frankel and Vidale, 1992; Stidham *et al.*, 1999), Los Angeles (Olsen and Archuleta, 1996; Wald and Graves, 1998; Graves *et al.*, 1998; Hartzell *et al.*, 1999; Olsen, 2000; Olsen *et al.*, 2003), San Bernardino (Frankel, 1993, 1994; Graves and Wald, 2004), Mexico (Furumura and Singh, 2002), Granada (Gil-Zepeda *et al.*, 2002), and Rome (Olsen *et al.*, 2006).

Considering the results of these studies, it is possible to come to some general conclusions about the nature of ground motions in basins. The first is that the 2-D and 3-D structures are important for trapping energy and increasing the amplitude and duration of the shaking above what is predicted by 1-D models (e.g., Hartzell *et al.*, 2000). A second conclusion is that the pattern of high ground motions depends strongly on the direction from which the seismic energy enters the basin, and hence the location of the earthquake. A third conclusion is that some basin phenomena remain to be explained, as for instance the Mexico City observations have not yet been fully modeled. However, perhaps the most important conclusion is that if the basin structure (geometry, P and S velocity, density, and Q) is known in sufficient detail, and if the earthquake source is also known, then the ground motion in basins can be synthesized at low frequencies. Measuring the basin structure in sufficient detail is not easy, and ground motions are quite sensitive to the velocity model (e.g., Graves and Wald, 2004). Thus, it is essential that basin models be tested and improved. An important challenge is to find efficient ways to use the weak motions from small and distant earthquakes to improve the velocity structure.

4.18.6.9 Long Linear Valleys Acting as Waveguides

Numerical simulations have recognized that a long, linear basin can also act as a waveguide, trapping surface-wave energy within the basin and guiding it with a relatively small decrease with distance of the amplitude of the shaking. This phenomenon has not been recognized in actual data, but shows up clearly in some recent synthetic ground motions for an

earthquake on the San Andreas Fault (Olsen *et al.*, 2006; **Figure 12**). Olsen simulated ground motions in southern California for two magnitude-7.7 earthquakes on the San Andreas Fault. The final fault slip is the same for both simulations, but in one case the rupture is simulated to propagate from northwest to southeast, and in the other from southeast to northwest. The simulated ground motions for the first case show large amplitudes in the Imperial-Mexicali Valley (**Figure 12(a)**), which is a sediment-filled northward extension of the Gulf of California (Sea of Cortez). Energy is effectively refracted out of the high-velocity mountains on either side of the valley, and is guided southeast along the axis of the valley. In the second case, rather than following the San Andreas Fault, the strongest shaking is in the sedimentary basin on the south side of the San Gabriel Mountains (**Figure 12(b)**). Movies showing the simulations make it clear that for the northwest epicenter, energy is trapped and propagated at high amplitude through the southern Imperial Valley, and that for the southeast epicenter, energy is trapped and propagated at high amplitude through the essentially continuous sedimentary valley south of the Sierra Madre Mountains into Los Angeles. The importance of this phenomenon is that it predicts ground motions in excess of 1 m s^{-1} throughout this entire heavily populated region of southern California.

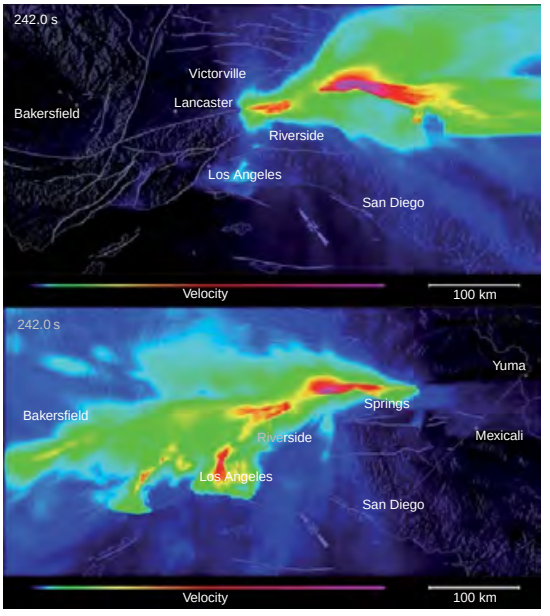


Figure 12 (a, b) Two maps showing the peak velocity of simulated ground motions for earthquakes on the southern segment of the San Andreas Fault in southern California. This is a map view, with northeast at the top. Topography is indicated by the raised-relief grayscale. Flat region in the foreground is the Pacific Ocean. Lines that criss-cross the region are major faults. In both cases the extent of fault rupture is the same, and can be approximately inferred from the portion of the San Andreas Fault highlighted by the large peak velocity in both figures. In case (a) the rupture nucleates at the northwest end (left end) of the rupture and propagates towards the southeast (right). In case (b), the rupture nucleates at the southeast end (right end) of the rupture and propagates towards the northwest (left). The velocity scale on both figures is the same. The colorscale at the bottom gives the peak velocity in m s^{-1}. Figures are from the Southern California Earthquake Center website and illustrate the results of Olsen *et al.* (2006). Image credit: Amit Chourasia, San Diego Supercomputer Center; Kem Bak Olsen, San Diego State University; Thomas Jordon, Southern california Earthquake Center.

4.18.6.10 Multipathing

With complexities in the velocity model, it is obviously possible for some circumstances to exist in which waves take multiple paths to arrive at the same location and at the same time. When this focusing effect occurs, amplitudes can be unusually large. One case where this has contributed significantly to amplitudes larger than expected is in the 1994 Northridge earthquake, where a lens effect focused energy in the vicinity of Santa Monica (Gao *et al.*, 1996; Davis *et al.*, 2000). Through analysis of data from a dense deployment of instruments to record aftershocks, these studies were able to show that aftershocks with different locations would have energy focused in different locations within Santa Monica in a manner predicted from classical lens theory. Graves *et al.* (1998) examined the focusing in this region using a finite-element wave propagation model. A second case is the 1995 Kobe earthquake, where damage was offset from the fault in a location where direct waves arriving through sediments coincided with signals that traveled through rock on the other side of the fault, and then converted to surface waves and propagated horizontally to the damage zone (e.g., Pitarka *et al.*, 1997, 1998).

4.18.6.11 Scattering

Anyone who has spent much time looking at rocks in the field recognizes that geological formations are not generally homogeneous. Rather, one often sees irregularities in the nature of the rocks on all

scales. These irregularities can cause small irregularities in the elastic moduli and in the density, which in turn cause seismic waves to be scattered. The scattering may be weak, with only a small fraction of the energy of a wave being scattered out of the main wave at any one irregularity. However, the net effect is sufficient to cause the coda that appears on the seismograms after the direct waves. The effect is illustrated in **Figure 13**. When each irregularity is excited by the seismic wave, it acts as a weak secondary source radiating waves, strongly preferring S waves (Aki, 1992; Zeng, 1993), in every direction. Although scattering may be weak, it is sufficient that the source radiation pattern (eqns [10] and [11]) is not generally observed at high frequencies (e.g., Liu and Helmberger, 1985; Satoh, 2002).

A basic theory for the time dependence of the seismic coda was described by Aki (1969) and expanded by Aki and Chouet (1975), Rautian and Khalturin (1978), and Phillips and Aki (1986). Aki modeled the seismic coda using a single-scattering approximation in a homogeneous medium. Under this model, the amplitude of the coda, $A(f,t)$, is given by

$$A(f,\,t) = S_{\mathrm{F}}(f)S_{\mathrm{SG}}(f)\phi(f,\,t) \qquad [30]$$

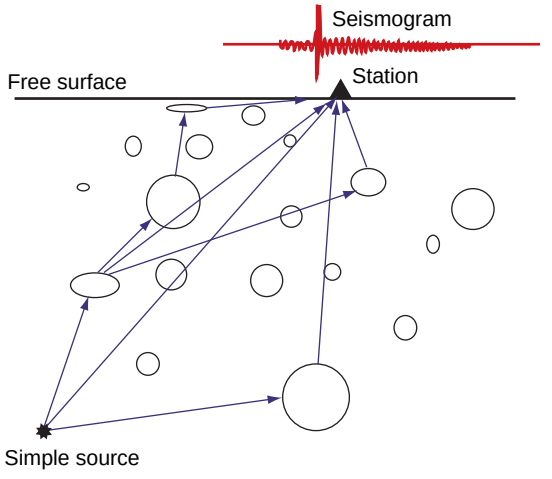


Figure 13 Schematic illustration of how scattering from random velocity irregularities in a medium will generate a coda on a seismograms. Zeng Y (2006) Broadband strong ground motion simulation using a composite source model, Proceedings of the 8th US National Conference on Earthquake Engineering, 18–22 April, 2006, San Francisco, CA, USA, Paper No. 1445.

where $S_{\mathrm{F}}(f)$ is the source spectrum as radiated at the fault, $S_{\mathrm{SG}}(f)$ is the modification to the spectrum by surface geology (also known as the 'site effect'), and $\phi(f,t)$ depends on how the seismic waves are scattered and attenuated. The functional form for $\phi(f,t)$ is predicted, under the single-scattering model, to be

$$\phi(f,\,t) = t^{-\upsilon}\exp[-\pi ft/Q_{\mathrm{c}}] \qquad [31]$$

where t is the elapsed time after the event, Q_{c} is coda Q, and $\upsilon = 1$ for scattering of body waves. The scattering model for the shape of the coda has subsequently been extended to multiple scattering (e.g., Zeng *et al.*, 1991; Zeng, 1991, 1993), scattering in multiple layers (e.g., Yomogida *et al.*, 1997) and multiple scattering including conversion from P to S energy (Zeng, 1993). The shape of the coda has also been investigated using an energy diffusion model (e.g., Frankel and Wennerberg, 1987) and by numerical modeling (e.g., Richards and Menke, 1983; Frankel and Clayton, 1984; Menke and Chen, 1984). The numerical studies show how scattering causes attenuation of the amplitude of the direct waves with some energy arriving later, perhaps especially at high frequencies. Frankel and Wennerberg (1987) conclude that Q_{c} is predominantly controlled by the intrinsic attenuation. It should be noted that scattering from major geological structures contributes very significant energy to the seismic coda (e.g., Spudich and Iida, 1993).

Study of the seismic coda can be useful for strong-motion seismology. According to eqn [31], the time dependence of the coda is the same at all stations in a small region, because at large times the waves contributing to the coda have all traveled large distances compared to the station separation. Rautian and Khalturian noted that $\phi(f,t)$ describes the time dependence for an impulsive source, so with multiple recordings at a single station, the relative values of $S_{\mathrm{F}}(f)$ among these multiple sources can be recovered. Motivated by Tsujiura (1978) and Tucker and King (1984), Phillips and Aki (1986) recognized that under these assumptions records from a single earthquake recorded at multiple stations could be used to estimate the site response. Some subsequent studies have supported that conclusion (e.g., Su *et al.*, 1998). Chin and Aki (1991) used differences in the site response between weak and strong motion to propose that the strong-motion site response was modified by nonlinear response of the near-surface geological deposits.

4.18.6.12 Correlation and Coherency

The effects of inhomogeneous geology on seismic waves, including scattering, focusing, and defocusing, and traveltime delays, cause ground motions to differ significantly even at stations with relatively small spatial separations. To quantify this it is useful to bring in the concept of the correlation and coherency of seismograms from nearby stations. In general, when two stations are located near to each other, the seismograms bear a closer resemblence at low frequencies than at high frequencies. Suppose that $u_A(\mathbf{x}_A, t)$ and $u_B(\mathbf{x}_B, t)$ are two nearby seismograms. The correlation coefficient, C_0, of these two seismograms, over the time interval $t_1 \le t \le t_2$, with no time lag, is estimated by

$$C_0 = \frac{\int_{t_1}^{t_2} u_A(\mathbf{x}_A, t) u_B(\mathbf{x}_B, t) \mathrm{d}t}{\left\{ \int_{t_1}^{t_2} u_A^2(\mathbf{x}_A, t) \mathrm{d}t \int_{t_1}^{t_2} u_B^2(\mathbf{x}_B, t) \mathrm{d}t \right\}^{1/2}} \quad [32]$$

The time interval can be taken to cover the entire seismogram, just the P wave, just the S wave, or other windows of interest.

The correlation coefficient in eqn [32] may be expected to depend on the direction to the source. Waves coming in perpendicular to the line between the two stations will arrive at nearly the same time, and thus could have a high correlation, while waves at other angles will have a phase delay due to the wave propagation speed that would decrease the correlation. A second way to quantify the similarity, especially appropriate for the P- and S-wave windows, is with the lagged correlation coefficient, C_L, which introduces a time shift to the seismograms accounting for this phase delay. Let us suppose that the direct arrival of a certain phase (e.g., the P or S wave) at the stations is t_A and t_B, respectively.

$$C_L(t_A - t_B) = \frac{\int_{t_1}^{t_2} u_A(\mathbf{x}_A, t - t_A) u_B(\mathbf{x}_B, t - t_B) \mathrm{d}t}{\left\{ \int_{t_1}^{t_2} u_A^2(\mathbf{x}_A, t - t_A) \mathrm{d}t \int_{t_1}^{t_2} u_B^2(\mathbf{x}_B, t - t_B) \mathrm{d}t \right\}^{1/2}} \quad [33]$$

When eqn [33] is applied, the resulting correlation coefficient is expected to be less dependent on the direction of wave arrival, and thus the function that is derived may have more capability to predict the correlation coefficient of these wave packets when waves come from a direction that is not previously observed. If $t_B = 0$ and t_A is considered a variable, the time of the largest positive peak in $C_L(t_A)$ is used to measure the difference in arrival times of a phase at the two stations. Note that for later arrivals, such as the P- or S-wave coda, the incoming waves tend to be highly scattered, and thus will be arriving at the station from many different directions. Thus the time shift for the direct arrivals does not apply to the later arrivals. Consequently for these later arrivals, eqn [33] is not expected to show much if any greater correlation than what is found in eqn [32].

The coherency is a frequency-domain parameter that measures the similarity of seismograms at two nearby stations over a frequency band of width Δf. The separation between the stations is $\Delta x = |\mathbf{x}_B - \mathbf{x}_A|$, where there is an implicit assumption of spatial stationarity in the region where the measurements are made. After Menke *et al.* (1990) and Abrahamson *et al.* (1991), the coherency can be written as

$$\gamma_{AB}(f, \Delta f, \Delta x) = \frac{\langle S_{AB}(f) \rangle}{(\langle S_{AA}(f) \rangle \langle S_{BB}(f) \rangle)^{1/2}} \quad [34]$$

The coherence is $|\gamma_{AB}(f, \Delta f, \Delta x)|^2$. For a real time series,

$$S_{AB}(f) = \bar{u}_A(f) u_B(f) \quad [35]$$

is the cross spectrum (the Fourier transform of the cross correlation) of the two seismograms. In eqn [35], $u_A(f)$ and $u_B(f)$ are the Fourier transforms of $u_A(\mathbf{x}_A, t)$ and $u_B(\mathbf{x}_B, t)$, respectively, and $\bar{u}_A(f)$ is the complex conjugate of $u_A(f)$. The coherency $\gamma_{AB}(f)$ is in general a complex function. For a single $u_A(\mathbf{x}_A, t)$ and $u_B(\mathbf{x}_B, t)$ and a single frequency, the magnitude of $\gamma_{AB}(f)$ is equal to unity. The notation $\langle S_{AB}(f) \rangle$ is used to indicate that $S_{AB}(f)$ is averaged over the frequency band of width Δf. Abrahamson *et al.* (1991) smooth $S_{AB}(f)$, which amounts to the same process. Thus $\gamma_{AB}(f)$ is calculated by taking an ensemble average. According to Abrahamson *et al.* (1991), the real part of $\gamma_{AB}(f)$ then gives the similarity of ground motions as a function of frequency with no adjustment for inclined wave propagation, and $|\gamma_{AB}(f)|$ gives the lagged coherency which does include that adjustment. Using the time-domain eqns [32] and [33], the frequency dependence of the lagged correlation coefficient can be extracted by applying a band-pass filter to the seismograms before evaluation of the equations.

In general, a way to present the results when multiple stations are available is to plot $|\gamma_{AB}(f, \Delta f, \Delta x)|$ as a function of the station separation Δx (e.g.,

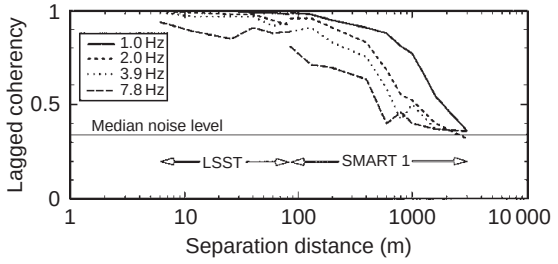


Figure 14 Lagged coherency of ground motions, as a function of station separation, at the SMART 1 array, sited on soft sediments in eastern Taiwan (Abrahamson *et al.*, 1991).

Figure 14). Figure 14 shows an example where lagged coherency is high for 1 Hz waves to separations of 100 m. Beyond this distance it starts to decline to a state of essentially no coherency at a separation of 2000 m. As frequencies increase, the coherency decreases. The coherency is highest when direct P or S waves are arriving, and decreases throughout the coda of the seismogram (e.g., Der *et al.*, 1984; Menke *et al.*, 1990; Vernon *et al.*, 1991). For a site in the Adirondack Mountains of northern New York State, Menke *et al.* (1990) suggest that the spatial coherence $|\gamma_{AB}(f, \Delta f, \Delta x)|^2 = \exp(-c_\gamma f \Delta x)$, where $c_\gamma = 0.4$–$0.7\,\mathrm{km}^{-1}\,\mathrm{Hz}^{-1}$. Based on a study of coherence in the San Fernando Valley of southern California, Hough and Field (1996) question whether this relationship is universal.

The spatial variability of the amplitude is also important, but not studied extensively. Steidl *et al.* (1996) suggest that when the input waves from the basement are known from a deep borehole up to 20 km away from a site of interest, and the geological structure is known at the site of interest, the amplitudes of the surface motions are highly predictable. Hartzell *et al.* (1997) characterize the standard deviation of ground motion as a function of station separation on a large scale (0–30 km) for stations in the Los Angeles, California, region that recorded aftershocks of the 1994 Northridge earthquake. On average, amplitudes are more variable for larger separations. Tsuda *et al.* (2006) look at predictability of the spectral amplification site response as a function of station separation and the similarity of V_{S30} on a scale of 0–5 km for the dense network of strong motion instruments in Yokohama, Japan. At low frequencies, when V_{S30} is similar, site response there is predictable within a factor of two for station separations up to 5 km. The predictability of the spectral site amplification function from nearby stations is lower and less correlated with V_{S30} at frequencies of 5–10 Hz. In the separation range of most interest for long engineered structures, such as bridges, this study did not have much resolution.

4.18.6.13 Effect of Topography

Topography can also have an effect on seismic waves at high frequencies. Consider for instance a linear mountain range with a base that is ~10 km wide, bounded by linear valleys on either side. If the shear velocity is $3\,\mathrm{km\,s}^{-1}$, waves with periods shorter than ~3 s, or frequencies greater than ~0.3 Hz, have wavelengths smaller than the size of the base of the mountain range. As a general rule, when the wavelength is comparable to or smaller than the size of the geological feature, it is quite reasonable to expect that feature to affect the ground motion. Sanchez-Sesma and Campillo (1991) show the range of numerical effects of topographic models. A good early observational example is described by Tucker *et al.* (1984), where the Fourier amplitude spectra at sites on ridges were amplified by as much as a factor of eight in some frequency bands. Thus topography does affect strong ground motions.

The effect of topography on strong motion is, however, not yet well understood, perhaps in part due to the difficulty in numerical modeling of the effects and in part due to the necessity to know complex subsurface geological structure (Bard and Tucker, 1985). Geli *et al.* (1988) and Assimaki and Gazetas (2004) provide thorough reviews of research. Geli *et al.* (1988) find that the amplification of ground motions on mountain tops does occur at wavelengths comparable to the width of the base of the mountains, but that often the observed amplifications tend to be larger than the predictions for a uniform medium with simple topographic structures on the surface. They attribute the differences to topographic complexity and to complex subsurface geological structure. Assimaki and Gazetas (2004) and Assimaki *et al.* (2005) concluded that amplitudes that were too high to be explained by topographic effects alone were related to soil conditions.

Some of the strongest ground motions have been observed on ridges. The Pacoima Dam accelerogram, with peak acceleration of 1.25*g* (Trifunac and Hudson, 1971) is in an area of complex topography on a ridge which forms the abutment of a dam. Anooshehpoor and Brune (1989) built a 3-D model of foam rubber to study the topographic effects, and concluded (consistent with the theoretical models cited by Geli *et al.* (1988)) that the net effect of the

topography was small in this case; the frequency of the peak ground motions did not match the freqeuncy of the topographic amplification and the complex topography shielded the site from waves from some parts of the source.

Hartzell *et al.* (1994) investigated severe damage on Robinwood Ridge in the Loma Prieta, California, earthquake by studying the topographic effects on ground motions in aftershock recordings (there were no records from this ridge in the main shock). They observed very complex effects, including propagation of significant amounts of energy parallel to the ridge even though the ridge is oblique to the direction from the source. High-frequency spectral peaks can be attributed to local variations in the site response. Although not stated in the paper, the observation of lower-frequency energy traveling parallel to the ridge suggests the possibility that energy is trapped in the ridge, perhaps in some ways analogous to energy trapped in basins, as the ridge also is likely to have lower velocities than the rock at greater depths.

A third important topographic effect was observed in the 1994 Northridge, California, earthquake. The strong-motion station at the Tarzana Cedar Hill Nursery recorded a peak acceleration of 1.78 g in the east–west component. Spudich *et al.* (1996) installed a dense network of stations to study the regional amplification on aftershock records. They found that the hill does amplify ground motions at 3.2 Hz, about the same frequency as was dominant on the accelerogram, by about a factor of two for the east–west component that is parallel to the long axis of the hill (**Figure 15**). The strongest amplifications in the aftershocks, however, were on the north–south motions, transverse to the long axis of the hill.

With the growth in computational capabilities, models of the effects of topography on ground motions are becoming more sophisticated (e.g., Bouchon and Barker, 1996; Ohminato and Chouet, 1997; Hayashi *et al.*, 2001). Larsen (2002) included topography in a finite-difference model of ground motions in southern Nevada. This study found that the complex topography of the region (Basin and Range) caused considerable scattering of surface waves. As these capabilities improve, significant progress in the understanding of topographic effects is to be expected.

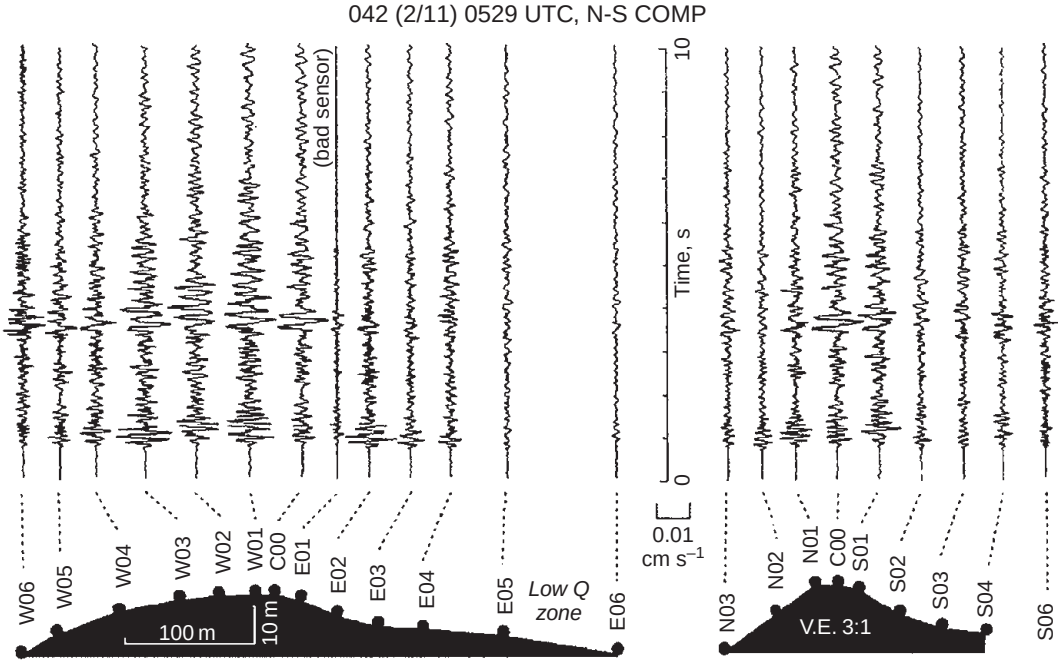


Figure 15 North-south component of seismograms from an aftershock of the 1994 Northridge earthquake, recorded at several stations on a 20-m-high topographic feature south of the earthquake. The seismograms on the left show a profile along the long, east–west axis of the hill, and the seismograms on the right show a north–south profile transverse to the hill. Note how the records show higher amplitudes on the hill than at stations at the base of the hill. From Spudich PA, Hellwig M, and Lee WHK (1996) Directional topographic response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake: Implications for mainshock motions. *Bulletin of the Seismological Society of America* 86(1B): S193–S208.

4.18.6.14 Q

When a seismic wave travels through even a homogeneous medium, it loses a small amount of energy. If the incoming amplitude is A_0, then a plane wave with a single-frequency (f) wave traveling a distance x through the medium with speed β has the amplitude

$$\frac{A}{A_0} = \exp\left(-\frac{\pi x f}{Q\beta}\right) = \exp\left(-\frac{\pi}{Q}\frac{x}{\lambda}\right) \quad [36]$$

In the distance of one wavelength, λ, the wave energy, which is proportional to the square of the amplitude, decreases by exp $(-2\pi/Q)$, where the normalization by 2π comes from one complete cycle in radians. Q is thus a dimensionless parameter, where higher Q or quality, implies lower losses.

In general, Q is different for P and for S waves, and it is a function of frequency, lithology, and depth. For the sake of finite-difference modeling at $f < 2$ Hz, Olsen *et al.* (2003) proposed the rule of thumb that, for V_S in meters per second, $Q_S/V_S = 0.02$ for V_S less than 1–2 km s^{-1}, and $Q_S/V_S = 0.1$ for layers with higher velocities. From studies of coda Q, Aki (1980) found that at frequencies greater than 1 Hz, Q shows clear frequency dependence, which he characterized as

$$Q = Q_0 f^{\delta} \quad [37]$$

where $0 < \delta < 1$. Erickson *et al.* (2004) examined Q in Lg waves, which represent S-wave energy trapped in the crust, for the frequency band from 0.5 to 15 Hz (**Figure 16**). At 1 Hz, typical values of crustal Q for shear waves are in the range of 100–200 in the western United States, and about 600–700 in the eastern United States. The high values of Q in the stable regions of eastern North America are responsible for earthquakes there being felt at significantly larger distances than the distances where the same-sized event is felt in the tectonically active west. In shallow sediments, shear-wave Q is much lower, in the range of 10–30 or less (e.g., Seale and Archuleta, 1989; Malagnini, 1996; Bonilla *et al.*, 2002).

The spectral decay parameter kappa (κ) has been found useful in engineering studies for characterizing the Fourier spectrum of ground motion. Anderson and Hough (1984) found that the high-frequency Fourier spectrum of acceleration at the surface for frequencies above 2 Hz during earthquakes with M > 5 behaves as

$$a(f) \sim a_0 \exp(-\pi\kappa f) \quad [38]$$

as shown in **Figure 17**, where a_0 gives the spectral level in the frequency band above the corner frequency but below the frequencies where the exponential decay starts to have a significant effect.

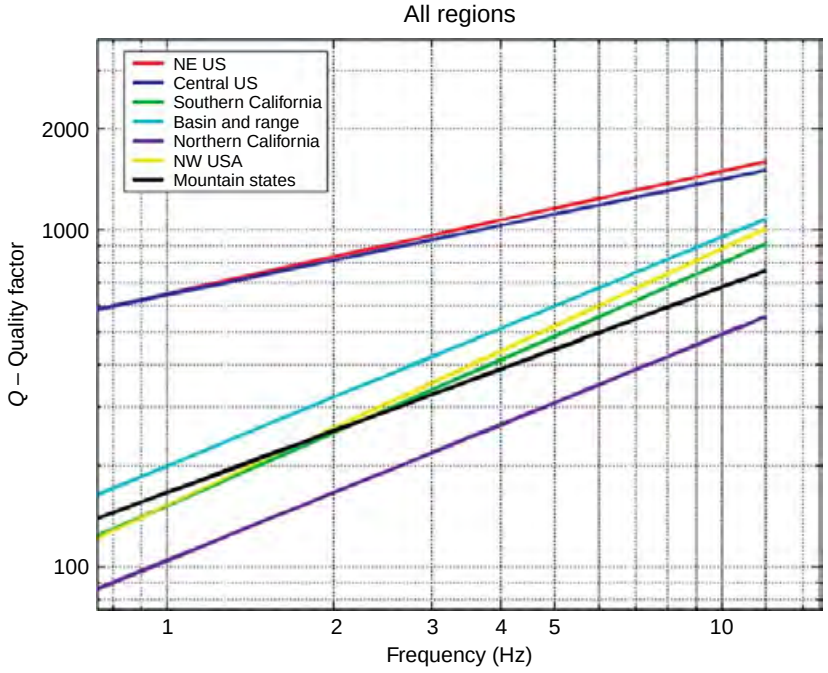


Figure 16 Quality factor Q for L_g waves in several regions of the United States. From Erickson *et al*. (2004).

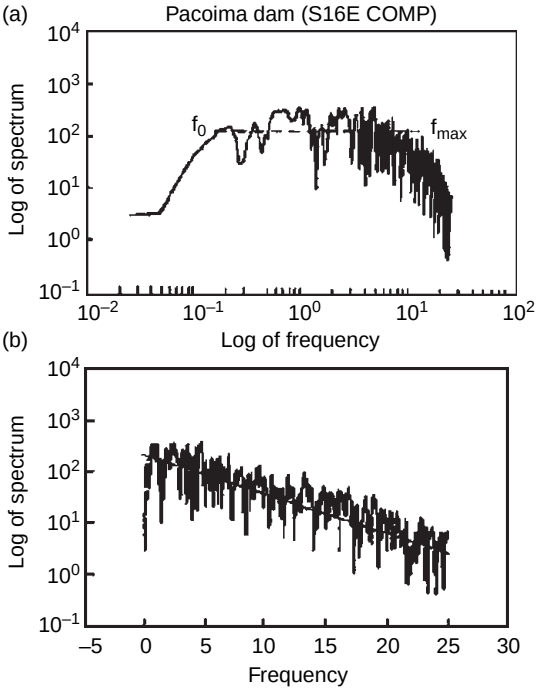


Figure 17 Figure showing measurement of kappa.

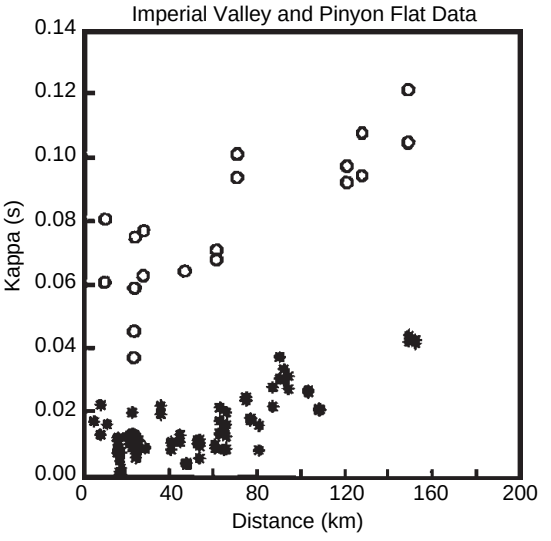


Figure 18 Kappa vs distance at two stations. Both stations are located in southern California. The Imperial Valley station (o) is located in El Centro, on deep sediments. The Anza station (*) is located 120 km northwest, on granite of the Penninsular Range. The distance dependence is similar, which Anderson interpreted as due to wave propagation affecting seismograms for both stations in the same way. The intercept is shifted, which Anderson interpreted as due to different values of Q in the shallow gological materials immediately below the stations (e.g., the upper 5 km). From Anderson JG (1986) Implication of attenuation for studies of the earthquake source. Ewing M (ed.) *Geophysical Monograph 37*, (*Maurice Ewing Series 6*): *Earthquake Source Mechanics*, pp. 311–318. Washington, DC: American Geophysical Union.

In addition, they found that κ depends on both the station and the distance from the source to the station (**Figure 18**). They suggested that the parameter can be modeled as

$$\kappa = \kappa_0 + \tilde{\kappa}(r) \qquad [39]$$

where κ_0 is a station parameter, and $\tilde{\kappa}(r)$ gives a distance dependence related to Q along the entire path.

The parameter κ enters the spectral amplitude in the same way as attenuation loss through eqn [36]. Furthermore, the distance dependency implies that it is at least in part related to Q. If two assumptions were true, κ would be entirely an attenuation parameter. The first is that the acceleration spectrum is flat at high frequencies (i.e., the displacement spectrum falls off as ω^{-2} above the corner, as suggested by Aki (1967), Brune (1970), and Hanks (1979, 1982). The second is that Q is independent of frequency. If both were true, then

$$\kappa = t^* = \int_{\text{path}} \frac{\mathrm{d}r}{Q(r)\beta(r)} \qquad [40]$$

where t^* is defined as a standard parameter for introducing Q losses to teleseismic body waves. Unfortunately, as discussed above, the two assumptions are not generally valid, but the relationship in eqn [40] is still useful.

There is no reason to abandon the assumption that the average displacement spectrum falls off as ω^{-2}, but deviation of individual sources from the ω^{-2} model is an explanation for the observation by Purvance and Anderson (2003) that there is a source contribution to κ.

As mentioned above, Aki (1980), and numerous subsequent papers, have found that above 1 Hz, in the crust Q depends on frequency as in eqn [37]. Hough and Anderson (1988) noted that over limited frequency ranges where eqn [37] is used to model the frequency dependence of Q, and since $0 < \delta < 1$, an alternative two-parameter model is

$$\frac{1}{Q_0 f^{\delta}} \approx \frac{1}{Q_i} + \frac{1}{Q_d f} \qquad [41]$$

where Q_i is a frequency-independent contribution to the total Q and $Q_d f$ is a frequency-dependent contribution. If eqn [41] is substituted into eqn [36], then

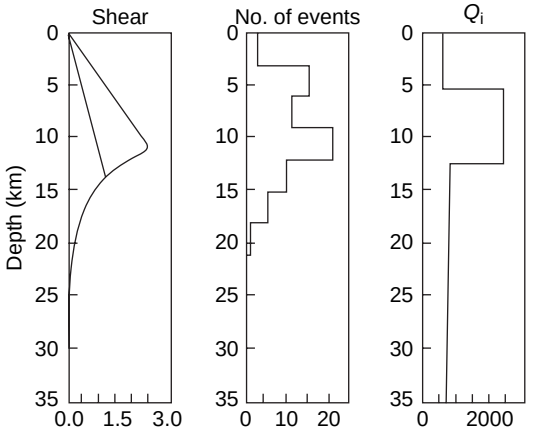


Figure 19 Q_i vs Z, as determined by Hough and Anderson (1988).

only the term in Q_i affects the shape of the spectrum, which is measured by κ. On this assumption, Hough and Anderson inverted the distance dependence of κ for a depth dependence of Q_i, with average results shown in **Figure 19**. They suggested that high Q_i occurs at the seismogenic depths in California, and that Q_i decreases at depths where seismogenic creep occurs instead of brittle fracture.

An approach that does not demand the approximation in eqn [41] can be described starting with eqn [4]. The effect of wave propagation on a phase (e.g., S wave) in the seismogram can be expanded as

$$S_{GF}(r,f) = S_P(r,f)S_{SG}(f)\exp(-\pi\kappa_0 f) \qquad [42]$$

where $S_p(r,f)$ gives the effect of the path, which may be interpreted as the propagation through seismic basement rock, from the source to below the site, $S_{SG}(f)$ describes the all the effect of propagation through the surface geology from seismic basement to the site described above excluding attenuation, and $\exp(-\pi\kappa_0 f)$ gives the effect of attenuation associated with the surface geology at the site and κ_0 has the same meaning as in eqn [39]. The path term can be further separated as

$$S_P(r,f) = g_{GS}(r)\exp\left(-\frac{\pi rf}{Q(f)\beta}\right) \qquad [43]$$

where $Q(f)$ takes on the more common functional form of eqn [37] and $g_{GS}(r)$ describes the geometrical spreading of the phase, for instance $g_{GS}(r) = 1/r$ for body waves in an infinite homogeneous medium (*see* Chapter 4.02). Some of the papers that utilize this approach for describing, and predicting, the spectra of ground motions include Margaris and Boore (1998), Malagnini and Herrmann (2000), Malagnini *et al.* (2000a,b), Akinci *et al.* (2001), Malagnini *et al.* (2002), Bay *et al.* (2003), Malagnini *et al.* (2004), and Scognamiglio *et al.* (2005).

4.18.6.15 Nonlinearity

The usual assumption in wave propagation is that the amplitudes of strains are small, and that the restoring stress in the deformed material is proportional to the strain applied by the wave. However, again it is a common experience that soils can be deformed permanently by various straightforward methods (e.g., compacting by stomping with a boot). In the case of compacting soils, the deformation exceeds the elastic limit, and as a consequence the elastic forces are not sufficient to cause the soil to return to its original shape. Under strong ground motion, it is possible for the waves to exceed the elastic limit of the material, where restoring stress is no longer proportional to the strain. Kramer (1996) provides a good textbook review of the topic.

In the linear models, the only material constants are the elastic moduli. In the nonlinear case, every material can be nonlinear in its own way, as the stress is no longer proportional to the strain. **Figure 20** shows one commonly used shape of nonlinear stress–strain curve, and defines some important variables. Those are the strain and stress at the corner, ε_C and σ_C, respectively; the low-strain shear modulus, μ_0; the secant shear modulus $\mu_{SEC} = \sigma_C/\varepsilon_C$; and the tangent shear modulus μ_{TAN} which depends on the local slope of the curve. The ratio of the secent elastic modulus to the low-strain shear modulus, μ_{SEC}/μ_0 is called the modulus reduction. **Figure 21** shows the secant modulus as a function of shear strain, corresponding to the stress–strain curve in **Figure 20**. The area inside the hysteresis loop is proportional to the energy loss in one deformation cycle. The damping ratio is defined by

$$\varsigma = \frac{W_D}{4\pi W_S} = \frac{1}{2\pi}\frac{A_{loop}}{\mu_{SEC}\varepsilon_C^2} \qquad [44]$$

where W_D is the dissipated energy, W_S is the maximum strain energy, A_{loop} is the area inside one complete cycle of the hysteresis loop, and ε_C is the peak strain in the deformation cycle (Kramer, 1996).

Comparing with the definition of Q (Eqn [36], see also Kramer, 1996), the damping ratio is related to Q by

$$Q = \frac{1}{2\varsigma} \qquad [45]$$

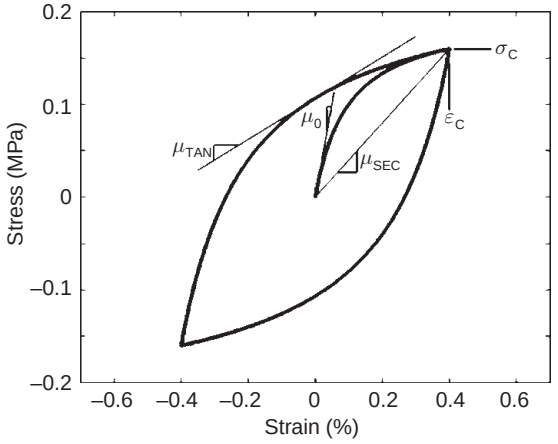


Figure 20 Definition of shear modulus parameters that are used in studies of nonlinear response of soils. This plot uses the functional form employed by Yu *et al.* (1993) to show the restoring stresses in a soil as a function of the deformation state during cyclic deformation. The soil starts in an undeformed state, strain is increased to ε_C, decreased to $-\varepsilon_C$, then increased again to ε_C, where in this case $\varepsilon_C = 0.006$.

Figure 21 thus also shows the quality factor Q corresponding to this model. In this model, Q is independent of frequency. The measurements of Q in soils cited above were taken at small amplitudes, and are much smaller than predicted by this model, perhaps indicating that additional phenomena contribute to the overall attenuation in soils besides the nonlinearity in the typical engineering models.

In the engineering literature, the shear modulus is typically given by the symbol G, while in seismology it is given the symbol μ. The dimensions of μ are force/area, and thus the same as pressure. Units and commonly used conversions are Pascal ($1\,\text{Pa} = 1\,\text{Newton m}^{-2}$), Megapascal ($1\,\text{MPa} = 10^6\,\text{Pa}$), dyne-cm^{-2} ($1\,\text{dyne-cm}^{-2} = 0.1\,\text{Pa}$), bars ($1\,\text{bar} = 0.1\,\text{MPa} = 10^6\,\text{dyne-cm}^{-2} \sim 1\,\text{atm}$), or pounds per square inch ($1\,\text{psi} = 6.895 \times 10^3\,\text{Pa}$).

Through the analysis of a vertical array of accelerometers, it is possible to infer the stress–strain relationships of the materials between pairs of instruments (e.g., Pavlenko and Irikura, 2002; Pavlenko and Loh, 2005). **Figure 22** shows an example for the site at Port Island that experienced strong liquefaction during the 1995 Kobe, Japan, earthquake. There are many interesting features to notice in **Figure 22**. The first observations from the variety of shapes is that nonlinear soil response can be exhibited in multiple ways. Not all of the stress–strain curves follow the pattern shown in **Figure 20**. In particular, at early times in the record, the curves at 16–18 m depth on the north–south component and at 31 m depth on the east–west component show a good example of a stress–strain relationship where μ_{TAN} initially decreases, but then increases with increasing strain. This phenomenon, known as strain hardening, is a

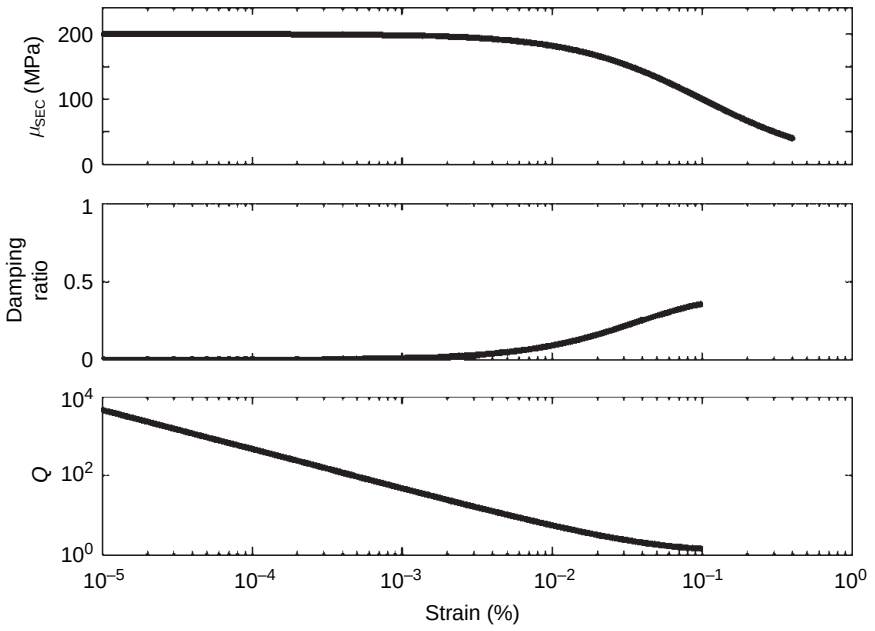


Figure 21 Values of μ_{SEC}, the damping ratio, and Q as a function of ε_C, for the stress-strain curves given in **Figure 18**.

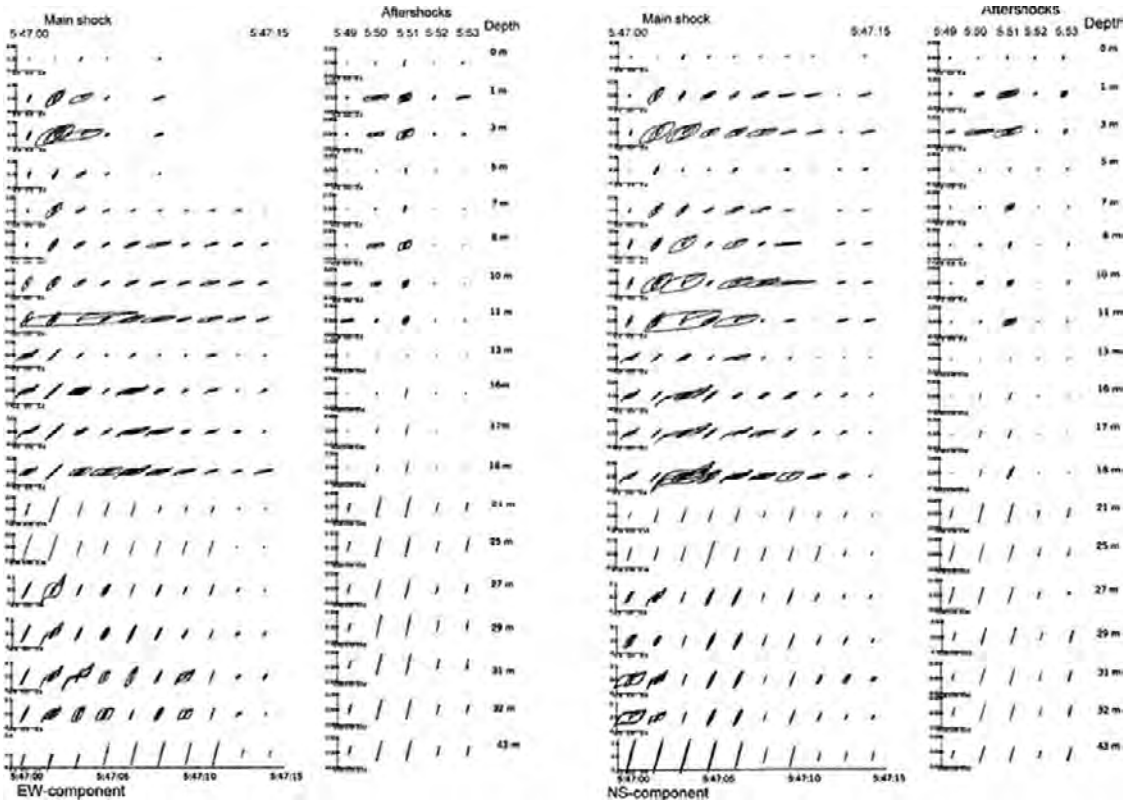


Figure 22 Stress–strain relations as a function of depth and time at Port Island during the $M_W = 6.9$, 1995 Kobe, Japan, earthquake and four of its aftershocks. From Pavlenko O and Irikura K (2002) Changes in shear moduli of liquefied and nonliquefied soils during the 1995 Kobe earthquake and its aftershocks at three vertical-array sites. *Bulletin of the Seismological Society of America* 92(5):1952–1969.

characteristic of sandy sites that are saturated with water (e.g., Yang *et al.*, 2002; Bonilla *et al.*, 2005). A second important point is that μ_{SEC} for many of the stress–strain curves decreases early in the earthquake, but then does not recover to the earliest values later in the record as the amplitude of shaking decreases. At some depths, it appears to not be fully recovered even in aftershocks up to 6 min later. A third point is that some layers, characterized by different lithologies and documented by Pavlenko and Irikura (2002), behave nearly linearly throughout the earthquake and its aftershocks.

A model for the average effects of nonlinearity on the amplitudes of ground motions have already been introduced in **Table 2** and **3**. These are based on empirical results for how much spectral amplitudes are reduced below the average amplitudes in the linear case. For site class D, under the strongest level of shaking, ground motion amplitudes are reduced to about two-third of what would be expected by simple linear scaling from the weak-motion case. Data from the Port Island site in Kobe and subsequent analysis suggests that nonlinearity is less effective in reducing the amplitudes of the vertical component of ground motions, so that the ratio of vertical to horizontal peak values might be increased (e.g., Yang and Sato, 2000; Yang *et al.*, 2002).

If μ_{SEC} is decreased during strong motion, as indicated on **Figures 20–22**, then the shear velocity would also be decreased. Consequently, based on eqn [28], the resonant frequency of a layer would be decreased. More generally, the effect of the surface geology, $S_{SG}(f)$ as defined in eqn [42], is expected to be different during strong motion with nonlinear response than it is during weak motion. Furthermore, the increase in ς or decrease in Q would result in decreased spectral amplitudes at high frequencies. O'Connell (1999) cautioned that linear wave propagation might mimic some of these effects. Nonetheless, several studies have identified those effects, and interpreted them as being a consequence of nonlinearity, in recent well-recorded earthquakes

(e.g., Chin and Aki, 1991; Satoh *et al.*, 1995; Aguirre and Irikura, 1997; Beresnev *et al.*, 1998a, 1998b; Cultrera *et al.*, 1999; Higashi and Sasatani, 2000; Satoh *et al.*, 2001; Frankel *et al.*, 2002; Rubinstein and Beroza, 2004; Tsuda *et al.*, 2006).

Several approaches are available to calculate the effect of nonlinearity on ground motions. In most cases, the overall idea is to assume by some means the ground motions that will be input to the bottom of a flat-layered geological structure as SH motion, and then calculate the response of the structure including nonlinearity. Approaches can be separated into two broad categories. The first is the 'equivalent linear' model, and the second is the 'full nonlinear' model. In the equivalent linear approach (e.g., Schnabel *et al.*, 1972), given initial input elastic parameters, one calculates the linear response of the layer (as in Section 4.18.6.5 above). Strains are then calculated, and the secant elastic moduli are determined from the modulus-reduction curves (**Figure 21**). These modified elastic moduli are used to recompute the layered response, and the system is iterated until it converges. In the equivalent linear approach, the nonlinearity increases the damping (decreases Q), and thus high frequencies are suppressed. In the full nonlinear approach (e.g., Joyner and Chen, 1975; Joyner, 1975; Elgamal, 1991; Yu *et al.*, 1993; Xu *et al.*, 1999; Archuleta *et al.*, 2003), the model uses a finite-element solution to the wave equation, and follows the stress–strain curves as in **Figure 20**, using the appropriate tangent elastic modulus to determine the stress–strain relationship at each point in the model. In the full nonlinear model, energy is converted to other frequencies (e.g., Yu *et al.*, 1993), and the high-frequency spectrum might increase in some cases. Hartzell *et al.* (2004) compare approaches to calculation of nonlinear effects, and present a compact summary of the dependence of modulus-reduction curves on material and depth.

4.18.7 Effects of Fault Finiteness

Large-magnitude earthquakes, which generate strong motions, are the result of slip over a large fault area. For instance, based on a regression of available data, Wells and Coppersmith (1994) found the relationship $M_W = 4.07 + 0.98 \log \Sigma_F$, where Σ_F is the area of the fault rupture, given in square kilometers. There is of course scatter around this mean trend. The standard deviation on this relationship is 0.24 magnitude units. Consider for instance the 1985 Michoacan, Mexico earthquake. The aftershock zone was about 180 km long and 50 km wide (Anderson *et al.*, 1986), giving an area of about $10^4\,\mathrm{km}^2$, approximately consistent with its magnitude $M_W = 8.1$. The 28 September 2004 Parkfield earthquake had a slip area about 30 km long and 10 km wide, giving an area of $300\,\mathrm{km}^2$, this time somewhat larger, but consistent within scatter, for $M_W = 6.0$ (Archuleta *et al.*, 2005). These examples make it clear that for a single site near the fault, where the ground motions are strongest, energy will be arriving from multiple directions, and direct waves from rupture on different parts of the fault will arrive at different times.

4.18.7.1 Spectral Shape

Eqns [13] and [14] provide a good starting point for considering the effects of the finite fault, even though the approximation they use of concentrating the total effect of the slip to a point is really only valid at a distance that is large compared to the source dimension. Recalling that rupture will propagate across the fault at a finite speed, typically smaller than the shear velocity, the duration of faulting T_F is controlled by the fault dimension. For a 60 km long strike-slip fault, the total duration of faulting will be $T_F \sim L/v_r \sim 20\,\mathrm{s}$ if $v_r \sim 3\,\mathrm{km\,s^{-1}}$. This implies that $\dot{M}_0(t)$ is a one-sided function with a duration of ~20 s. From the theory of the Fourier spectrum, a one-sided pulse transforms to a spectral amplitude that has its peak at zero frequency, is relatively flat until a corner frequency, and then falls off above the corner frequency. The corner frequency is inversely proportional to the pulse duration, that is, $f_c \sim 1/(2T_F)$. Since the area under the pulse is proportional to the moment, this shows that as the magnitude of an earthquake increases, the low-frequency spectral level increases but the corner frequency decreases. This scaling law was recognized by Aki (1967). **Figure 4** illustrates this effect, and the considerations in this section explain why the magnitude-dependent shapes in **Figure 4** are an inevitable consequence of the physics controlling strong motion.

4.18.7.2 Pulse Duration and Directivity

Directivity has a first-order effect on the ground motions. Directivity is the effect on the seismograms of rupture propagation along the fault. It impacts both the high-frequency and the low-frequency

seismograms. At low frequencies, directivity can be responsible for coherent, potentially very destructive pulses with large amplitudes. At high frequencies also, it can cause energy to arrive in short time intervals through its control over the the duration of shaking. To illustrate the effect, consider first **Figure 23**. This figure assumes an observation point O is located at an arbitrary location near a vertical strike-slip fault. For simplicity, consider a case where the strong motions near the source are dominated by direct S waves. The first S wave from the rupture arrives at O at time $t_1 = r_1/\beta$. The last S wave from the rupture arrives at time $t_2 = L/v_r + t_r + r_2/\beta$, the sum of the time it takes the rupture to reach the far end of the fault, plus the rise time, plus the traveltime of the ray from the end of the fault to the observation point O. Since the fault will be radiating energy continuously as the rupture progresses, S-wave energy will be arriving directly from the fault for the entire time interval from t_1 to t_2. Thus the duration of the S wave is

$$T_S = t_2 - t_1 = \frac{L}{v_r} + t_r + \frac{(r_2 - r_1)}{\beta} \qquad [46]$$

When T_S is averaged over observations that sample the S wave at many azimuths, it is reasonable to expect that $\bar{T}_S \sim T_F$, but at any one direction it may differ significantly.

Consider for instance the case where the observation is taken at a point near the origin of the rupture (E), t_r is small, and $v_r \approx 0.9\beta$. In this case, r_1 is small and r_2 is comparable to L, so the time interval over which direct arrivals of S waves will be arriving is on the order of $2L/v_r$. On the other hand, for a station is located near the far end of the rupture (F), the duration terms in L and in r_1 approximately cancel, r_2 is small, and the duration of the seismogram is short. At high frequencies, directivity shows up as a short, intense accelerogram at the far end of the fault, in contrast with a lower-amplitude, long-duration accelerogram near the origin of rupture. This is illustrated in **Figure 24**, where the Caleta de Campos record is near the epicenter while the La Union and Zihuatanejo stations are located in the forward directivity direction. Additional examples where directivity has been recognized at high frequencies are given in Boatwright and Boore (1982) and Frankel *et al.* (1986).

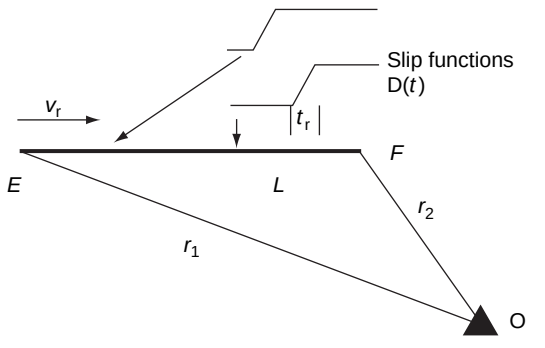


Figure 23 Geometry for consideration of the effect of directivity. This illustrates a vertical, strike-slip fault in map view, recorded at observation point O. The rupture is a Haskell propagating ramp with rupture velocity v_r and rise time τ. The shear velocity is β. Rupture is assumed to originate below an epicenter (point E), at one end of the fault, and propagate to the far end (point F). The duration of direct arrivals at O is given by the eqn [46].

Directivity is related to the isochrone velocity (e.g., Spudich and Fraser, 1984). The speed L/T_S is the mean isochrone velocity for this site. A high isochrone velocity results in energy from a large part of the fault arriving in a short time interval, as occurs near the far end of the fault in **Figure 23**. Conversely at a station that is located where the isochrone velocity is low the seismic energy will be spread over a longer time interval, resulting in lower amplitudes.

At low frequencies, directivity associated with propagating rupture causes constructive interference of long-period waves and coherent velocity pulses with large amplitudes. Such pulses have time-function characteristics that are correlated with their polarizations. Near a strike-slip fault, the fault-perpendicular component of ground velocity pulses observed in the region toward which the rupture is propagating are typically two-sided pulses (little net displacement), whereas the fault-parallel component of ground velocity observed near the fault as the rupture passes by are one-sided pulses having significant net displacement associated with the static fault offset, as mentioned in Section 4.18.4.2 This was shown by Archuleta and Hartzell (1981), Luco and Anderson (1983), Anderson and Luco (1983a, 1983b), and recently Somerville *et al.* (1997), who have developed empirical ground motion prediction relations that include this effect. Heaton *et al.* (1995) and Hall *et al.* (1995) summarized such pulses observed up to that time, which had peak velocities of up to 170 cm s^{-1}, and they showed that such pulses could be especially dangerous to buildings 10–25-stories tall. **Figure 25**, from Somerville *et al.* (1997), is selected to illustrate these phenomena. A feature of

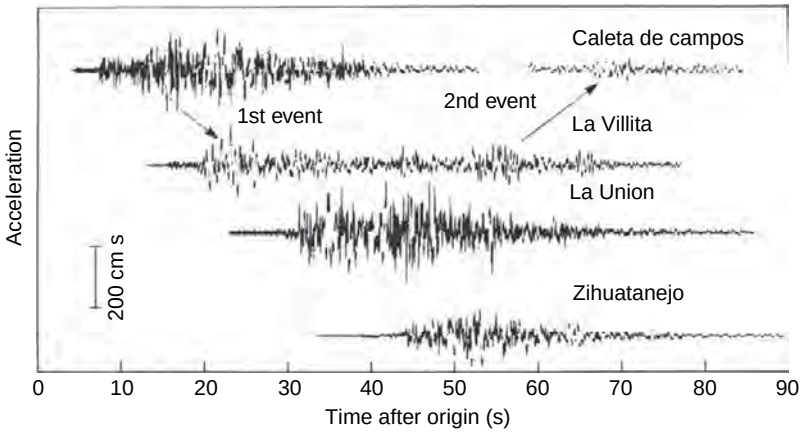


Figure 24 Seismograms from the 1985 Michoacan, Mexico earthquake ($M_W = 8.1$). Caleta de Campos is nearly above the hypocenter of this subduction-thrust earthquake, and the other stations are located on an approximately linear array in the direction of rupture. Thus, Caleta de Campos experienced backwards directivity. All four are above the rupture of the fault. The instrument at Caleta de Campos turned off and restarted after a 6 s gap. Note how the duration of strong shaking decreases at the stations in the direction of forward directivity. From Anderson JG, Bodin P, Brune J, Prince J, and Singh SK (1986) Strong ground motion and source mechanism of the Mexico earthquake of September 19, 1985. *Science* 233: 1043–1049.

the intense velocity pulse, as seen in **Figure 25**, is that it tends to be strongest on the component perpendicular to the fault. This result, which may initially be surprising, is predicted by dislocation modeling. It was first recognized and explained in the interpretation of accelerograms from the 1966 Parkfield, California, earthquake (Aki, 1967; Haskell, 1969).

4.18.8 Predicting Strong Ground Motions

4.18.8.1 Predicting Ground Motion Synthetic Seismograms

Combining the physics that is summarized in the previous sections, it is possible to generate synthetic ground motions. A vision for strong-motion seismology is that these synthetics should be sufficiently reliable to be used for engineering design of structures. Bolt and Abrahamson (2003) review this topic from the engineering perspective. Another important use is the development of scenarios for possible future earthquakes, to advance preparedness. Faccioli and Pessina (2003) have reviewed this topic from the viewpoint of this application.

The representation theorem, eqn [1], and its expression in the frequency domain (e.g., eqn [4]) provides the mathematical foundation for these approaches. Equation 1 requires two inputs: the Green's function and a representation of the seismic source. Thus it would seem that one way to organize this topic would be around these two requirements. In practice, that becomes difficult because many of the methods in current use have intertwined the two in ways that the individual contributions are not intuitively separable. While keeping the representation theorem in mind, this review organizes the alternatives in order of increasing complexity. In general, the more complex methods are more computer intensive, and the relationships of the components of the more complex models to the representation theorem become more explicit. At this more complex level, where the source and the Green's function are treated as independent modules, there are many possible combinations of source type and Green's function type that have not been tested.

The difficulty in reviewing the various methods is mild compared with the difficulty in evaluating these methods. The models can be evaluated by how well they predict statistical characteristics of the ground motion, or they can be evaluated by how well they are able to match the waveforms of observed strong motions in past earthquakes. To be useful for engineering applications, the statistical characteristics that should be predicted well include Fourier and response spectral amplitudes across the entire frequency band of interest (i.e., 0.1–20 Hz), peak velocity, and duration of shaking in different frequency bands (e.g., Abrahamson *et al.*, 1990;

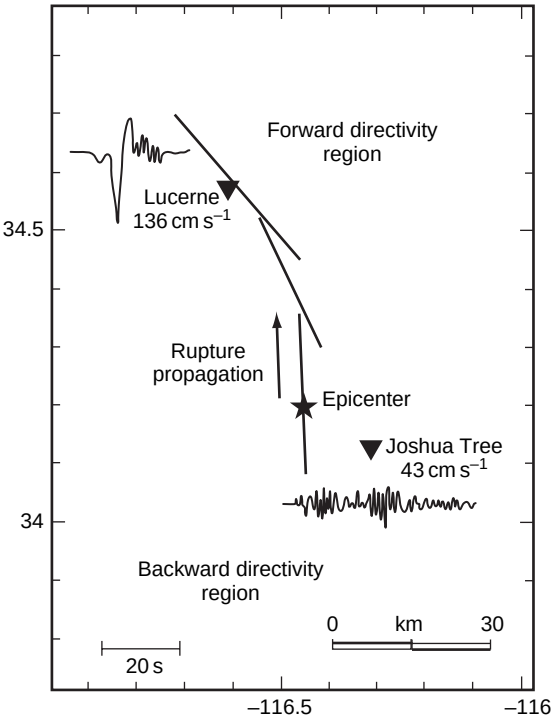


Figure 25 Somewhat schematic illustration of the origin of directivity in the 1992 Landers, California, earthquake ($M_W = 7.3$). The epicenter, identified by the solid star, was near the southern end of the fault. The strong-motion stations at Joshua Tree and Lucerne, identified by solid triangles, were the two nearest to the rupture. The Lucerne station experienced a coherent velocity pulse with duration under 20 s and peak velocity of 136 cm s^{-1}. The Joshua Tree seismogram, shown below the station, recorded farther from the fault and in the backwards directivity region, has a duration closer to 40 s and relatively low amplitudes peaking at 43 cm s^{-1}. The timescale for the seismograms is given at the bottom-left corner of the figure. From Somerville PG, Smith NF, Graves RW, Abrahamson NA (1997) Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismological Research Letters* 68(1): 199–222.

Schneider *et al.*, 1993; Hartzell *et al.*, 1999, 2005). Waveform matching is not expected for engineering applications, since the location of the hypocenter of future large earthquakes is not predictable. However, many but not all of the methods aim for waveform matches at low frequencies (see references in Section 4.18.8.1.3). Successful waveform matches demonstrate that the particular source model and Green's function combination are capable of predicting actual seismograms – an important feature when the source is a composite kinematic slip function that, while motivated by some physical model for the source, may have a tenuous relationship with physically realizable slip functions (e.g., Zeng and Chen, 2001). Anderson (2004) suggested a score that combines both statistical parameters and waveform matching, and as a means of calibrating what a good statistical match might be between a synthetic and observed seismogram, he shows statistics of how well the east and north components of a set of seismograms match each other.

Determining a statistic, or set of statistics, to compare data and predictions is only the smallest part of the problem with evaluating prediction methods. The greater challenge comes from calibrating new ideas and determining if they are an improvement over existing methods. To be thorough, a new source model, say, should be compared with an older source model using the same Green's functions, over the set of all of the well-recorded and/or most important strong motion records. Since a new source model requires calibration of some model parameters first, the computational task is enormous and very time consuming. Then in order to test whether it is reliable for engineering applications the new method, along with all of the existing methods, should be used to make blind prediction of new data whenever another strong earthquake occurs and is well recorded. The lack of an efficient system to carry out these functions is currently inhibiting progress in improving broadband synthetic seismogram models.

The main approaches can be grouped as methods that use stochastic simulations, that use empirical Green's functions, that use synthetic Green's functions with a kinematic source model, and that use synthetic Green's functions with a dynamic source model in a 3-D representation of the geological structure.

4.18.8.1.1 Stochastic simulations

One important approach to generating synthetic seismograms was proposed by Boore (1983). In this approach, one generates a completely random time series, applies a shape function in the time domain so that the time series has a reasonable envelope of amplitude with time, transforms to the frequency domain, modifies the Fourier amplitude spectrum to match the amplitude and shape of a spectral model, and then transforms back to the time domain to obtain the synthetic seismogram. The representation theorem is thus applied in the freqeuency domain in this approach (eqn [4]). Improvements to this approach are made thorough improved models of

the source spectrum (as discussed in Section 4.18.5.3) and the effects of wave propagation to make increasingly sophisticated spectral shapes. Atkinson and Boore (2000) provide a concise summary of why this approach is widely considered useful for engineering seismology applications.

Spectral models may start with point-source approximations, especially the Brune (1970) model, modified for attenuation along the path and at the site, and for site amplification (e.g., Boore, 1983; Boore and Atkinson, 1987; Boore and Joyner, 1997). Appropriate spectral shapes and the parameters that control them have received considerable attention, especially for eastern North America where observations of strong motions are not available for seismic hazard analyses (e.g., Boore, 1986; Boore and Atkinson, 1989, 1992; Boatwright and Choy, 1992; Atkinson 1993a, 1993b, 1996; Atkinson and Mereu, 1992; Haddon, 1996, 1997; Atkinson and Boore, 1998; Atkinson, 2004), but also for other regions (e.g., Atkinson, 1995; Sokolov, 1997; Margaris and Boore, 1998; Beresnev and Atkinson, 2002). With this calibration, the model has been used for engineering applications, especially to generate synthetic seismograms for a range of magnitudes and distances. The synthetic seismograms have been used as a basis for development of GMPEs (Atkinson, 1984; Boore and Atkinson, 1987; Atkinson and Somerville, 1994; Atkinson, 1995; Atkinson and Boore, 1995).

A finite-fault is a more realistic model for larger earthquakes, and causes a more complex spectral shape, with multiple corner frequencies (e.g., Haskell, 1964; Atkinson and Cassidy, 2000). Following the conceptual framework given in eqn [1], for larger events close to the fault, Silva *et al.* (1990) and Atkinson and Silva (1997) summed the contribution of subfaults, where each subfault contribution is treated as a smaller earthquake with energy release timed according to the rupture velocity. Like the point-source model, the stochastic finite-source model has been useful to help develop ground motion predictions (e.g., Atkinson and Silva, 2000;) and to interpret well-recorded earthquakes in several regions (e.g., Beresnev and Atkinson, 1998, 1999; Atkinson and Beresnev, 2002; Gregor *et al.*, 2002; Junn *et al.*, 2002; Roumelioti and Kiratzi, 2002; Roumelioti and Beresnev, 2003; Silva *et al.*, 2003; Benetatos and Kiratzi, 2004; Roumelioti *et al.*, 2004).

4.18.8.1.2 Empirical Green's functions

Section 4.18.4.1 showed how eqn [8] can be derived from eqn [1] to accommodate empirical Green's functions. Eqn [8] expressed the representation theorem in terms of empirical Green's functions. To use an empirical Green's function to generate a synthetic for a larger event, next it is necessary to define the number of times, N_k, that an empirical Green's function must be added to represent the slip in the kth element of the fault during the larger event.

$$N_k = \frac{M_{0k}(\boldsymbol{\xi}, \infty)}{M_{0k}^{\mathrm{EGF}}} \qquad [47]$$

Since the total slip in any fault element in the main shock is generally much larger than the slip in a small event, N_k is rounded to the nearest integer. A numerical procedure is needed to approximate the convolution in eqn [8]; it is not appropriate to simply multiply the EGF by N_k because the EGF has a very short rise time compared to the rise time of the slip on the fault element during the main shock. One approach is to define a set of N_k random delay times, τ_{mk}, according to a probability distribution based on the assumed slip model in the kth fault element. Then Equation 8 becomes:

$$u_n(\mathbf{x}, t) = \sum_{k=1}^{K} \sum_{m=1}^{N_k} u_n^{\mathrm{EGF}}(\mathbf{x}, t - \tau_{mk}; \boldsymbol{\xi}) \qquad [48]$$

If N_k is sufficiently large, this approach is no different from a convolution with the assumed slip function $[u_k(\boldsymbol{\xi}, \tau)]$ in the kth element, but if N_k is small the randomness in the realization of lag times is equivalent to introducing some high-frequency irregularity to the slip function, increasing the high-frequency spectral level in the resulting synthetic.

Several methods have been proposed to assure that the resulting synthetic seismograms match both the moment (which is guaranteed by eqn [48]) and the high-frequency spectral level of the target event (e.g., Irikura, 1983; Joyner and Boore, 1986; Wennerberg, 1990; Dan *et al.*, 1990; Kanamori *et al.*, 1993; Hutchings, 1991, 1994; Frankel, 1995; Ordaz *et al.*, 1995). These methods also deal with the issues that arise when the corner frequency of the EGF falls within the frequency range of interest. Matching the high-frequency spectral level puts constraints on the probability distribution allowed for τ_{mk} and on whether $u_n^{\mathrm{EGF}}(\mathbf{x}, t - \tau_{mk}; \boldsymbol{\xi})$ should be multiplied by a constant (that would also affect N_k) before the summation in eqn [48].

In a strict, formal sense, the need for records from the targeted station limits the applicability of empirical Green's functions. Most of the studies cited above

validate the EGF technique with records from smaller events recorded on the same instrument that recorded a major earthquake. Theoretically, at the future site of a critical engineering structure, a seismic station might be installed with the intention of recording empirical Green's functions specifically for that structure. Less formally, a nearby station on similar site conditions could also be used; the uncertainties introduced with this approximation have not been quantified.

The availability of small earthquakes with epicenters on the targeted fault also, in a formal sense, limits use of the EGF method. Fault segments that rupture in a major earthquake may have few small earthquakes in between the major events. It has long been remarked, for instance, that the segment of the San Andreas Fault in southern California that broke in 1857 has remarkably few small earthquakes located directly on the fault (e.g., Brune *et al.*, 1967; Joswig, 2001). Thus, the same EGF may need to be used for several of the fault elements. However, most of the studies cited above use individual foreshocks or aftershocks to reproduce the statistical character of observed strong motions. Reasons for the success are probably related to the relatively high influence of the near-surface geology on the ground motions, as found by Mori (1996).

Given the dependence of ground motions on focal mechanism and the direction from the fault to the station, lack of similarity of the focal mechanism of the main shock and the EGF is potentially limiting. Some studies have suggested that the S-wave radiation pattern breaks down at high frequencies, possibly but not necessarily due to strong scattering. Liu and Helmberger (1985) noted unusual complexity in the radiation pattern of a small earthquake in the Imperial Valley, California. Vidale (1989) found the focal mechanism of the 1987 Whittier Narrows, California, earthquake ($M_L = 5.9$) and its $M_L = 5.3$ aftershock, to be important in controlling the distribution of peak acceleration (at frequencies of 3–6 Hz). Satoh (2002) found a transition from amplitude controlled by the radiation pattern at 3 Hz to isotropic radiation at 6 Hz in the 1998 Miyagiken-Nanbu, Japan, earthquake ($M_W = 5.0$). Pulido and Kubo (2004) found the transition occurred between 1 Ha and 4 Hz in data from the 2000 Tottori, Japan, earthquake ($M_W = 6.8$). Considering that it is commonly possible to determine the focal mechanism of microearthquakes from high-frequency P-wave arrivals, these differing conclusions are consistent with the idea that scattering more effectively disrupts the mean radiation pattern of larger earthquakes at high frequencies, as later scattered arrivals become more important than direct arrivals in much of the wave train. To the extent that is true, the limitations placed on the use of EGF with a different focal mechanism might be muted when simulating ground motions for a large earthquake.

Finally, the empirical Green's functions need an adequate signal-to-noise ratio at low frequencies. The importance of this was described in Section 4.18.4.1.1.

4.18.8.1.3 Synthetic Green's functions with kinematic and dynamic sources

Several studies have used synthetic Green's functions to generate synthetic strong ground motion seismograms. The advantage of these methods is that the seismograms that are generated should be fully consistent with solutions to the wave equation for a distributed source, while the seismograms generated with the previous methods have generally lost information about the phase. Section 4.18.4.2 introduced the main concepts for computing synthetic Green's functions. When these are used, eqn [1] is employed directly as the framework for combining the source and the synthetic Green's functions. Issues in such applications include the choice and validation of the Green's function and, for predicting ground motions for an arbitrary fault, the source time function to be used. The use of the finite-difference or finite-element method (reviewed below) to generate Green's functions is preferred at low frequencies. Some studies in this category do not attempt to generate broadband synthetic seismograms, but rather are limited to the frequencies that can be achieved using the finite-difference solutions. Among the studies that do generate broadband seismograms, multiple approaches seem to work to generate the higher frequencies. For instance, Somerville *et al.* (1991) and Zeng *et al.* (1994) use of flat-layered Green's functions with a velocity model appropriate for the station, which Zeng *et al.* (1995) modified by a scattering model. Kamae *et al.* (1998), Hartzell *et al.* (1999, 2002), Pitarka *et al.* (2000), and Mena *et al.* (2006) use a hybrid approach, where the high frequencies are generated by a stochastic model similar to that described in the previous section.

A complete review of the finite-element and finite-difference approaches to the generation of synthetic seismograms is beyond the scope of this chapter. The methods themselves are referenced in the articles that report results. The papers by Vidale

et al. (1985) and Levander (1988) are among the earliest to use the finite-difference technique to generate synthetic seismograms in a complex geological structure. Several of the initial applications were to study the effects of a basin on the incoming seismic energy, and, with improving computer capabilities, the entire process from the source to the basin (e.g., Frankel and Vidale, 1992; Frankel, 1993; Graves, 1993; Yomogida and Etgen, 1993; Olsen *et al.*, 1995; Olsen and Archuleta, 1996; Inoue and Miyatake, 1997; Wald and Graves, 1998; Furumura and Koketsu, 1998; Pitarka *et al.*, 1998; Graves, 1998; Sato *et al.*, 1999; Satoh *et al.*, 2001; Furumura and Singh, 2002; Pitarka *et al.*, 2002; Furumura *et al.*, 2003; Yoshimura *et al.*, 2003; Olsen *et al.*, 2006).

At this level of sophistication, explicit kinematic models for the slip on the source are needed. Several approaches have been applied (e.g., Irikura, 1983; Tumarkin *et al.*, 1994; Tumarkin and Archuleta, 1994; Zeng *et al.*, 1994; Hartzell *et al.*, 1999, 2005). For one approach to describing the source, Somerville *et al.* (1999) estimated a scaling relation for the size of asperities, in which the asperities are about one-quarter of the total fault rupture area. Miyake *et al.* (2003) define the 'strong motion generation area' as the part of the fault which is the source of large slip velocities, large stress release (~10 MPa), and large amplitudes of high-frequency radiation (0.2–10 Hz). Similarly, Miyake *et al.* (2003) find that a scaling relation for the rise time obtained by Somerville *et al.* (1999) is consistent with the rise time of the strong-motion-generating area.

To give a sense of how these models might work, we consider briefly a different model, by, Zeng *et al.* (1994). Their kinematic 'composite source model' consists of a large number of subevents, motivated by Frankel (1991). The sizes of the subevents follow a power-law distribution, and each subevent is located at random on the fault. Rupture starts at a designated hypocenter and spreads at a designated rupture velocity. Each subevent radiates when the rupture front reaches its center. **Figure 26** shows an example of seismograms generated by this method, compared with observations from the 1994 Northridge, California earthquake (Anderson and Yu, 1996). **Figure 27** shows an example of the bias and standard deviation achieved by Hartzell *et al.* (1999) for a model of the 1994 Northridge, California ($M_W = 6.7$), ground motions.

The highest level of complexity in source models is the introduction of dynamic sources (e.g., Beroza and Mikumo, 1996; Miyatake, 2000; Nielsen and Olsen, 2000; Aagaard *et al.*, 2001; Oglesby and Day, 2002; Oglesby *et al.*, 2003, 2004; Guatteri *et al.*, 2003, 2004; Peyrat and Olsen, 2004; Duan and Oglesby, 2005; Olsen *et al.*, 2006). These synthetics draw on current models for the dynamics of slip on the fault (Chapter 4.02). **Figure 12** was generated by this approach. These source models do not generate high-frequency energy (e.g., above 1 Hz), but with careful definition of initial stress conditions they are capable of generating synthetic seismograms that match low frequencies of the observations. Once dynamic models are better characterized, it will be possible to develop efficient kinematic models that reproduce their characteristics, potentially to include high-frequency radiation.

Since the models described in this section follow the physics of the problem closely, one might expect that they would ultimately be the most general methods for ground motion predictions. In the hands of the developers, several of these models have been tuned to give good, unbiased predictions for selected earthquakes. However, the challenge of demonstrating their reliability, as outlined in the introduction to this section, has not yet been fully solved. These models generally have several free source parameters that are available to be adjusted, and blind testing by users who are independent of the model developers has not yet taken place. Their continued development is one of the most exciting frontiers of strong-motion seismology.

4.18.8.2 Ground Motion Prediction Equations

A longstanding goal in engineering seismology has been to develop simple equations that can be used to estimate the magnitude of various parameters associated with strong ground motions. Initially, most such equations attempted to predict peak acceleration as a function of magnitude and distance. With more and higher-quality digital strong-motion data, and efficient computer processing, recent efforts also predict peak velocity and response spectral amplitudes at a range of frequencies, and possibly additional parameters. As the ability improves to incorporate more physics into generation of synthetic seismograms, generation of synthetic seismograms may in part supersede the need for GMPEs. However, these equations will remain useful for the forseeable future for probabilistic seismic hazard analyses and for applications that do not require specialized studies.

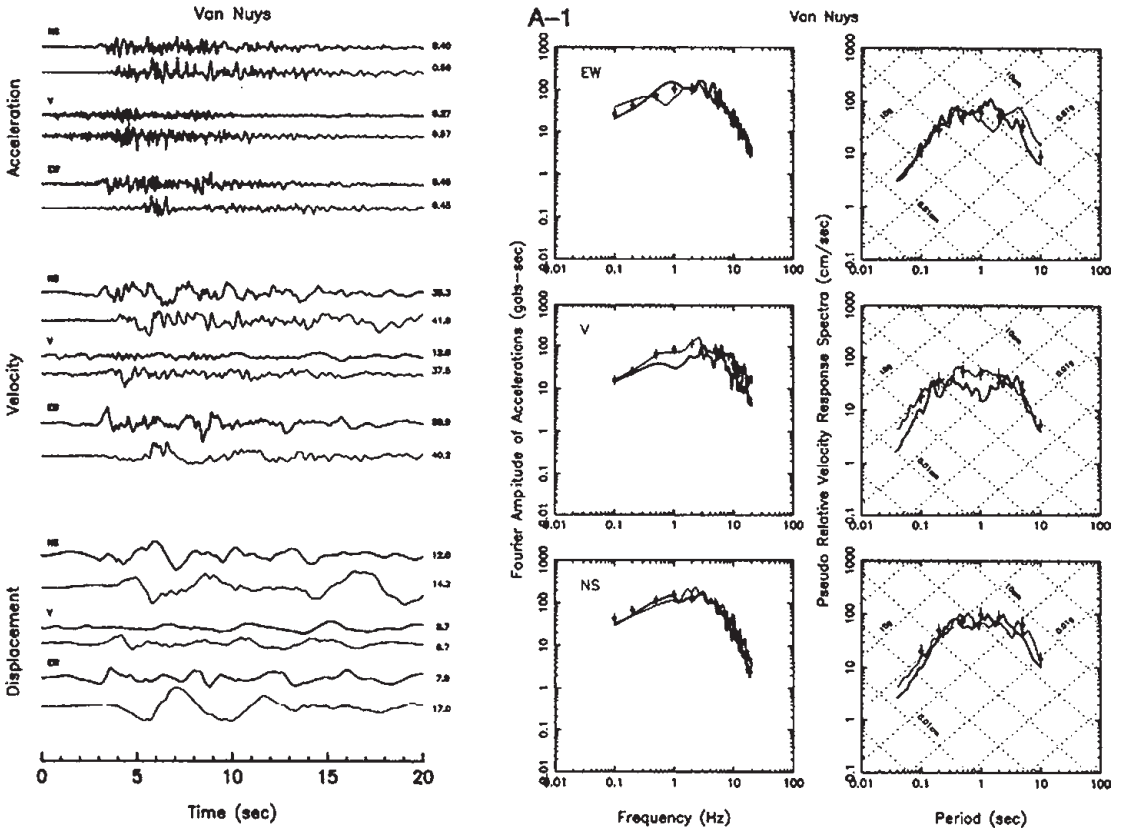


Figure 26 Observed and synthetic seismograms at the Van Nuys strong-motion station, from the 1994 Northridge earthquake, $M_W = 6.7$. Each record pair on the left shows data above and the synthetic below. In the plots of spectra, the heavy line is the data, the light line is one specific realization of the composite source model, and the points with error bars show the average of several realizations. From Anderson JG and Yu G (1996) Predictability of strong motions from the Northridge, California, earthquake. *Bulletin of the Seismological Society of America* 86: S100–S114.

This chapter refers to these equations as GMPEs. These are sometimes referred to as 'regressions' or 'regression equations', after the mathematical technique used to estimate the values of the coefficients in the equations. They are also sometimes referred to as 'attenuation equations', 'attenuation models' or 'attenuation curves'. Formally attenuation refers to the rate at which the amplitude of motion decreases with distance. The terminology 'ground motion prediction equations' is more descriptive in defining the purpose of the equations.

A recent review of GMPEs has been prepared by Campbell (2003). Bolt and Abrahamson (2003) include this topic also in their more extensive review of procedures for estimating strong motions. Several articles describing GMPEs are published in a single issue of *Seismological Research Letters* (Abrahamson and Shedlock, 1997; Atkinson and Boore, 1997a, 1997b; Toro *et al.*, 1997; Youngs *et al.*, 1997; Anderson, 1997; Abrahamson and Silva, 1997; Boore *et al.*, 1997; Campbell, 1997; Sadigh *et al.*, 1997; Spudich *et al.*, 1997; Somerville *et al.*, 1997), thus providing a handy reference to the subject. In the literature, there are large numbers, probably in the high hundreds if not the thousands, of GMPEs. One factor that accounts for the diversity is that whenever significant new data is recorded, it provides the opportunity to improve on past equations. The second factor is that, because of the differences among tectonic environments, many scientists have attempted to develop equations that apply to the region that they are studying, based on only the local data.

There are some common features to GMPEs. The first is that nearly all are written in the form

$$\ln Z = F_{GM}(M_W, r_{flt}, \ldots) + e_F + e_S \qquad [49]$$

where Z is the ground motion parameter to be predicted, $F_{GM}(.)$ is a function of several variables that will be discussed subsequently, e_F is a source term (i.e., the mean bias of motions for a particular

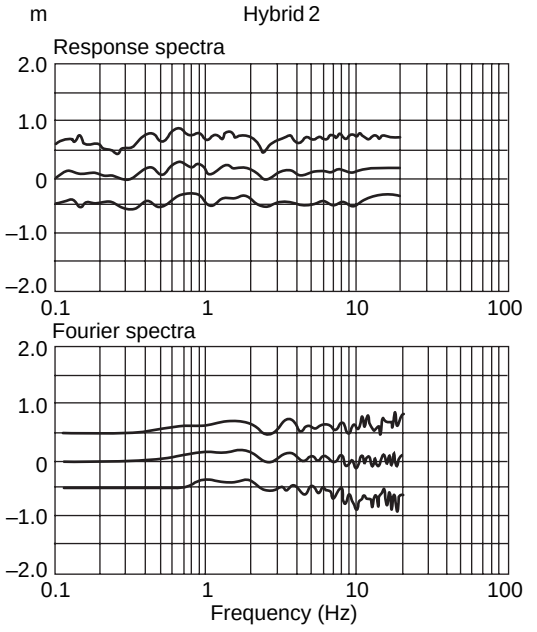


Figure 27 Bias with ±1 standard error of the simulation for the horizontal component of the acceleration response spectra (top) and Fourier spectra (bottom) of 29 near-field strong-motion stations that recorded the 1994 Northridge, California, earthquake. From Hartzell S, Harmsen S, Frankel A, and Larsen S (1999) Calculation of broadband time histories of ground motion: Comparison of methods and validation using strong-ground motion from the 1994 Northridge earthquake. *Bulletin of the Seismological Society of America* 89(6): 1484–1504.

earthquake), and e_S is the residual for the site. Typical parameters are peak acceleration, peak velocity, peak displacement, and acceleration response spectra with 5% damping for periods from 10 to 0.01 s (100 Hz). Because of the assymptotic properties of the response spectrum, at a frequency of 100 Hz, it is identical to peak acceleration.

Functional forms for $F_{GM}(M_W, r_{flt}, \ldots)$ in eqn [49] generally include a linear term in M_W, and often some additional magnitude dependence terms that allow Z to 'saturate', that is, reach a maximum value, as M_W increases. Some older models use a different magnitude scale instead of M_W, but M_W is now almost universally preferred. Also eqn [49] has replaced the hypocentral distance r with the distance to the nearest point on the extended fault, r_{flt}. While r_{flt} is not universally accepted as the best way to measure the distance to an extended fault, there is universal agreement that the distance to the hypocenter or epicenter is not appropriate. Noting that eqns [13] and [14] have distance terms in the form of r^{-1}, that taking the log of eqns [13] or [14] will result in terms of the form $-\ln r$, that wave propagation is more complex in a layered medium than in a homogeneous space, and that the finite fault size affects the rate of ground motion attenuation, many models have terms in the form $-b_r \ln r_{flt}$ setting the distance dependence to an unknown parameter that is estimated by regression.

Anderson and Quaas (1988) show a good example of how the amplitude of peak acceleration decreases much more slowly as distance increases for large-magnitude earthquakes than for small-magnitude earthquakes. Rogers and Perkins (1996) and Anderson (2000) present differing explanations for this phenomenon. Rogers and Perkins point out that when more parts of the fault contribute, the peak acceleration will follow the statistics of extreme values. Anderson points out that for the larger fault, with a long source duration, the late arrivals from parts of the fault that rupture first arrive at the same time as the earlier arrivals from later-rupturing parts of the fault. Both phenomena would contribute. Neither study, however, mentions nonlinear soil response, which can depress the amplitude of high-frequency motions near the source but not at large distances. When this occurs, the apparent attenuation rate is decreased compared to the situation where the ground response is linear, and since the larger earthquakes tend to cause greater nonlinearity, the shape of the curve is more affected. The more recent GMPE studies find the magnitude-dependent shape in the data.

Several additional variables are generally included. The site condition was often entered as a categorical variable in earlier models (e.g., Abrahamson and Silva, 1997). With improved data about the site conditions, new models now in preparation mostly use V_{S30}, thus aligning the equations better with the engineering application. Other variables, generally entered as categorical variables, include fault type (strike-slip, normal, or thrust) and an adjustment for the difference between horizontal and vertical motions. Another correction, introduced by Abrahamson and Silva, adjusts for whether the station is on the hanging wall of the fault.

The terms e_F and e_S are usually assumed to be normally distributed, with standard deviations of say σ_F and σ_S, respectively. For a random future earthquake, when the source term is not known, the total standard deviation is $\sigma_{GM} = \left(\sigma_F^2 + \sigma_S^2\right)^{1/2}$.

Figure 28 shows the predictions of the Abrahamson and Silva (1997) model. **Figure 29**

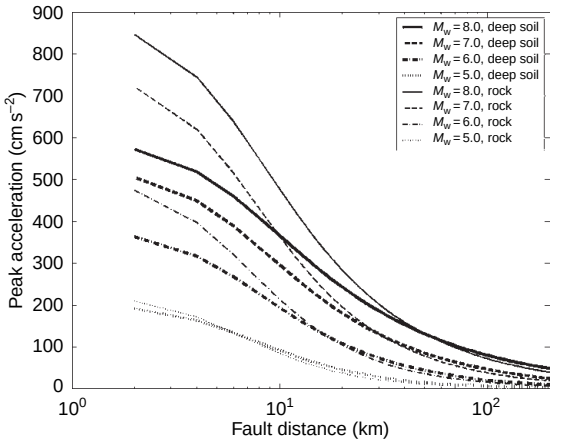


Figure 28 Median estimates of peak acceleration as a function of distance to the nearest point on the fault, according to Abrahamson and Silva (1997). These ground motion predictions are for strike-slip earthquakes in the continental crust, for moment magnitudes of 5–8, and for two different kinds of near-surface geology: deep soil and rock/shallow soil.

shows the standard deviation as a function of magnitude for four models for peak acceleration. **Figure 30** shows predicted acceleration response spectra for four magnitude events.

Figures 29 and **30** show total standard deviations in the prediction of ground motion in the range from 0.4 to 0.8 natural logarithm units. Since this is a lognormal distribution, a standard deviation of 0.69 corresponds to a multiplicative factor of two.

4.18.8.3 Future Directions for GMPEs

A major revision of GMPEs for the United States (and perhaps for earthquakes in shallow continental crust in general) is nearing completion. This is refered to as the NGA project, and is organized by PEER (Power *et al.*, 2006). This project introduces several innovations in the process of developing GMPEs.

One innovation in the process of developing the GMPE is the application of a team approach. The development of the database of ground motion observations is carried out as a separate project (Chiou *et al.*, 2006). This database has drawn on a data set of source models for major earthquakes (Mai *et al.*, 2006). Synthetic seismograms at low frequencies (Day) and high (e.g., Zeng, 2006) frequencies have been developed to provide theoretical guidance for the modeling as desired. Five teams of modelers were formed to draw upon this data. Each team generates its own GMPE, but the teams have interacted considerably. Finally, throughout the project, there have been frequent open meetings that provide for interaction among the NGA project team, scientists, and users (Power *et al.*, 2006).

One difficult issue has been the question of 'oversaturation' at high frequencies. Ground motions from some strong recent earthquakes (M_W 7.4 Izmit, Turkey (e.g., Anderson, 2000); M_W 7.6 Chi-Chi, Taiwan (e.g., Boore, 2001); and M_W 7.9 Denali, Alaska (e.g., Ellsworth *et al.*, 2004; Frankel, 2004) all gave peak accelerations that were about a factor of two smaller than those expected from the previous generation of GMPE, dominated by data from western North America. Optimizing the fit to this data can drive a GMPE to predict that peak acceleration is greatest for strong earthquakes ($M \sim 6.5$) and decreasing above that magnitude, a phenomenon called 'oversaturation'. If oversaturation is not allowed, the data from the largest magnitudes force the models to lower amplitudes than previous models for the moderate-sized earthquakes.

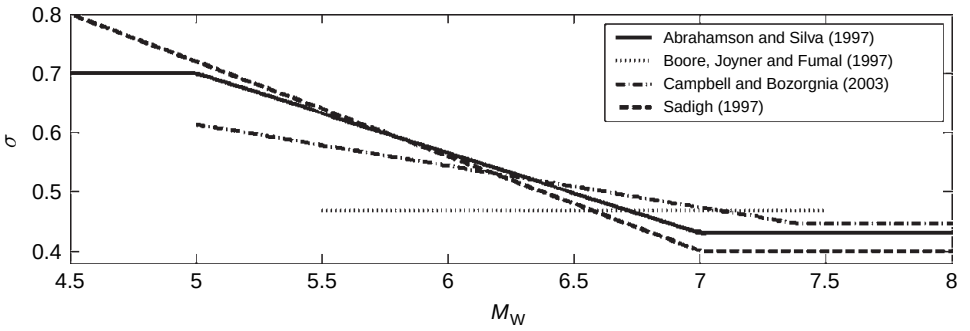


Figure 29 Standard deviation, σ_{GM} of the distribution of the data minus predictions from four GMPE models for natural log of peak acceleration, as a function of M_W. The GMPE shown here are those of Abrahamson and Silva (1997), Boore *et al.* (1997), Campbell and Bozorgnia (2003), and Sadigh *et al.* (1997).

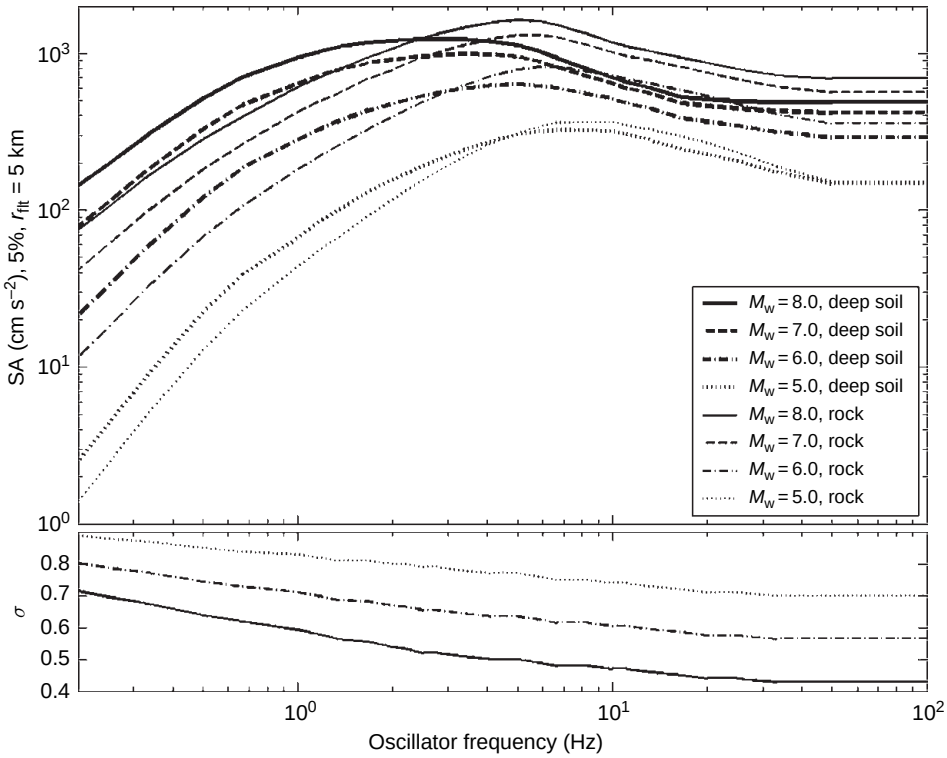


Figure 30 Ground motion predictions of Abrahamson and Silva (1997) for acceleration response spectra (SA) with 5% damping. The acceleration response at the oscillator frequency f_0 gives the peak response of a damped single-degree-of-freedom oscillator to the accelerogram. Because of its usefulness for approximating the response of a simple structure, SA is commonly used in engineering seismology instead of the Fourier spectrum (FS). At high frequencies, SA is assymptotic to peak acceleration, in contrast to the Fourier amplitude spectrum that falls off rapidly. At low frequencies, the assymptote is controlled by the peak displacement on the integrated accelerogram. At frequencies between those controlled by the assymptotes, SA is related to $(2\pi f_0)\mathrm{FS}(f_0)$ (Udwadia and Trifunac, 1973). At the bottom, the figure shows the standard deviation of the prediction equations, in natural logarithm units. In this model, the standard deviation is the same for both soil and rock sites, and the standard deviation for $M_W = 7$ is the same as for $M_W = 8$.

There are perhaps three alternative hypotheses to explain these low values. One (H1) is that they occur in different tectonic provinces from North America, so they are irrelevant for predicting motions in North America; the second (H2) is that even if that does not matter there are special characteristics of these three earthquakes that might not be typical; and the third (H3) is that they are typical of this magnitude of event.

Abrahamson and Shedlock (1997) proposed that earthquakes can be divided into three categories for the purposes of development or selection of an appropriate GMPE: shallow earthquakes in active tectonic regions, subduction zones, and stable continental interiors. Support for their viewpoint would be based on an appeal to fundamental physics. Considering that the chemical composition of the continental crust and the thermal structure of the crust are not that variable, it is difficult to see why the physics of slip on a fault or wave propagation should be different depending on the continent. The physical argument for separating stable continental interiors from active tectonic regions would be based on much older, cooler, thicker, and less-fractured crust in the stable regions. Thus the developer teams in NGA seemed to be rejecting H1. In support of H2, Anderson *et al.*, (2002) noted that all three of these large earthquakes occurred on faults with high slip rates. Based on Wesnousky (1988), faults with high slip rates generally have larger total slip, and show less irregularity at the surface than faults with low slip rates. Thus in support of H2, one might propose that when an earthquake in the M 7.5–8 range occurs on a fault with a lower slip rate, one should expect stronger ground motions, perhaps in the range of the extrapolations from the smaller

earthquakes (Anderson *et al.*, 2002). For the purposes of NGA, the data on total slip or slip rate was not sufficient to test this model. In support of H3, Wang *et al.* (2004) noted that five aftershocks of the Chi-Chi earthquake in the magnitude range from 5.8 to 6.3 were much more consistent with ground motions predicted from the North American GMPEs. One possibility to explain the discrepancy could be pervasive nonlinear soil response (Su *et al.*, 2006) which, if true, would be a universal feature of large earthquakes and might potentially cause oversaturation on soft soils.

This entire problem emphasizes the critical importance of continuing to gather strong-motion data from large earthquakes. Continental earthquakes with $M > 7$ are relatively rare events, as pointed out by Anderson *et al.* (2005a, 2005b), and every time one occurs outside of a strong-motion network, another opportunity to learn and improve our capability to make informed decisions on how to build to protect people from earthquakes has been lost.

The issue of uncertainty in GMPE, coupled with the issue of extreme ground motions, is a second challenge, although it only partly overlaps the scope of NGA. Civilian Radioactive Waste Management System Management and Operating Contractor (1998) and Stepp *et al.* (2001) present a probabilistic seismic hazard analysis for a site at Yucca Mountain, Nevada, which the United States has designated as the site for a high-level nuclear repository. When this model is extrapolated to the smallest annual probabilities of interest to the project, which are 10^{-8} per year, the model predicts peak accelerations of about 11*g*, and peak velocity of about $1300\,\mathrm{cm\,s^{-1}}$ on a Class A site condition. These values are much greater than those in **Figures 1** and **2**.

The issue of extreme ground motions is whether it is possible for motions at this level to occur in earthquakes, and if it is possible to place a practical upper bound on earthquake motions. The issue of uncertainty arises because these values are a direct outcome of the uncertainty in GMPEs, and if the uncertainty is overestimated or should be handled differently, then these values could turn out to be numerical artifacts in an evolving field of science. The issue is not only important for Yucca Mountain. The 1996 and 2002 national probabilistic seismic hazard maps for the United States (Frankel *et al.*, 2002) predicts levels of ground motions close to the San Andreas Fault which Brune (1996) has argued are not consistent with observations of precariously balanced rocks. Several subsequent studies have strengthened the scientific basis for the use of precarious rocks to test hazard curves at low probabilities (e.g., Shi *et al.*, 1996; Bell *et al.*, 1998; Anooshehpoor *et al.*, 1999, 2004; Anderson *et al.*, 2000; Brune, 2000, 2001, 2002, 2003; Anooshehpoor and Brune, 2002; Stirling *et al.*, 2002; Brune *et al.*, 2003, 2004, 2005, 2006; Purvance, 2005). Anderson and Brune (1999b) also found inconsistencies with the national map in Nevada. For the New Madrid region of the central United States, where the earthquake occurrence rate is very much smaller than in the active western part of the country, the national seismic hazard map also predicts rather large ground motions. Both of these results also might be the result of uncertainties in the GMPE. Anderson and Brune (1999a) suggest that the cause is an ergodic assumption in the probabilistic seismic hazard analysis methodology: the uncertainties in the GMPE are essentially spatial uncertainties, but in the PSHA the type of uncertainty that is needed is the uncertainty on the ground motion for an individual site. A final comment on uncertainty is that with the new data mentioned above, preliminary versions of some of the new relations were predicting ground motions at high magnitudes and short distances that were largely outside of the uncertainties of the previous generation of GMPE, raising the issue of how to estimate the uncertainty well. Resolution of these problems is an important future direction.

4.18.9 Conclusions

It is appropriate to return to the questions raised at the beginning of this chapter: is the physics of strong ground motion understood, and is it possible to estimate the amplitudes of ground motions well enough for engineering applications? From this review, it should be clear that the main elements of the physics are reasonably well understood. There is a major gap still in understanding the physics at the earthquake source, as there are large uncertainties in the physics of slip on the fault. The mechanism of high-frequency energy generation is not well understood. The three-dimensional regional velocity structures in most seismic regions are not known with the level of detail needed for finite-element modeling. The effects of nonlinear stress–strain curves introduce more uncertainties. Those uncertainties translate into uncertainties in prediction of ground motion near faults and at large magnitudes as well. It should be clear that calibration of all of the physical models

depends on gathering good data, and at small distances and for large magnitudes, there is still too little data to promote much confidence in estimates of the median or the variability of possible outcomes.

At larger distances (e.g., 20+km) and smaller magnitudes ($M < 6.5$), the database is larger and the uncertainties are smaller. In this distance range, the forces that need to be resisted by earthquakes can be estimated with an uncertainty of a factor of about two once the important faults have been identified, the magnitudes of earthquakes that can be caused by those faults have been estimated, and the site has been investigated sufficiently. Thus, while there is much to be done, it is safe to say that most earthquake disasters do not occur because of a lack of knowledge of what the ground motions will be. They occur because existing knowledge of probable ground motions, and existing knowledge of how to construct structures to resist those motions, is not being applied. There are certainly many legacy structures, but right now the world is in the midst of the largest burst of human population growth that it will ever have. Reasonable levels of ground motions to be resisted by new construction can be estimated in almost all cases. The only reasons this information is not applied are either related to priorities and politics (a failure to learn the nature of the inevitable future earthquakes and then invest in the modest incremental cost needed to build earthquake-resistant structures) or related to the extreme poverty that is far too prevalent.

Acknowledgment

New figures used strong-motion data from Guerrero that was gathered with partial support of the National Science Foundation under Grant CMS 0000050.

References

Aagaard BT, Heaton TH, and Hall JF (2001) Dynamic earthquake ruptures in the presence of lithostatic normal stresses: Implications for friction models and heat production. *Bulletin of the Seismological Society of America* 91: 1765–1796.

Abercrombie R and Mori J (1994) Local observations of the onset of a large earthquake - 28 June 1992 Landers, California. *Bulletin of the Seismological Society of America* 84(3): 725–734.

Abrahamson NA and Shedlock KM (1997) Overview. *Seismological Research Letters* 67: 9–23.

Abrahamson NA and Silva WJ (1997) Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismological Research Letters* 68: 94–127.

Abrahamson NA, Somerville PG, and Cornell CA (1990) Uncertainty in numerical strong motion predictions, *Proceedings of the Fourth U.S. National Conference on Earthquake Engineering*, vol. 1, 407–416.

Abrahamson NA, Schneider JF, and Stepp JC (1991) Empirical spatial coherency functions for application to soil-structure interaction analysis. *Earthquake Spectra* 7: 1–27.

Aguirre J and Irikura K (1997) Nonlinearity, liquefaction, and velocity variation of soft soil layers in Port Island, Kobe, during the Hyogoken Nanbu earthquake. *Bulletin of the Seismological Society of America* 87: 1244–1258.

Aki K (1967) Scaling law of seismic spectrum. *Journal of Geophysical Research* 72(4): 1217–1231.

Aki K (1969) Analysis of the seismic coda of local earthquakes as scattered waves. *Journal of Geophysical Research* 74: 615–631.

Aki K (1980) Attenuation of shear waves in the lithosphere for frequencies from 0.05 to 20 Hz. *Physics of the Earth and Planetary Interiors* 21: 50–60.

Aki K (1992) Scattering conversion P to S vs. S to P. *Bulletin of the Seismological Society of America* 82: 1969–1972.

Aki K and Chouet B (1975) Origin of coda waves: Source, attenuation and scattering effects. *Journal of Geophysical Research* 89: 3322–3342.

Aki K and Richards PG (1980) *Quantitative Seismology Theory and Methods*. San Francisco, CA: Freeman W H and Co.

Aki K and Richards PG (2002) *Quantitative Seismology,* 2nd edn. 700p. Sausalito, CA: University Science Books.

Akinci A, Malagnini L, Herrmann RB, Pino NA, Scognamiglio L, and Eyidogan H (2001) High-frequency ground motion in the Erzincan region, Turkey: Inferences from Small Earthquakes. *Bulletin of the Seismological Society of America* 91: 1446–1455.

Anderson JG (1986) Implication of attenuation for studies of the earthquake source. In: Ewing M (ed.) *Geophysical Monograph 37*, (*Maurice Ewing Series 6*): *Earthquake Source Mechanics*, pp. 311–318. Washington, DC: American Geophysical Union.

Anderson JG (1997) Nonparametric description of peak acceleration above a subduction thrust. *Seismological Research Letters* 67: 86–93.

Anderson JG (2000a) Expected shape of regressions for ground-motion parameters on rock. *Bulletin of the Seismological Society of America* 90: S43–S52.

Anderson JG (coordinator) (2000b) Implications for seismic hazard analysis. In: Kocaeli, Turkey, Earthquake of August 17, 1999 Reconnaissance Report, Youd TL, Bardet J-P, and Bray JD (eds.) Earthquake Spectra 16, Supplement A, 113–137.

Anderson JG (2003) Strong-motion seismology. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 937–965. New York: Academic Press.

Anderson JG (2004) Quantitative measure of the goodness-of-fit of synthetic seismograms. In Proceedings, 13th World Conference on Earthquake Engineering, Vancouver, British Columbia, August 1–6, 2004, Paper No. 243.

Anderson JG and Brune JN (1999a) Probabilistic seismic hazard analysis without the ergodic assumption. *Seismological Research Letters* 70: 19–28.

Anderson JG and Brune JN (1999b) Methodology for using precarious rocks in Nevada to test seismic hazard models. *Bulletin of the Seismological Society of America* 89: 456–467.

Anderson JG, Bodin P, Brune J, Prince J, and Singh SK (1986) Strong ground motion and source mechanism of the Mexico earthquake of September 19, 1985. *Science* 233: 1043–1049.

Anderson JG, Brune JN, Anooshehpoor R, and Ni S-D (2000) New ground motion data and concepts in seismic hazard analysis. *Current Science* 79(9): 1278–1290 (NOV 10 2000).

Anderson JG, Brune JN, Purvance M, and Anooshehpoor A (2005b) Data needs for improved probabilistic seismic hazard analysis. In: Gulkan P and Anderson JG (eds.) *NATO Science Series, IV. Earth and Environmental Sciences, vol. 58: Directions in Strong Motion Instrumentation*, pp. 1–24. Dordrecht: Springer.

Anderson JG, Brune JN, and Wesnousky SG (2002) Physical Phenomena Controlling High-Frequency Seismic Wave Generation In Earthquakes in Proceedings of the Seventh US National Conference of Earthquake Engineering, 21–25 July, Boston, Massachusetts, CD Rom, Paper # 00272, Earthquake Engineering Research Institute, Palo Alto, California.

Anderson JG and Hough S (1984) A Model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies. *Bulletin of the Seismological Society of America* 74: 1969–1994.

Anderson JG and Luco JE (1983a) Parametric study of near-field ground motion for a strike slip dislocation model. *Bulletin of the Seismological Society of America* 73: 23–43.

Anderson JG and Luco JE (1983b) Parametric study of near field ground motions for oblique-slip and dip-slip dislocation models. *Bulletin of the Seismological Society of America* 73: 45–57.

Anderson JG, Purvance M, Brune JN, and Anooshehpoor A (2005a) Developing Constraints on Extreme Ground Motions Based on Seismic and Geological Observations (extended abstract), 2nd International Workshop, Strong Ground Motion Prediction and Earthquake Tectonics in Urban Areas, 25–27 October, 2005, Earthquake Research Institute, University of Tokyo, Japan, 63–66.

Anderson JG and Quaas R (1988) Effect of magnitude on the character of strong ground motion: An example from the Guerrero, Mexico, strong motion network. *Earthquake Spectra* 4: 635–646.

Anderson JG, Sucuoglu H, Erbernik A, *et al.* (2000) Implications for seismic hazard analysis. In: Kocaeli, Turkey, Earthquake of August 17, 1999, Reconnaissance Report, Youd TL, Bardet JP, and Bray JD (eds.) Earthquake Spectra 16 (supplement A), 113–137.

Anderson JG, Wesnousky SG, and Stirling MW (1996) Earthquake size as a function of fault slip rate. *Bulletin of the Seismological Society of America* 86: 683–690.

Anderson JG and Yu G (1996) Predictability of strong motions from the Northridge, California, earthquake. *Bulletin of the Seismological Society of America* 86: S100–S114.

Anooshehpoor A and Brune JN (1989) Foam rubber modeling of topographic and dam interaction effects at Pacoima Dam. *Bulletin of the Seismological Society of America* 79(5): 1347–1360 (October 1989).

Anooshehpoor A and Brune JN (2002) Verification of precarious rock methodology using shake table tests of rock models. *Soil Dynamics and Earthquake Engineering Journal* 22: 917–922.

Anooshehpoor A, Brune JN, and Zeng Y (2004) Methodology for obtaining constraints on ground motion from precariously balanced rocks. *Bulletin of the Seismological Society of America* 94: 285–303.

Anooshehpoor A, Heaton TH, Shi BP, and Brune JN (1999) Estimates of the ground accelerations at Point Reyes Station during the 1906 San Francisco earthquake. *Bulletin of the Seismological Society of America* 89(4): 845–853 (AUG 1999).

Apsel RJ and Luco JE (1983) On the Green's functions for a layered half-space. Part II. *Bulletin of the Seismological Society of America* 73(4): 931–951 (Aug 1983).

Archuleta RJ and Hartzell SH (1981) Effects of fault finiteness on near-source ground motion. *Bulletin of the Seismological Society of America* 71: 939–957.

Archuleta RJ, Liu P, Steidl JH, Bonilla LF, Lavallée D, and Heuze F (2003) Finite-fault site-specific acceleration time histories that include nonlinear soil response. *Physics of the Earth and Planetary Interiors* 137: 153–181.

Archuleta R, Steidl J, and Squibb M (2004) Cosmos virtual data center: A web-based portal to strong motion data, 13th World Conference on Earthquake Engineering, Vancouver, BC (Canada), 1–6 Aug 2004, Venue West Conference Services Ltd, Vancouver, B.C, Canada. Report Number 1566.

Archuleta RJ, Ustodio S, and Liu P (2005) Resolving the source parameters of the Parkfield earthquake by multiple inversions of different data sets (extended abstract), 2nd International Workshop, Strong Ground Motion Prediction and Earthquake Tectonics in Urban Areas, 25–27 October, 2005, Earthquake Research Institute, University of Tokyo, Japan, 41–44.

Assimaki D and Gazetas G (2004) Soil and topographic amplification on canyon banks and the 1999 Athens earthquake. *Journal of Earthquake Engineering* 8(1): 1–43.

Assimaki D, Gazetas G, and Kausel E (2005) Effect of local soil conditions on the topographic aggravation of seismic motion: Parametric investigation and recorded field evidence from the 1999 Athens earthquake. *Bulletin of the Seismological Society of America* 95(3): 1059–1089.

Atkinson G (1993a) H/V ratio for hard-rock sites. *Bulletin of the Seismological Society of America* 83: 587–596.

Atkinson G (1993b) Source spectra for earthquakes in eastern North America. *Bulletin of the Seismological Society of America* 83: 1778–1798.

Atkinson GM (1984) Attenuation of strong ground motion in Canada from a random vibrations approach. *Bulletin of the Seismological Society of America* 74(6): 2629–2653 (December 1984).

Atkinson GM (1995) Attenuation and source parameters of earthquakes in the Cascadia region. *Bulletin of the Seismological Society of America* 85(5): 1327–1342 (October 1995).

Atkinson GM (1996) The high-frequency shape of the source spectrum for earthquakes in Eastern and Western Canada. *Bulletin of the Seismological Society of America* 86(1A): 106–112 (February 1996).

Atkinson GM (2004) Empirical attenuation of ground-motion spectral amplitudes in Southeastern Canada and the Northeastern United States. *Bulletin of the Seismological Society of America* 94(3): 1079–1095 (June 2004).

Atkinson GM and Beresnev IA (2002) Ground motions at Memphis and St. Louis from M 7.5–8.0 earthquakes in the New Madrid seismic zone. *Bulletin of the Seismological Society of America* 92(3): 1015–1024 (April 2002).

Atkinson G and Boore D (1995) New ground motion relations for eastern North America. *Bulletin of the Seismological Society of America* 85: 17–30.

Atkinson GM and Boore DM (1995) Ground motion relations for Eastern North America. *Bulletin of the Seismological Society of America* 85: 17–30.

Atkinson GM and Boore DM (1997a) Some comparisons between recent groung-motion relations. *Seismological Research Letters* 67: 24–40.

Atkinson GM and Boore DM (1997b) Stochastic point-source modeling of ground motions in the Cascadia region. *Seismological Research Letters* 67: 74–85.

Atkinson GM and Boore DM (2000) Reply to comment by Haddon RAW 'Evaluation of models for earthquake source spectra in Eastern North America' by Gail M. Atkinson and David M. Boore. *Bulletin of the Seismological Society of America* 90(5): 1339–1341 (October 2000).

Atkinson GM and Boore DM (1998) Evaluation of models for earthquake source spectra in Eastern North America. *Bulletin of the Seismological Society of America* 88: 917–934.

Atkinson G and Cassidy J (2000) Integrated use of seismograph and strong motion data to determine soil amplification in the Fraser Delta: Results from the Duvall and Georgia Strait earthquakes. *Bulletin of the Seismological Society of America* 90: 1028–1040.

Atkinson G and Mereu R (1992) The shape of ground motion attenuation curves in Southeastern Canada. *Bulletin of the Seismological Society of America* 82: 2014–2031.

Atkinson G and Somerville P (1994) Calibration of time history simulation methods. *Bulletin of the Seismological Society of America* 84: 400–414.

Atkinson GM and Silva W (1997) An empirical study of earthquake source spectra for California earthquakes. *Bulletin of the Seismological Society of America* 87: 97–113.

Atkinson GM and Walter S (2000) Stochastic modeling of California ground motions. *Bulletin of the Seismological Society of America* 90(2): 255–274 (April 2000).

Bao HS, Bielak J, Ghattas O, *et al.* (1998) Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers. *Computer Methods in Applied Mechanics and Engineering* 152(1–2): 85–102.

Bard P-Y and Bouchon M (1985) The two-dimensional resonance of sediment-filled valleys. *Bulletin of the Seismological Society of America* 75: 519–541.

Bard P-Y and Tucker BE (1985) Underground and ridge effects: Comparing observations with theory. *Bulletin of the Seismological Society of America* 75: 905–922.

Bay F, Fäh D, Malagnini L, and Giardini D (2003) Spectral shear-wave ground-motion scaling in Switzerland. *Bulletin of the Seismological Society of America* 93: 414–429.

Beeler NM, Wong T-F, and Hickman SH (2003) On the expected relationships among apparent stress, static stress drop, effective shear fracture energy, and efficiency. *Bulletin of the Seismological Society of America* 93: 1381–1389.

Bell JW, Brune JN, Liu TZ, Zreda M, and Yount JC (1998) Dating precariously balanced rocks in seismically active parts of California and Nevada. *Geology* 26(6): 495–498 (Jun 1998).

Benites R and Olsen KB (2005) Modeling strong ground motion in the Wellington Metropolitan Area, New Zealand. *Bulletin of the Seismological Society of America* 95(6): 2180–2196.

Beresnev I and Atkinson G (1999) Generic finite-fault model for ground motion prediction in Eastern North America. *Bulletin of the Seismological Society of America* 89: 608–625.

Beresnev IA and Atkinson GM (1998) Stochastic finite-fault modeling of ground motions from the 1994 Northridge, California, earthquake. Part I: Validation on rock sites. *Bulletin of the Seismological Society of America* 88(6): 1392–1401 (Dec 1998).

Beresnev IA and Atkinson GM (2002) Source parameters of earthquakes in Eastern and Western North America based on finite-fault modeling. *Bulletin of the Seismological Society of America* 92(2): 695–710 (March 2002).

Beresnev IA, Atldnson GM, Johnson PA, and Field EH (1998a) Stochastic finite-fault modeling of ground motions from the 1994 Northridge, California earthquake. Part II: Widespread nonlinear response at soil sites. 88: 1402–1410.

Beresnev IA, Field EH, Van Den Abeele K, and Johnson PA (1998b) Magnitude of nonlinear sediment response in Los Angeles basin during the 1994 Northridge, California, earthquake. *Bulletin of the Seismological Society of America* 88: 1079–1084.

Bernard P and Madariaga R (1984) A new assymptotic method for the modeling of near-field accelerograms. *Bulletin of the Seismological Society of America* 74: 539–557.

Beroza GC and Mikumo T (1996) Short slip duration in dynamic rupture in the presence of heterogeneous fault properties. *Journal of Geophysical Research* 101: 22449–22460.

Benetatos CA and Kiratzi AA (2004) Stochastic strong ground motion simulation of intermediate depth earthquakes: The cases of the 30 May 1990 Vrancea (Romania) and of the 22 January 2002 Karpathos Island (Greece) earthquakes. *Soil Dynamics and Earthquake Engineering* 24(1): 1–9 (Jan 2004).

Bindi D, Spallarossa D, Augliera P, and Cattaneo M (2001) Source parameters estimated from the aftershocks of the 1997 Umbria–Marche (Italy) seismic sequence. *Bulletin of the Seismological Society of America* 91: 448–455.

Boatwright J and Boore DM (1982) Analysis of the ground accelerations radiated by the 1980 Livermore Valley earthquakes for directivity and dynamic source characteristics. *Bulletin of the Seismological Society of America* 72(6): 1843–1865 (December 1982).

Boatwright J and Choy GL (1992) Acceleration source spectra anticipated for large earthquakes in Northeastern North America. *Bulletin of the Seismological Society of America* 82: 660–682.

Bolt B (1969) Duration of strong motion. *Proceedings of the 4th World Conference on Earthquake Engineering,* Santiago, Chile, 1304–1315.

Bolt BA and Abrahamson NA (2003) Estimation of strong seismic ground motions. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 983–1001. New York: Academic Press.

Bonilla LF, Steidl JH, Gariel J-C, and Archuleta RJ (2002) Borehole response studies at the Garner Valley Downhole Array, Southern California. *Bulletin of the Seismological Society of America* 92: 3165–3179.

Bonilla LF, Archuleta RJ, and Lavallée D (2005) Hysteretic and dilatant behavior of cohesionless soils and their effects on nonlinear site response: Field data observations and modeling. *Bulletin of the Seismological Society of America* 95(6): 2373–2395 (doi:10.1785/0120040128).

Boore DM (1983) Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bulletin of the Seismological Society of America* 73(6): Part A, 1865–1894 (Dec 1983).

Boore DM (1986) Short-period P- and S-wave radiation from large earthquakes: Implications for spectral scaling relations. *Bulletin of the Seismological Society of America* 76(1): 43–64 (February 1986).

Boore DM (2001) Comparisons of ground motions from the 1999 Chi–Chi Earthquake with empirical predictions largely based on data from California. *Bulletin of the Seismological Society of America* 91: 1212–1217.

Boore DM and Atkinson GM (1987) Stochastic prediction of ground motion and spectral response parameters at hard-rock sites in eastern North America. *Bulletin of the Seismological Society of America* 77: 440–467.

Boore D and Atkinson G (1989) Spectral scaling of the 1985–1988 Nahanni, Northwest Territories, earthquakes. *Bulletin of the Seismological Society of America* 79: 1736–1761.

Boore D and Atkinson G (1992) Source spectra for the 1988 Saguenay, Quebec earthquakes. *Bulletin of the Seismological Society of America* 82: 683–719.

Boore DM and Joyner WB (1997) Site amplifications for generic rock sites. *Bulletin of the Seismological Society of America* 87: 327–341.

Boore DM, Joyner WB, and Fumal TE (1997) Equations for estimating horizontal response spectra and peak accelerations from Western North American earthquakes: A summary of recent work. *Seismological Research Letters* 68: 128–153.

Boore DM, Stephens CD, and Joyner WB (2002) Comments on baseline correction of digital strong-motion data: Examples from the 1999 Hector Mine, California, Earthquake. *Bulletin of the Seismological Society of America* 92(4): 1543–1560 (May 2002).

Borcherdt RD and Glassmoyer G (1992) On the characteristics of local geology and their influence on ground motions

generated by the Loma Prieta earthquake in the San Francisco Bay region, California. *Bulletin of the Seismological Society of America* 82: 603–641.

Borcherdt RD and Glassmoyer GN (1994) Influences of local geology on strong and weak ground motions recorded in the San Francisco Bay region and their implications for site-specific building-code provisions, U. S. Geological Survey Professional Paper, Report: P 1551-A, pp. A77–A108.

Bouchon M and Barker JS (1996) Seismic response of a hill: The example of Tarzana, California. *Bulletin of the Seismological Society of America* 86: 66–72.

Bouchon M, Schulz CA, and Toksöz MN (1996) Effect of threedimensional topography on seismic motion. *Journal of Geophysical Research* 101: 5835–5846.

Braunmiller J, Nabelek J, Leitner B, and Qamar A (1995) The 1993 Klamath-Falls, Oregon, earthquake sequence - source mechanisms from regional data. *Geophysical Research Letters* 22(2): 105–108.

Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquake. *Journal of Geophysical Research* 75: 4997–5009.

Brune JN (1971) Correction. *Journal of Geophysical Research* 76: 5002.

Brune JN (1976) The physics of earthquake strong motion. In: Lomnitz C and Rosenblueth E (eds.) *Seismic Risk and Engineering Decisions*, pp. 141–177. New York: Elsevier.

Brune JN (1996) Precariously balanced rocks and ground-motion maps for southern California. *Bulletin of the Seismological Society of America* 86(1A): 43–54.

Brune JN (2000) Precarious rock evidence for low ground shaking on the footwall of major normal faults. *Bulletin of the Seismological Society of America* 90(4): 1107–1112.

Brune JN (2001) Shattered rock and precarious rock evidence for strong asymmetry in ground motions during thrust faulting. *Bulletin of the Seismological Society of America* 91(3): 441–447.

Brune JN (2002) Precarious-rock constraints on ground motion from historic and recent earthquakes in Southern California. *Bulletin of the Seismological Society of America* 92(7): 2602–2611.

Brune JN (2003) Precarious rock evidence for low near-source accelerations for trans-tensional strike-slip earthquakes. *Physics of the Earth and Planetary Interiors* 137(1-4): 229–239.

Brune JN and Allen CR (1967) A micro-earthquake survey of the San Andreas fault system in Southern California. *Bulletin of the Seismological Society of America* 57(2): 277–296.

Brune JN, Anooshehpoor A, Purvance MD, and Brune RJ (2006) Band of precariously balanced rocks between the Elsinore and San Jacinto, California, fault zones: Constraints on ground motion for large earthquakes. *Geology* 34(3): 137–140.

Brune JN, Anooshehpoor A, Shi BP, and Zeng Y (2004) Precarious rock and overturned transformer evidence for ground shaking in the M-s 7.7 Kern County earthquake: An analog for disastrous shaking from a major thrust fault in the Los Angeles basin. *Bulletin of the Seismological Society of America* 94(6): 1993–2003.

Brune JN, Smith KD, and Zeng YH (2005) Precarious rock evidence for seismic shaking during and prior to the 1992 M-L 5.6 Little Skull Mountain, Nevada, earthquake. *Earthquake Spectra* 21(4): 967–985.

Brune JN, von Seggern D, and Anooshehpoor A (2003) Distribution of precarious rocks at the Nevada Test Site: Comparison with ground motion predictions from nuclear tests. *Journal of Geophysical Research – Solid Earth* 108(36), doi: 10.1029/2002JB002000.

Building Seismic Safety Council (1997) *NEHRP Recommended Provisions for Seismic Regulations for New Buildings and other Structures, 1997 Edition. Part 1: Provisions, (FEMA 302)*, 336 p. Washington, D.C: Building Seismic Safety Council.

Campbell (1997) Empirical near-source attenuation relationships for horizontal and vertical components of peak ground acceleration, peak ground velocity, and pseudo-absolute acceleration response spectra. *Seismological Research Letters* 67: 154–179.

Campbell KW (2003) Strong-motion attenuation equations. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 1003–1012. New York: Academic Press.

Campbell KW and Bozorgnia Y (2003) Updated near-source ground-motion (Attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra. *Bulletin of the Seismological Society of America* 93(1): 314–331.

Castro RR, Anderson JG, and Singh SK (1990) Site response, attenuation and source spectra of S-waves along the Guerrero, Mexico subduction zone. *Bulletin of the Seismological Society of America* 79: 1481–1503.

Chen S-Z and Atkinson GM (2002) Global comparisons of earthquake source spectra. *Bulletin of the Seismological Society of America* 92(3): 885–895.

Chi WC, Dreger D, and Kaverina A (2001) Finite-source modeling of the 1999 Taiwan (Chi–Chi) earthquake derived from a dense strong-motion network. *Bulletin of the Seismological Society of America* 91(5): 1144–1157.

Chin BH and Aki K (1991) Simultaneous study of source, path and site effects on strong ground motion during the 1989 Loma Prieta earthquake: A preliminary result on pervasive nonlinear site effects. *Bulletin of the Seismological Society of America* 81: 1859–1884.

Chiou B, Darragh R, and Silva W (2006) An overview of the NGA strong-motion dataset, Abstracts of the Centennial Meeting of the Seismological Society of America, 18–22 April, 2006, San Francisco, CA, USA, Abstract No. 783.

Christensen N (1989) Seismic properties of rocks. In: James DE (ed.) *The Encyclopedia of Solid Earth Geophysics*, pp. 1093–1098. New York: Van Nostrand Reinhold Company.

Civilian Radioactive Waste Management System Management and Operating Contractor (1998) Probabilistic seismic hazard analyses for fault displacement and vibratory ground motion at Yucca Mountain, Nevada, Wong IG and Stepp JC, report coordinators, US Department of Energy, Oakland, CA.

Cultrera GC, Boore DM, Joyner WB, and Dietel CM (1999) Nonlinear soil response in the vicinity of the Van Norman Complex following the 1994 Northridge, California, earthquake. *Bulletin of the Seismological Society of America* 89: 1214–1231.

Dan K, Watanabe T, Tanaka T, and Sato R (1990) Stability of earthquake ground motion synthesized by using different small-event records as empirical Green's functions. *Bulletin of the Seismological Society of America* 80: 1433–1455.

Davis PM, Rubinstein JL, Liu KH, Gao SS, and Knopoff L (2000) Northridge earthquake damage caused by geologic focusing of seismic waves. *Science* 289: 1746–1750.

Day SM, Bielak J, Dreger D, *et al.* (2006) Numerical simulation of basin effects on long-period ground motion, Proceedings of the 8th US National Conference on Earthquake Engineering, April 18–22, 2006, San Francisco, CA, USA, Paper No. 1857.

Der ZA, Marshall ME, O'donnell A, and Mcelfresh TW (1984) Spatial coherence structure and attenuation of the Lg phase, site effects, and the interpretation of the Lg coda. *Bulletin of the Seismological Society of America* 74(4): 1125–1147.

Dreger D (1997) The large aftershocks of the Northridge earthquake and their relationship to mainshock slip and

fault-zone complexity. *Bulletin of the Seismological Society of America* 87(5): 1259–1266.
Duan B and Oglesby DD (2005) The dynamics of thrust and normal faults over multiple earthquake cycles: Effects of dipping fault geometry. *Bulletin of the Seismological Society of America* 95: 1623–1636.
Elgamal AW (1991) Shear hysteretic elastoplastic earthquake response of soil systems. *Earthquake Engineering and Structural Dynamics* 20(4): 371–387.
Ellsworth WL, Celebi M, Evans JR, Jensen EG, Nyman DJ, and Spudich P (2004) Near-field ground motions of the 2002 Denali fault, Alaska, earthquake recorded at Pump Station 10. *Earthquake Spectra* 20: 597–616.
Erickson D, McNamara DE, and Benz HM (2004) Frequency-dependent *Lg Q* within the Continental United States. *Bulletin of the Seismological Society of America* 94: 1630–1643.
Faccioli E and Pessina V (2003) Use of engineering seismology tools in ground shaking scenarios. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 1031–1048. New York: Academic Press.
Frankel A (1991) High frequency spectral falloff of earthquakes, fractal dimension of complex rupture, b-value, and the scaling of stress on faults. *Journal of Geophysical Research* 96: 6291–6302.
Frankel A (1993) Three-dimensional simulations of ground motions in the San Bernardino Valley, California, for hypothetical earthquakes on the San Andreas fault. *Bulletin of the Seismological Society of America* 83: 1020–1041.
Frankel A (1994) Dense array recordings in the San Bernardino Valley of landers-big bear aftershocks: Basin surface waves, Moho reflections, and three-dimensional simulations. *Bulletin of the Seismological Society of America* 84: 613–624.
Frankel A (1995) Simulating strong motions of large earthquakes using recordings of small earthquakes: The Loma Prieta mainshock as a test case. *Bulletin of the Seismological Society of America* 85: 1144–1160.
Frankel A (2004) Rupture process of the M 7.9 Denali Fault, Alaska, Earthquake: Subevents, directivity, and scaling of high-frequency ground motions. *Bulletin of the Seismological Society of America* 94(6B): S234–S255.
Frankel A and Clayton RW (1984) A finite-difference simulation of wave propagation in two-dimensional random media. *Bulletin of the Seismological Society of America* 74: 2167–2186.
Frankel A, Carver D, Cranswick E, Bice T, Sell R, and Hanson S (2001) Observations of basin ground motions from a dense seismic array in San Jose, California. *Bulletin of the Seismological Society of America* 91: 1–12.
Frankel AD, Carver DL, and Williams RA (2002) Nonlinear and linear site response and basin effects in Seattle for the M 6.8 Nisqually, Washington, earthquake. *Bulletin of the Seismological Society of America* 92: 2090–2109.
Frankel A, Fletcher J, Vernon F, *et al.* (1986) Rupture characteristics and tomographic source imaging of the ML < 3 earthquakes near Anza, California. *Journal of Geophysical Research* 91: 2633–2650.
Frankel A, Hough S, Friberg P, and Busby R (1991) Observations of Loma Prieta aftershocks from a dense array in Sunnyvale, California. *Bulletin of the Seismological Society of America* 81: 1900–1922.
Frankel AD, Petersen MD, Mueller CS *et al.* (2002b) Documentation for the 2002 Update of the National Seismic Hazard Maps, Open File Report 02-420, US Geological Survey, Denver.
Frankel A and Stephenson W (2000) Three-dimensional simulations of ground motions in the Seattle region for earthquakes in the Seattle Fault Zone. *Bulletin of the Seismological Society of America* 90: 1251–1267.
Frankel A and Vidale J (1992) A 3-dimensional simulation of seismic-waves in the Santa-Clara Valley, California, from a Loma-Prieta aftershock. *Bulletin of the Seismological Society of America* 82(5): 2045–2074.
Frankel A and Wennerberg L (1987) Energy-flux model of seismic coda: Separation of scattering and intrinsic attenuation. *Bulletin of the Seismological Society of America* 77: 1223–1251.
Furumura T and Chen L (2005) Parallel simulation of strong ground motions during recent and historical damaging earthquakes in Tokyo, Japan. *Parallel Computing* 31(2): 149–165.
Furumura T and Koketsu K (1998) Specific distribution of ground motion during the 1995 Kobe earthquake and its generation mechanism 25: 785–788.
Furumura T and Singh SK (2002) Regional Wave Propagation from Mexican Subduction Zone Earthquakes: The attenuation functions for interplate and inslab events. *Bulletin of the Seismological Society of America* 92: 2110–2125.
Furumura T, Kennett BLN, and Koketsu K (2003) Visualization of 3D wave propagation from the 2000 Tottori-ken Seibu, Japan, earthquake: Observation and numerical simulation. *Bulletin of the Seismological Society of America* 93(2): 870–881 (online movies available for the 23 October 2004 Chuetsu earthquake ($M = 6.8$) at http://www.eri.u-tokyo.ac.jp/furumura/04Chuetsu/index.html and for the 16 August 2005 Miyagioki, Japan, earthquake ($M = 7.2$) at http://www.eri.u-tokyo.ac.jp/furumura/05Miyagi/).
Fukuyama E (1991) Analysis and interpretation of the heterogeneous rupture process. *Tectonophysics* 197(1): 1–17.
Gao S, Liu H, Davis PM, and Knopoff L (1996) Localized amplification of seismic waves and correlation with damage due to the Northridge earthquake: Evidence for focusing in Santa Monica. *Bulletin of the Seismological Society of America* 86: S209–S230.
Garcia D, Singh SK, Herraiz M, Pacheco JF, and Ordaz M (2004) Inslab earthquakes of Central Mexico: Q, source spectra, and stress drop. *Bulletin of the Seismological Society of America* 94(3): 789–802.
García D, Singh SK, Herráiz M, Ordaz M, and Pacheco JF (2005) Inslab earthquakes of Central Mexico: Peak ground-motion parameters and response spectra. *Bulletin of the Seismological Society of America* 95: 2272–2282.
Geli L, Bard PY, and Julien B (1988) The effect of topography on earthquake ground motion: A review and new results. *Bulletin of the Seismological Society of America* 78: 42–63.
Gil-Zepeda SA, Francisco L, Jorge A, *et al.* (2002) 3D Seismic response of the deep basement structure of the Granada Basin (southern Spain). *Bulletin of the Seismological Society of America* 92: 2163–2176.
Graves RW (1993) Modeling three-dimensional site response effects in the Marina district basin, San Francisco, California. *Bulletin of the Seismological Society of America* 83: 1042–1063.
Graves RW (1996) Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society of America* 86(4): 1091–1106.
Graves RW (1998) Three-dimensional finite-difference modeling of the San Andreas fault: Source parameterization and ground-motion levels. *Bulletin of the Seismological Society of America* 88: 881–897.
Graves RW and Wald DJ (2004) Observed and simulated ground motions in the San Bernardino Basin region for the Hector Mine, California, Earthquake. *Bulletin of the Seismological Society of America* 94: 131–146.
Graves RW, Pitarka A, and Somerville PG (1998) Ground motion amplification in the Santa Monica area: Effects of shallow

basin edge structures. *Bulletin of the Seismological Society of America* 88: 1224–1242.

Gregor NJ, Silva WJ, Wong IG, and Youngs RR (2002) Ground-Motion Attenuation Relationships for Cascadia Subduction Zone Megathrust Earthquakes Based on a Stochastic Finite-Fault Model. *Bulletin of the Seismological Society of America* 92(5): 1923–1932.

Guatteri M, Mai PM, Beroza GC, and Boatwright J (2003) Strong ground-motion prediction from stochastic-dynamic source models. *Bulletin of the Seismological Society of America* 93: 301–313.

Guatteri M, Mai PM, and Beroza GC (2004) A pseudo-dynamic approximation to dynamic rupture models for strong ground motion prediction. *Bulletin of the Seismological Society of America* 94: 2051–2063.

Haddon R (1996) Earthquake source spectra in Eastern North America. *Bulletin of the Seismological Society of America* 86: 1300–1313.

Haddon R (1997) Reply to comment on 'Earthquake sottrce spectra in Eastern North America' by Atkinson, Boore, and Boatwright. *Bulletin of the Seismological Society of America* 87: 1703–1708.

Hall JF, Heaton TH, Halling MW, and Wald DJ (1995) Near-source ground motion and its effects on flexible buildings. *Earthquake Spectra* 11(4): 569–605.

Hanks TC (1979) b-values and ∼0 -∼seismic source models: Implications for tectonic stress variations along active crustal fault zones and the estimation of high-frequency strong ground motion. *Journal of Geophysical Research* 84: 2235–2242.

Hanks TC (1982) f_{max}. *Bulletin of the Seismological Society of America* 72: 1867–1879.

Hanks TH and Bakun WH (2002) A bilinear source-scaling model for M-log A observations of continental earthquakes. *Bulletin of the Seismological Society of America* 92: 1841–1846.

Hanks TC and Johnson DA (1976) Geophysical assessment of peak accelerations. *Bulletin of the Seismological Society of America* 66(3): 959–968.

Hanks TC and Kanamori H (1979) A moment magnitude scale. *Journal of Geophysical Research* 84: 2348–2350.

Hardebeck JL and Hauksson E (1997) Static stress drop in the 1994 Northridge, California, aftershock sequence. *Bulletin of the Seismological Society of America* 87: 1495–1501.

Hartzell SH (1978) Earthquake aftershocks as Green's functions. *Geophysical Research Letters* 5: 1–4.

Hartzell S, Bonilla LF, and Williams RA (2004) Prediction of nonlinear soil effects. *Bulletin of the Seismological Society of America* 94(5): 1609–1629.

Hartzell S, Carver D, Cranswick E, and Frankel A (2000) Variability of site response in Seattle, Washington. *Bulletin of the Seismological Society of America* 90: 1237–1250.

Hartzell SH, Carver DL, and King KW (1994) Initial investigation of site and topographic effects at Robinwood ridge, California. *Bulletin of the Seismological Society of America* 84: 1336–1349.

Hartzell S, Cranswick E, Frankel A, Carver D, and Meremonte M (1997) Variability of site response in the Los Angeles urban area. *Bulletin of the Seismological Society of America* 87: 1377–1400.

Hartzell S, Harmsen S, Frankel A, and Larsen S (1999) Calculation of broadband time histories of ground motion: Comparison of methods and validation using strong-ground motion from the 1994 Northridge earthquake. *Bulletin of the Seismological Society of America* 89(6): 1484–1504.

Hartzell S, Leeds A, Frankel A, *et al.* (2002) Simulation of broadband ground motion including nonlinear soil effects for a magnitude 6.5 earthquake on the Seattle Fault, Seattle, Washington. *Bulletin of the Seismological Society of America* 92: 831–853.

Hartzell S, Guatteri M, Mai PM, Liu P-C, and Fisk M (2005) Calculation of broadband time histories of ground motion. Part II: Kinematic and dynamic modeling using theoretical Green's functions and comparison with the 1994 Northridge earthquake. *Bulletin of the Seismological Society of America* 95: 614–645.

Haskell NA (1953) The dispersion of surface waves on multilayered media. *Bulletin of the Seismological Society of America* 43: 17–34.

Haskell NA (1960) Crustal reflection of plane SH waves. *Journal of Geophysical Research* 65: 4147–4150.

Haskell NA (1962) Crustal reflection of plane P and SV waves. *Journal of Geophysical Research* 67: 4751–4767.

Haskell NA (1964) Total energy and energy spectral density of elastic wave radiation from propagating faults. *Bulletin of the Seismological Society of America* 54: 1811–1841.

Haskell NA (1969) Elastic displacements in the near-field of a propagating fault. *Bulletin of the Seismological Society of America* 59: 865–908.

Hayashi K, Burns DR, and Toksoz MN (2001) Discontinuous-grid finite-difference Seismic Modeling including surface topography. *Bulletin of the Seismological Society of America* 91(6): 1750–1764.

Heaton TH (1990) Evidence for and implications of self-healing pulses of slip in earthquake rupture. 64: 1–20.

Heaton TH, Hall JF, Wald DJ, and Halling MJ (1995) Response of high-rise and base-isolated buildings to a hypothetical M (sub W) 7.0 blind thrust earthquake. *Science* 267(5195): 206–211.

Higashi S and Sasatani T (2000) Nonlinear site response in Kushiro during the 1994 Hokkaido Toho-oki Earthquake. *Bulletin of the Seismological Society of America* 90(4): 1082–1095.

Hisada Y and Bielak J (2003) A theoretical method for computing near-fault ground motions in layered half-spaces considering static offset due to surface faulting, with a physical interpretation of fling step and rupture directivity. *Bulletin of the Seismological Society of America* 93(3): 1154–1168.

Horike M (1985) Inversion of phase velocity of long-period microtremors to the S-wave-velocity structure down to the basement in urbanized areas. 33: 59–96.

Hough SE and Anderson JG (1988) High-frequency spectra observed at Anza, California: Implications for Q structure. *Bulletin of the Seismological Society of America* 78: 692–707.

Hough SE and Dreger DS (1995) Source parameters of the 23 April 1992 M 6.1 Joshua Tree, California, earthquake and its aftershocks: Empirical green′s function analysis of GEOS and TERRAscope data. *Bulletin of the Seismological Society of America* 85(6): 1576–1590.

Hough SE and Field EH (1996) On the coherence of ground motion in the San Fernando Valley. *Bulletin of the Seismological Society of America* 86(6): 1724–1732.

Hough SE and Kanamori H (2002) Source properties of earthquakes near the Salton Sea triggered by the 16 October 1999 M 7.1 Hector Mine, California, earthquake. *Bulletin of the Seismological Society of America* 92(4): 1281–1289.

Hough SE, Seeber L, Lerner LA, Armbruster JC, and Guo H (1991) Empirical Greens-function analysis of Loma-Prieta aftershocks. *Bulletin of the Seismological Society of America* 81(5): 1737–1753.

Hough SEL, Seeber A, Rovelli L, *et al.* (1992) Ambient noise and weak-motion excitation of sediment resonances: Results from the Tiber Valley, Italy. *Bulletin of the Seismological Society of America* 82: 1186–1205.

Hudson DE (1979) *Reading and interpreting strong motion accelerograms*, 112p. Berkeley, CA: Earthquake Engineering Research Institute.

Humphrey JR, Jr. and Anderson JG (1992) Shear wave attenuation and site response in Guerrero, Mexico. *Bulletin of the Seismological Society of America* 82: 1622–1645.

Hutchings L (1991) 'Prediction' of strong ground motion for the 1989 Loma Prieta earthquake using empirical Green′s functions. *Bulletin of the Seismological Society of America* 81(5): 1813–1837.

Hutchings L (1994) Kinematic earthquake models and synthesized ground motion using empirical Green's functions. *Bulletin of the Seismological Society of America* 84: 1028–1050.

Hwang R-D, Wang J-H, Huang B-S, *et al.* (2001) Estimates of stress drop of the Chi–Chi, Taiwan, earthquake of 20 September 1999 from near-field seismograms. *Bulletin of the Seismological Society of America* 91: 1158–1166.

Inoue T and Miyatake T (1997) 3-D simulation of near-field strong ground motion: Basin edge effect derived from rupture directivity. *Geophysical Research Letters* 24: 905–908.

Irikura K (1983) Semi-empirical estimation of strong ground motions during large earthquakes. *Bulletin of the Disaster Prevention Research Institute* 33(2): 63–104.

Jaeger JC and Cook NGW (1976) *Fundamentals of Rock Mechanics,* 2nd edn, 585p. London: Chapman and Hall.

Jennings PC (2003) An introduction to the earthquake response of structures. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 1097–1125. New York: Academic Press.

Jin A, Moya CA, and Ando M (2000) Simultaneous Determination of site responses and source parameters of small earthquakes along the Atotsugawa Fault Zone, central Japan. *Bulletin of the Seismological Society of America* 90: 1430–1445.

Jones LE and Helmberger DV (1998) Earthquake source parameters and fault kinematics in the Eastern California shear zone. *Bulletin of the Seismological Society of America* 88: 1337–1352.

Joswig M (2001) Mapping seismic quiescence in California. *Bulletin of the Seismological Society of America* 91(1): 64–81.

Joyner WB (1975) A method for calculating nonlinear seismic response in two-dimensions. *Bulletin of the Seismological Society of America* 65: 1337–1357.

Joyner WB (1984) A scaling law for the spectra of large earthquakes. *Bulletin of the Seismological Society of America* 74: 1167–1188.

Joyner WB and Boore DM (1986) On simulating large earthquakes by Green's-function addition of smaller earthquakes. In: Das S, Boatwright J, and Scholz CH (eds.) *Geophysical Monograph, vol. 37, Earthquake Source Mechanics*, pp. 267–274. Washington, DC: American Geophysical Union.

Joyner WB and Chen TF (1975) Calculation of nonlinear ground response in earthquakes. *Bulletin of the Seismological Society of America* 65: 1315–1336.

Junn J-G, Jo N-D, and Baag C-E (2002) Stochastic prediction of ground motions in southern Korea. *Geosciences Journal (Seoul)* 6(3): 203–214.

Kamae K, Irikura K, and Pitarka A (1998) A technique for simulating strong ground motion using hybrid Green's function. *Bulletin of the Seismological Society of America* 88: 357–367.

Kanai K (1983) *Engineering Seismology*, 251p. Tokyo: University of Tokyo Press.

Kanamori H (1977) The energy release in great earthquakes. *Journal of Geophysical Research* 82: 2981–2987.

Kanamori H and Allen CR (1986) Earthquake repeat time and average stress drop. In: Das S, Boatwright J, and Scholz CH (eds.) *Geophysical Monograph 37*: *Earthquake Source Mechanics*, pp. 227–235. Washington, DC: American Geophysical Union.

Kanamori H and Anderson DL (1975) Theoretical basis of some empirical relations in seismology. *Bulletin of the Seismological Society of America* 65: 1073–1095.

Kanamori H and Rivera L (2004) Static and dynamic scaling relations for earthquakes and their implications for rupture speed and stress drop. *Bulletin of the Seismological Society of America* 94(1): 314–319.

Kanamori H, Thio HK, Dreger D, Hauksson E, and Heaton T (1992) Initial investigation of the Landers, California, Earthquake of 28 June 1992 using terrascope. *Geophysical Research Letters* 19(22): 2267–2270.

Kanamori H, Jennings PC, Singh SK, and Astiz L (1993) Estimation of strong ground motions in Mexico City expected for large earthquakes in the Guerrero seismic gap. *Bulletin of the Seismological Society of America* 83(3): 811–829.

Kawase H (1996) The cause of the damage belt in Kobe; 'the basin-edge effect,' constructive interference of the direct S-wave with the basin-induced diffracted/Rayleigh waves. *Seismological Research Letters* 67(5): 25–34.

Kawase H (2003) Site effects on strong ground motions. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 1013–1030. New York: Academic Press.

Kawase H, Matsushima S, Graves RW, and Somerville PG (1998) Three-dimensional wave propagation analysis of simple two-dimensional basin structures with special reference to ''the basin-edge effect'', *Zishin (Journal of Seismological Society of Japan)* 50: 431–449 (in Japanese).

Kawase H, Satoh T, Iwata T, and Irikura K (1998) S-wave velocity structures in the San Fernando and Santa Monica areas. In *Proceedings of the 2nd International Symposium on Effects of Surface Geology on Seismic Motions*, Tokyo, Japan, vol. 2, 733–740.

Kinoshita S (2003) Kyoshin Net (K_NET), Japan. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 1049–1056. New York: Academic Press.

Kinoshita S and Ohike M (2002) Scaling relations of earthquakes that occurred in the upper part of the Philippine sea plate beneath the Kanto region, Japan, estimated by means of borehole recordings. *Bulletin of the Seismological Society of America* 92: 611–624.

Kramer SL (1996) *Geotechnical Earthquake Engineering*, 653p. Upper Saddle River, NJ: Prentice Hall.

Komatitsch D and Tromp J (1999) Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International* 139(3): 806–822.

Komatitsch D, Qinya L, Jeroen T, Peter S, Christiane S, and John HS (2004) Simulations of ground motion in the Los Angeles Basin based upon the Spectral-Element Method. *Bulletin of the Seismological Society of America* 94(1): 187–206.

Larsen (2002) Las Vegas Basin seismic response project: 3-D Finite-difference ground motion simulations. 83(47), Fall Meeting Supplement, Abstract S12B-1188, 2002.

Levander AR (1988) 4th-order finite-difference P-SV seismograms. *Geophysics* 53(11): 1425–1436.

Lee WHK, Kanamori H, and Jennings PC (eds.) (2003) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B. New York: Academic Press.

Liu H-L and Helmberger DV (1985) The 23:19 aftershock of the 15 October 1979 Imperial Valley earthquake: More evidence for an asperity. *Bulletin of the Seismological Society of America* 75: 689–708.

Louie JN (2001) Faster, better: Shear-wave velocity to 100 meters depth from refraction microtremor arrays. *Bulletin of the Seismological Society of America* 91: 347–364.

Luco JE and Anderson JG (1983) Steady state response of an elastic half-space to a moving dislocation of finite width. *Bulletin of the Seismological Society of America* 73: 1–22.

Luco JE and Apsel RJ (1983) On the Green's functions for a layered half-space. Part I: *Bulletin of the Seismological Society of America* 73(4): 909–929.

Ma K-F, Mori J, Lee SJ, and Yu SB (2001) Spatial and temporal distribution of slip for the 1999 Chi–Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America* 91(5): 1069–1087.

Mai P, Jonsson S, and Bethmann F (2006) SRCMOD - an online resource for finite-source rupture models of past earthquakes, Abstracts of the Centennial Meeting of the Seismological Society of America, April 18-22, 2006, San Francisco, CA, USA, Abstract No. 295.

Mai PM and Beroza GC (2000) Source scaling properties from finite-fault-rupture models. *Bulletin of the Seismological Society of America* 90(3): 604–615.

Malagnini L (1996) Velocity and attenuation structure of very shallow soils: Evidence for a frequency-dependent Q. *Bulletin of the Seismological Society of America* 86: 1471–1486.

Malagnini L and Herrmann RB (2000) Ground-motion scaling in the region of the 1997 Umbria–Marche earthquake (Italy). *Bulletin of the Seismological Society of America* 90: 1041–1051.

Malagnini L, Herrmann RB, Mercuri A, Opice S, Biella G, and de Franco R (1997) Shear-wave velocity structure of sediments from the inversion of explosion-induced Rayleigh waves: Comparison with cross-hole measurements. *Bulletin of the Seismological Society of America* 87: 1413–1421.

Malagnini L, Herrmann RB, and Koch K (2000a) Regional ground-motion scaling in Central Europe. *Bulletin of the Seismological Society of America* 90: 1052–1061.

Malagnini L, Herrmann RB, and Di Bona M (2000b) Ground-motion scaling in the Apennines (Italy). *Bulletin of the Seismological Society of America* 90: 1062–1081.

Malagnini L, Akinci A, Herrmann RB, Pino NA, and Scognamiglio L (2002) Characteristics of the ground motion in northeastern Italy. *Bulletin of the Seismological Society of America* 92: 2186–2204.

Malagnini L, Mayeda K, Akinci A, and Bragato PL (2004) Estimating absolute site effects. *Bulletin of the Seismological Society of America* 94: 1343–1352.

Margaris BN and Boore DM (1998) Determination of $\Delta\sigma$ and κ_0 from response spectra of large earthquakes in Greece. *Bulletin of the Seismological Society of America* 88: 170–182.

Margaris BN and Hatzidimitriou PM (2002) Source spectral scaling and stress release estimates using strong-motion records in Greece. *Bulletin of the Seismological Society of America* 92: 1040–1059.

Matsushima T and Okada H (1990) Determination of deep geological structures under urban areas using long-period microtremors. *ButsuriTansa* 43: 21–33.

Matsuzawa T, Takeo M, Ide S, *et al.* (2004) S-wave energy estimation of small-earthquakes in the western Nagano region, Japan. *Geophysical Research Letters* 31(3): L03602.

McGarr A and Fletcher JB (2003) Maximum slip in earthquake fault zones, apparent stress, and stick-slip friction. *Bulletin of the Seismological Society of America* 93: 2355–2362.

McGarr A, Green RWE, and Spottiswoode SM (1981) Strong ground motion of mine tremors: Some implications for near-source ground motion parameters. *Bulletin of the Seismological Society of America* 71: 295–319.

Mena B, Durukal E, and Erdik M (2006) Effectiveness of hybrid Green's function method in the simulation of near-field strong motion: An application to the 2004 Parkfield earthquake. *Bulletin of the Seismological Society of America* 96(4B): S183–S205 (doi:10.1785/0120050814).

Menke W and Chen R (1984) Numerical studies of the coda falloff rate of multiply scattered waves in randomly layered media. *Bulletin of the Seismological Society of America* 74: 1605–1614.

Menke W, Lerner-Lam AL, Dubendorff B, and Pacheco J (1990) Polarization and coherence of 5 to 30 hz seismic wave Fields at a hard-rock site and their relevance to velocity heterogeneities in the crust. *Bulletin of the Seismological Society of America* 80(2): 430–449.

Meremonte M, Frankel A, Cranswick E, Carver D, and Worley D (1996) Urban seismology—Northridge aftershocks recorded by multi-scale arrays of portable digital seismographs. *Bulletin of the Seismological Society of America* 86: 1350–1363.

Midorikawa S (1994) Case histories of upthrow of objects during earthquakes (in Japanese), *Jishin (Journal of the Seismological Society of Japan)* 47: 333–340.

Miyatake T (2000) Computer simulation of strong ground motion near a fault using dynamic fault rupture modeling: Spatial distribution of the peak ground velocity vectors. *Pure and Applied Geophysics* 157: 2063–2081.

Miyake H, Iwata T, and Irikura K (2003) Source characterization for broadband ground-motion simulation; kinematic heterogeneous source model and strong motion generation area. *Bulletin of the Seismological Society of America* 93(6): 2531–2545.

Miyake H, Iwata T, and Irikura K (2003) Source characterization for broadband ground-motion simulation: Kinematic heterogeneous source model and strong motion generation area. *Bulletin of the Seismological Society of America* 93(6): 2531–2545.

Mori J (1996) Rupture directivity and slip distribution of the M 4.3 foreshock to the 1992 Joshua Tree earthquake, Southern California. *Bulletin of the Seismological Society of America* 86(3): 805–810.

Nazarian S and Stokoe KH, II (1984) *In situ* shear wave velocities from spectral analysis of surface waves. In: *Proceedings of the World Conference on Earthquake Engineering*, vol. 8, San Francisco, July 21–28.

Nielsen SB and Olsen KB (2000) Constraints on stress and friction from dynamic rupture models of the 1994 Northridge, California, earthquake. *Pure and Applied Geophysics* 157: 2029–2046.

O'Connell DRH (1999) Replication of apparent nonlinear seismic response with linear wave propagation models. *Science* 283: 2045–2050.

Oglesby DD and Day SM (2002) Stochastic fault stress: Implications for fault dynamics and ground motion. *Bulletin of the Seismological Society of America* 92: 3006–3021.

Oglesby DD, Douglas DS, Harris R, Ratchkovski N, and Hasen R (2004) Inverse kinematic and forward dynamic models of the 2002 Denali Fault earthquake, Alaska. *Bulletin of the Seismological Society of America* 94: S214–S233.

Oglesby DD, Steven MD, Li Y-G, and Vidale JE (2003) The 1999 Hector Mine earthquake: The dynamics of a branched fault system. *Bulletin of the Seismological Society of America* 93: 2459–2476.

Oldham RD (1899) Report on the great earthquake of 12th June 1897. *Memoirs of the Geological Survey of India* 29: 379.

Olsen KB (2000) Site amplification in the Los Angeles Basin from three-dimensional modeling of ground motion. *Bulletin of the Seismological Society of America* 90: S77–S94.

Olsen KB, Archuleta RJ, and Matarese JR (1995) Three-dimensional simulation of a magnitude 7.75 earthquake on the San Andreas fault. *Science* 270: 1628–1632.

Olsen KB and Archuleta RJ (1996) Three-dimensional simulation of earthquakes on the Los Angeles fault system. *Bulletin of the Seismological Society of America* 86: 575–596.

Olsen KB, Day SM, and Bradley CR (2003) Estimation of *Q* for Long-Period (>2 sec) Waves in the Los Angeles Basin. *Bulletin of the Seismological Society of America* 93: 627–638.

Olsen KB, Day SM, Minster JB, *et al.* (2006) Strong shaking in Los Angeles expected from Southern San Andreas earthquake. *Geophysical Research Letters* 33: L07305 (doi:10.1029/2005GL025472).

Ohmachi T and Midorikawa S (1992) Ground-motion intensity inferred from upthrow of Boulders during the 1984 western Nagano prefecture, Japan, earthquake. *Bulletin of the Seismological Society of America* 82(1): 44–60.

Ohminato T and Chouet BA (1997) A free surface boundary condition for including 3D topgraphy in the finite-difference method. *Bulletin of the Seismological Society of America* 87: 494–515.

Ordaz M, Arboleda J, and Singh SK (1995) A scheme of random summation of an empirical Green's function to estimate ground motions from future large earthquakes. *Bulletin of the Seismological Society of America* 85(6): 1635–1647.

Pavlenko O and Irikura K (2002) Changes in shear moduli of liquefied and nonliquefied soils during the 1995 Kobe earthquake and its aftershocks at three vertical-array sites. *Bulletin of the Seismological Society of America* 92(5): 1952–1969.

Pavlenko O and Loh C-H (2005) Nonlinear identification of the soil response at Dahan downhole array site during the 1999 Chi–Chi earthquake. *Soil Dynamics and Earthquake Engineering* 25: 241–250.

Peyrat S and Olsen KB (2004) Nonlinear dynamic rupture inversion of the 2000 western Tottori, Japan, earthquake. *Geophysical Research Letters* 31: L05604.

Phillips WS and Aki K (1986) Site amplification of coda waves from local earthquake in central California. *Bulletin of the Seismological Society of America* 76: 627–648.

Pitarka A and Irikura K (1996) Basin structure effects on long-period strong motions in the San Fernando Valley and the Los Angeles Basin from the 1994 Northridge earthquake and an aftershock. *Bulletin of the Seismological Society of America* 86: S126–S137.

Pitarka AS, Takenaka H, and Suetsugu D (1994) Modeling strong motion in the Ashigara Valley for the 1990 Odawara, Japan, earthquake. *Bulletin of the Seismological Society of America* 84: 1327–1335.

Pitarka A, Irikura K, and Iwata T (1997) Modelling of ground motion in the Higashinada (Kobe) area for an aftershock of the 1995 January 17 Hyogo-ken Nanbu, Japan, earthquake. *Geophysical Journal International* 131(2): 231–239.

Pitarka A, Irikura K, Iwata T, and Sekiguchi H (1998) Three dimensional simulation of the near fault ground motion for the 1995 Hyogoken Nanbu (Kobe), Japan, earthquake. *Bulletin of the Seismological Society of America* 88: 428–440.

Pitarka A, Somerville P, Fukushima Y, Uetake T, and Irikura K (2000) Simulation of near-fault strong ground motion using hybrid Green's functions. *Bulletin of the Seismological Society of America* 90: 566–586.

Pitarka A, Somerville PG, Fukushima Y, and Uetake T (2002) Ground-motion attenuation from the 1995 Kobe earthquake based on simulations using the hybrid Green's function method. *Bulletin of the Seismological Society of America* 92: 1025–1031.

Pitarka A, Graves R, and Somerville P (2004) Validation of a 3D velocity model of the Puget sound region based on modeling ground motion from the 28 February 2001 Nisqually Earthquake. *Bulletin of the Seismological Society of America* 94: 1670–1689.

Plicka V and Zahradnik J (2002) The eGf method for dissimilar focal mechanisms: The Athens 1999 earthquake. *Tectonophysics* 359(1–2): 81–95.

Power M, Chiou B, Abrahamson N, and Roblee C (2006) The ''next generation of ground motion attenuation models'' (NGA) project: An overview, Proceedings of the 8th U.S. National Conference on Earthquake Engineering, April 18–22, 2006, San Francisco, CA, USA, Paper No. 2022.

Prejean SG and Ellsworth WL (2001) Observations of Earthquake Source Parameters at 2 km Depth in the Long Valley Caldera, Eastern California. *Bulletin of the Seismological Society of America* 91: 165–177.

Pulido N and Kubo T (2004) Near-fault strong motion complexity of the 2000 Tottori earthquake (Japan) from a broadband source asperity model. *Tectonophysics* 390(1–4): 177–192.

Purvance MD (2005) Overturning of slender blocks: Numerical investigation and application to precariously balanced rocks in Southern California, *PhD Dissertation*, University of Nevada, Reno.

Purvance MD and Anderson JG (2003) A comprehensive study of the observed spectral decay in strong-motion accelerations recorded in Guerrero, Mexico. *Bulletin of the Seismological Society of America* 93: 600–611.

Rautian TC and Khalturin VI (1978) The use of coda for determination of the earthquake source spectrum. *Bulletin of the Seismological Society of America* 68: 923–948.

Richards PG and Menke W (1983) The apparent attenuation of a scattering medium. *Bulletin of the Seismological Society of America* 73: 1005–1021.

Rogers AM and Perkins DM (1996) Monte Carlo simulation of peak-acceleration attenuation using a finite-fault uniform-patch model including isochrone and extremal characteristics. *Bulletin of the Seismological Society of America* 86: 79–92.

Roumelioti Z and Beresnev IA (2003) Stochastic finite-fault modeling of ground motions from the 1999 Chi–chi, Taiwan, earthquake; application to rock and soil sites with implications for nonlinear site response. *Bulletin of the Seismological Society of America* 93(4): 1691–1702.

Roumelioti Z, Dreger D, Kiratzi A, and Theodoulidis N (2003) Slip distribution of the 7 September 1999 Athens earthquake inferred from an empirical Green's function study. *Bulletin of the Seismological Society of America* 93(2): 775–782.

Roumelioti Z and Kiratzi A (2002) Stochastic simulation of strong-motion records from the 15 April 1979 (M 7.1) Montenegro earthquake. *Bulletin of the Seismological Society of America* 92(3): 1095–1101.

Roumelioti Z, Kiratzi A, and Theodulidis N (2004) Stochastic strong ground-motion simulation of the 7 September 1999 Athens (Greece) earthquake. *Bulletin of the Seismological Society of America* 94(3): 1036–1052.

Rubinstein JL and Beroza GC (2004) Evidence for widespread nonlinear strong ground motion in the *M*W 6.9 Loma Prieta earthquake. *Bulletin of the Seismological Society of America* 94(5): 1595–1608.

Sadigh K, Chang C-Y, Egan JA, Makdisi F, and Youngs RR (1997) Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismological Research Letters* 68: 180–189.

Sánchez-Sesma FJ and Campillo M (1991) Diffraction of *P*, *SV*, and Rayleigh waves by topographic features: A boundary integral formulation. *Bulletin of the Seismological Society of America* 81: 2234–2253.

Sato T, Helmberger DV, Somerville PG, Graves RW, and Saikia CK (1998a) Estimates of regional and local strong

motions during the great 1923 Kanto, Japan, earthquake (*Ms* 8.2). Part 1: Source estimation of a calibration event and modeling of wave propagation paths. *Bulletin of the Seismological Society of America* 88: 183–205.

Sato T, Graves RW, Somerville PG, and Kataoka S (1998b) Estimates of regional and local strong motions during the great 1923 Kanto, Japan, earthquake (*Ms* 8.2). Part 2: Forward simulation of seismograms using variable-slip rupture models and estimation of near-fault long-period ground motions. *Bulletin of the Seismological Society of America* 88: 206–227.

Sato T, Graves RW, and Somerville PG (1999) Three-dimensional finite-difference simulations of long-period strong motions in the Tokyo metropolitan area during the 1990 Odawara earthquake (*Mj* 5.1) and the Great 1923 Kanto earthquake (*Ms* 8.2) in Japan. *Bulletin of the Seismological Society of America* 89(3): 579–607.

Sato T, Kawase H, Matsui M, and Kataoka S (1991) Array measurement of high frequency microtremors for underground structure estimation. In *Proceedings of the 4th International Conference on Seismic Zonation*, Odawana, Japan, vol. 2, 409–416.

Satoh T (2002) Empirical frequency-dependent radiation pattern of the 1998 Miyagiken–Nanbu earthquake in Japan. *Bulletin of the Seismological Society of America* 92(3): 1032–1039.

Satoh T, Fushimi M, and Tatsumi Y (2001) Inversion of strain-dependent nonlinear characteristics of soils using weak and strong motions observed by borehole sites in Japan. *Bulletin of the Seismological Society of America* 91: 365–380.

Satoh T, Kawase H, Sato T, and Pitarka A (2001) Three-dimensional finite-difference waveform modeling of strong motions observed in the Sendai Basin, Japan. *Bulletin of the Seismological Society of America* 91(4): 812–825.

Satoh T, Sato T, and Kawase H (1995) Nonlinear behavior of soil sediments identified by using borehole records observed at the Ashigara valley, Japan. *Bulletin of the Seismological Society of America* 85: 1770–1789.

Savage JC (1972) Relation of corner frequency to fault dimensions. *Journal of Geophysical Research* 77: 3788–3795.

Schnabel PB, Lysmer J, and Seed HB (1972) SHAKE: A computer program for earthquake response analysis of horizontally layered sites, Report EERC 72-12, University of California, Berkely.

Schneider JF, Silva WJ, and Stark C (1993) Ground motion model for the 1989 M 6.9 Loma Prieta earthquake including effects of source, path, and site. *Earthquake Spectra* 9: 251–287.

Scholz CH, Aviles CA, and Wesnousky SG (1986) Scaling differences between large intraplate and interplate earthquakes. *Bulletin of the Seismological Society of America* 76: 65–70.

Scognamiglio L, Malagnini L, and Akinci A (2005) Ground-motion scaling in eastern Sicily, Italy. *Bulletin of the Seismological Society of America* 95: 568–578.

Seale SH and Archuleta RJ (1989) Site amplification and attenuation of strong ground motion. *Bulletin of the Seismological Society of America* 79: 1673–1696.

Sekiguchi H and Iwata T (2001) The source process of the 1999 Chi–Chi, Taiwan, earthquake in semi-long period (2–20s). *Annual Report on Active Fault and Paleoearthquake Researches* 1: 315–324.

Shakal AF, Huang MJ, and Graizer VM (2003) Strong-Motion Data Processing. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol 81B, pp. 967–981. Amsterdam: Academic Press.

Shi B, Anooshehpoor A, Zeng Y, and Brune JN (1996) Rocking and overturning of precariously balanced rocks by earthquakes. *Bulletin of the Seismological Society of America* 86: 1364–1371.

Shi J, Kim W-Y, and Richards PG (1998) The corner frequencies and stress drops of intraplate earthquakes in the Northeastern United States. *Bulletin of the Seismological Society of America* 88: 531–542.

Shin TC, Tsai YB, Yeh YT, Liu CC, and Wu YM (2003) Strong-motion instrumentation programs in Taiwan. In: Lee WHK, Kanamori H, Jennings PC, and Kiss linger (eds.) *International Handbook of Earthquake and Engineering Seismology*, vol. 81B, pp. 1057–1062. New York: Academic Press.

Silva W, Darragh R, Stark C, *et al.* (1990). A methodology to estimate design response spectra in the near-source region of large earthquakes using the Band-Limited- White-Noise ground motion model. In: Proceedings of Fourth US National Conference on Earthquake Engineering, vol. 1, 487–494.

Silva W, Ivan Wong, Timothy Siegel, Nick Gregor, Robert Darragh, and Richard Lee (2003) Ground motion and liquefaction simulation of the 1886 Charleston, South Carolina, earthquake. *Bulletin of the Seismological Society of America* 93(6): 2717–2736.

Somerville P, Sen M, and Cohee B (1991) Simulation of strong ground motions recorded during the 1985 Michoacan, Mexico and Valparaiso, Chile earthquakes. *Bulletin of the Seismological Society of America* 81(1): 1–27.

Somerville PG, Smith NF, Graves RW, and Abrahamson NA (1997) Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismological Research Letters* 68(1): 199–222.

Somerville P, Irikura K, Graves R, *et al.* (1999) Characterizing crustal earthquake slip models for the prediction of strong ground motion. *Seismological Research Letters* 70: 59–80.

Spudich P and Frazer LN (1984) Use of ray theory to calculate high-frequency radiation from earthquake sources having spatially variable rupture velocity and stress drop. *Bulletin of the Seismological Society of America* 74: 2061–2082.

Spudich P and Frazer LN (1987) Erratum. *Bulletin of the Seismological Society of America* 77: 2245.

Spudich P and Iida M (1993) The seismic coda, site effects, and scattering in alluvial basins studied using aftershocks of the 1986 North Palm Springs, California, earthquake as source arrays. *Bulletin of the Seismological Society of America* 83: 1721–1743.

Spudich PA, Hellwig M, and Lee WHK (1996) Directional topographic response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake: Implications for mainshock motions. *Bulletin of the Seismological Society of America* 86(1B): S193–S208.

Spudich P, Fletcher JB, Hellwieg M, *et al.* (1997) SEA96- A new predictive relation for earthquake ground motions in extensional tectonic regimes. *Seismological Research Letters* 67: 190–198.

Steidl JH, Tumarkin AG, and Archuleta RJ (1996) What is a reference site? *Bulletin of the Seismological Society of America* 86: 1733–1748.

Stepp JC, Wong I, Whitney J, *et al.* (2001) Probabilistic seismic hazard analyses for fault displacement and ground motions at Yucca Mountain, Nevada. *Earthquake Spectra* 17: 113–152.

Stidham C, Antolik M, Dreger D, Larsen S, and Romanowicz B (1999) Three-dimensional structure influences on the strong-motion wavefield of the 1989 Loma Prieta earthquake. *Bulletin of the Seismological Society of America* 89: 1184–1202.

Stirling MW, Anooshehpoor A, Brune JN, Biasi GP, and Wesnousky SG (2002) Assessment of the site conditions of precariously balanced rocks in the Mojave Desert, Southern

California. *Bulletin of the Seismological Society of America* 92(6): 2139–2144.
Stork AL and Ito H (2004) Source parameter scaling for small earthquakes observed at the western Nagano 800-m-deep borehole, central Japan. *Bulletin of the Seismological Society of America* 94: 1781–1794.
Su F, John GA, and Yuehua Z (1998) Study of weak and strong ground motion including nonlinearity from the Northridge, California, earthquake sequence. *Bulletin of the Seismological Society of America* 88(6): 1411–1425.
Su F, Anderson JG, and Zeng Y (2006) Characteristics of ground motion response spectra from recent large earthquakes and their comparison with IEEE Standard 693, Proceedings of the 8th US National Conference on Earthquake Engineering, 18–22 April, 2006, San Francisco, California, USA, Paper No. 1360.
Takahashi R (1956) The SMAC. Strong motion accelerograph and other latest instruments for measuring earthquakes and building vibrations. Proceedings of the First World Conference in Earthquake Engineering, Berkeley, California 1956, 3: 1–11.
Takahashi T, Sato H, Ohtake M, and Obara K (2004) Scale dependence of apparent stress for earthquakes along the subducting pacific plate in northeastern Honshu, Japan. *Bulletin of the Seismological Society of America* 95: 1334–1345.
Takemura M, Motosaka M, and Yamanaka H (1995) Strong motion seismology in Japan. *Journal of Physics of the Earth* 43: 211–257.
Tooby P (2004) Terra Shake: simulating the big one on the San Andreas fault, Envision, Special Issue for SC2004, San Diego Supercomputer Center 4-7.
Toro GR, Abrahamson NA, and Schneider JF (1997) Model of strong ground motions from earthquakes in central and eastern North America: Best estimates and uncertainties. *Seismological Research Letters* 67: 41–57.
Trifunac MD and Hudson DE (1971) Analysis of the Pacoima Dam accelerogram, San Fernando, California, earthquake of 1971. *Bulletin of the Seismological Society of America* 61: 1393–1411.
Trifunac MD and Brady AG (1975) A study of the duration of strong earthquake ground motion. *Bulletin of the Seismological Society of America* 65: 581–626.
Trifunac MD and Todorovska MI (2001) Evolution of accelerographs, data processing, strong motion arrays and amplitude and spatial resolution in recording strong earthquake motion. *Soil Dynamics and Earthquake Engineering* 21: 537–555.
Tsuda K, Archuleta RJ, and Koketsu K (2006) Quantifying the spatial distribution of site response by use of the Yokohama high-density strong-motion network. *Bulletin of the Seismological Society of America* 96: 926–942.
Tsuda K, Steidl J, Archuleta R, and Assimaki D (2006) Site-response estimation for the 2003 Miyagi-Oki earthquake sequence considering nonlinear site response. *Bulletin of the Seismological Society of America* 96(4A): 1474–1482 (doi:10.1785/0120050160).
Tsujiura M (1978) Spectral analysis of the coda waves from local earthquakes. *Bulletin of Earthquake Research Institute, Tokyo University* 53: 1–48.
Tucker BE and King JL (1984) Dependence of sediment-filled valley response on input amplitude and valley properties. *Bulletin of the Seismological Society of America* 74: 153–165.
Tucker BE, King JL, Hatzfeld D, and Nersesov IL (1984) Observations of hard rock site effects. *Bulletin of the Seismological Society of America* 74: 121–136.
Tumarkin AG and Archuleta RJ (1994) Empirical ground motion prediction. *Annali di Geofisica* 37(6): 1691–1720.
Tumarkin AG, Archuleta RJ, and Madariaga R (1994) Scaling relations for composite earthquake models. *Bulletin of the Seismological Society of America* 84(4): 1279–1283.
Udwadia FE and Trifunac MD (1973) Damped fourier spectrum and response spectra. *Bulletin of the Seismological Society of America* 63: 1775–1783.
Utsu T (2002) Relationships between magnitude scales. In: *International Handbook of Earthquake and Engineering Seismology*, volume 81A, pp. 733–746. Int'l Assoc. Seismol. & Phys. Earth's Interior.
Velasco AA, Ammon CJ, Farrell J, and Pankow K (2004) Rupture directivity of the 3 November 2002 Denali fault earthquake determined from surface waves. *Bulletin of the Seismological Society of America* 94(6): S293–S299, Part B Suppl. S.
Venkataraman A, Rivera L, and Kanamori H (2002) Radiated energy from the 16 October 1999 Hector Mine earthquake: Regional and teleseismic estimates. *Bulletin of the Seismological Society of America* 92(4): 1256–1265.
Vernon FL, Fletcher J, Carroll L, Chave A, and Sembera E (1991) Coherence of seismic body waves from local events as measured by a small-aperture array. *Journal of Geophysical Research* 96: 11981–11996.
Vidale JE, Helmberger DH, and Clayton RW (1985) Finite-difference seismograms for *SH* waves. *Bulletin of the Seismological Society of America* 75: 1765–1782.
Vidale J (1989) Influence of focal mechanism on peak accelerations on strong motions of the Whittier Narrows, California, earthquake and an aftershock. *Journal of Geophysical Research* 94: 9607–9613.
Vladmir SY (1997) Empirical models for estimating Fourier-amplitude spectra of ground acceleration in the Northern Caucasus (Racha seismogenic zone). *Bulletin of the Seismological Society of America* 87(6): 1401–1412.
Wald DJ and Graves RW (1998) The seismic response of the Los Angeles Basin, California. *Bulletin of the Seismological Society of America* 88: 337–356.
Wang G-Q, Boore DM, Igel H, and Zhou X-Y (2004) Comparisons of ground motions from five aftershocks of the 1999 Chi–Chi, Taiwan, earthquake with empirical predictions largely based on data from California. *Bulletin of the Seismological Society of America* 94: 2198–2212.
Weisstein EW (2005) 'Cross-Correlation Theorem.' From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Cross-CorrelationTheorem.html, accessed Feb 2007.
Wells DL and Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America* 84: 974–1002.
Wennerberg LG (1990) Stochastic summation of empirical Green's functions. *Bulletin of the Seismological Society of America* 80(6): 1418–1432.
Wesnousky SG (1986) Earthquakes, quaternary faults, and seismic hazard in California. *Journal of Geophysical Research* 91: 12587–12631.
Wesnousky SG (1988) Seismological and structural evolution of strike-slip faults. *Nature* 335: 340–342.
Wu CJ, Takeo M, and Ide S (2001) Source process of the Chi–Chi earthquake: A joint inversion of strong motion data and global positioning system data with a multifault model. *Bulletin of the Seismological Society of America* 91(5): 1128–1143.
Xie JK, Liu ZY, Herrmann RB, and Cranswick E (1991) Source processes of 3 aftershocks of the 1983 Goodnow, New-York, earthquake - high-resolution images of small, symmetrical ruptures. *Bulletin of the Seismological Society of America* 81(3): 818–843.

Xu H, Day SM, and Minster J-BH (1999) Two-dimensional linear and nonlinear wave propagation in a half-space. *Bulletin of the Seismological Society of America* 89(4): 903–917.
Youngs RR, Chiou S-J, Silva WJ, and Humphrey JR (1997) Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismological Research Letters* 67: 58–73.
Yang J, Sato T, Savidis S, and Li XS (2002) Horizontal and vertical components of earthquake ground motions at liquefiable sites. *Soil Dynamics and Earthquake Engineering* 22: 229–240.
Yomogida K, Aki K, and Benites R (1997) Coda Q in two-layer random media. *Geophysical Journal International* 128(2): 425–433.
Yomogida K and Etgen JT (1993) 3-D wave propagation in the Los Angeles Basin for the Whittier–Narrows earthquake. *Bulletin of the Seismological Society of America* 83: 1325–1344.
Yoshimura C, Bielak J, Hisada Y, and Fernández A (2003) Domain reduction method for three-dimensional earthquake modeling in localized regions. Part II: Verification and applications. *Bulletin of the Seismological Society of America* 93: 825–841.
Yang J and Sato T (2000) Interpretation of seismic vertical amplification observed at an array site. *Bulletin of the Seismological Society of America* 90: 275–285.
Yang J, Sato T, Savidis S, and Li XS (2002) Horizontal and vertical components of earthquake ground motions at liquefiable sites. *Soil Dynamics and Earthquake Engineering* 22: 229–240.
Yu G, Anderson JG, and Siddharthan R (1993) On the characteristics of nonlinear soil response. *Bulletin of the Seismological Society of America* 83: 218–244.
Zeng Y (1991) Compact solutions of multiple scattered wave energy in time domain. *Bulletin of the Seismological Society of America* 81: 1022–1029.
Zeng Y (1993) Theory of scattered P- and S-wave energy in a random isotropic scattering medium. *Bulletin of the Seismological Society of America* 83: 1264–1276.
Zeng Y (2006) Broadband strong ground motion simulation using a composite source model, Proceedings of the 8th US National Conference on Earthquake Engineering, 18–22 April, 2006, San Francisco, CA, USA, Paper No. 1445.
Zeng Y, Anderson JG, and Su F (1995) Subevent rake and random scattering effects in realistic strong ground motion simulation. *Geophysical Research Letters* 22(1): 17–20.
Zeng Y, Anderson JG, and Yu G (1994) A Composite source model for computing realistic synthetic strong ground motions. *Geophysical Research Letters* 21(8): 725–728.
Zeng Y, Su F, and Aki K (1991) Scattered wave energy propagation in random isotropic scattering medium. Part I: Theory. *Journal of Geophysical Research* 96: 607–619.
Zeng YH and Chen CH (2001) Fault rupture process of the 20 September 1999 Chi–Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America* 91(5): 1088–1098.

Relevant Websites

http://26.cosmos-eq.org/ – COSMOS Virtual Data Center.
http://www.Seismo.ethz.ch – Database of Finite-Source Rupture Models.
http://www.globalcmt.org – Global CMT Catalog Search.
http://peer.berkeley.edu – Pacific Earthquake Engineering Research Center: NGA Database.
http://www.scec.org – Southern California Earthquake Center.
http://neic.usgs.gov – USGS: Earthquake Hazards Program.

4.19 Historical Seismicity – Paleoseismology

L. Grant, University of California, Irvine, CA, USA

4.19.1 Introduction

4.19.1.1 Paleoseismology and the Significance of Paleoearthquakes

Paleoseismology is the study of earthquakes, decades, centuries, or millennia after their occurrence (Yeats and Prentice, 1996). Paleoseismic studies typically focus on prehistoric or pre-instrumental earthquakes (Wallace, 1981; Sieh, 1981; McCalpin and Nelson, 1996) to supplement the historic record of seismicity. The time between surface ruptures of active faults is on the order of decades for the fastest-moving faults to thousands or tens of thousands of years for more numerous but less-active faults (Sibson, 2002). Modern scientific observation and analysis of earthquakes have occurred for a small fraction of this time, so there are few observations on the spatial and temporal characteristics of fault ruptures over multiple seismic cycles (Sieh, 1996). Paleoseismic investigations provide data on earthquake occurrence and the seismogenic behavior of individual faults that complement modern observations of earthquakes, with important implications for scientific models and practical applications for seismic hazard assessment (Reiter, 1995).

Moderate to large earthquakes may generate surface rupture or induce permanent changes in the landscape and local environment. Under suitable conditions, evidence of earthquakes is preserved in the geologic record. 'Paleoearthquakes' are recognized by detailed observations and analyses of geologic or environmental conditions within fault zones or in tectonically or seismically active regions. Paleoseismic investigations yield information about the recency of fault movement, dates of previous earthquakes, recurrence times, average slip rate, and earthquake effects over time intervals ranging from decades to thousands of years. Paleoseismic data are applied toward estimating the magnitudes of past earthquakes, forecasting the magnitudes of future earthquakes (WGCEP, 1988; Reiter, 1991), and identifying areas susceptible to fault rupture.

Paleoseismology is a subspecialty of seismology in that it focuses on earthquakes. However, the primary methods of paleoseismic data collection and analysis are drawn from geology and are distinctly different from analyses of seismic waves conducted by geophysicists. Therefore, paleoseismology is sometimes considered to be a subspecialty of active tectonics or earthquake geology. Paleoseismology began to emerge as a distinct discipline in the late 1960s to mid-1980s. Summaries, compendia, or review articles focusing on paleoseismology have been generated by Wallace (1981), Sieh (1981), Wallace (1986), Crone and Omdahl (1987), Vittori *et al.* (1991), Pantosti and Yeats (1993), Prentice *et al.* (1994), Serva and Slemmons (1995), Yeats and Prentice (1996), Pavlides *et al.* (1999), Grant (2002), and Grant and Lettis (2002). Yeats *et al.* (1997) provide a comprehensive text on the geology of earthquakes with global coverage of active tectonics and paleoseismic studies. For a comprehensive text and reference on paleoseismology, the reader is referred to McCalpin (1996).

4.19.1.2 Purpose and Scope of this Chapter

This is an updated and expanded revision of a review and tutorial on paleoseismology (Grant, 2002), first written in 2000 and published in 2002 as Chapter 30 in the *International Handbook of Earthquake and Engineering Seismology* (Lee *et al.*, 2002). Expansion and revision are warranted by the increasing importance of paleoseismology to the understanding of fault rupture and the assessment of seismic hazard, in addition to significant increases in the amount and quality of paleoseismic data worldwide. The purpose of this chapter is to summarize methods, contributions, and issues in paleoseismology. Topics include paleoseismic investigation methods, models, and applications. Issues in research and application are discussed in each section. Examples and references are chosen to highlight major concepts and findings rather than attempting to provide a comprehensive or geographically inclusive summary. Most examples are from western North America, and heavy emphasis is placed on studies of the San Andreas Fault system. Discussions of faulting and seismic hazard assessment are treated more fully by Sibson (2002) and Somerville and Moriwaki (2002).

4.19.2 Methods

With few exceptions, earthquakes are generated by the movement of faults (Bolt, 1999). Therefore, paleoseismic research is directly or indirectly a study of faults and their surface expression. Paleoseismic investigation methods employ techniques from several geologic subdisciplines, including stratigraphic analysis, Quaternary geology, soil science, geomorphology, engineering geology, geochronology, and structural geology. Major objectives of paleoseismic studies include identification of recently active or seismogenic faults, measurement of displacement (coseismic and long-term average), and establishing catalogs of paleoearthquakes. This section is organized by common goals of paleoseismic studies, rather than by disciplinary methods, because most investigations employ a variety of techniques.

4.19.2.1 Fault Identification

Because most earthquakes are generated by faults, the first step in most paleoseismic investigations is identification of a fault for study. An active fault is a fault that may have displacement within a future period of concern to humans (Wallace, 1981). Identification of active faults requires recognizing previous displacement and constraining the age of displacement. Therefore, an active fault is usually identified by associating it with tectonically deformed Quaternary-age materials or surfaces (e.g., Ziony and Yerkes, 1985).

4.19.2.1.1 Recognition of tectonic deformation

Most paleoseimic investigation methods focus on faults that cause recognizable tectonic deformation or environmental changes (e.g., submergence, erosion, anomalous deposition or mortality) at or near the Earth's surface. Faults may be classified by average slip rate (Sibson, 2002), type of displacement (strike slip, dip slip normal, dip-slip reverse, thrust, or oblique slip), and depth (surface or blind). The depth, rate, and type of fault slip determine the amount and nature of fault-induced deformation at the surface. The faster the slip rate, the more active the fault and the more likely the fault can be recognized and characterized by paleoseismologists (Slemmons and dePolo, 1986). Faults that reach the surface are easier to recognize than blind faults, but they may rupture to the surface infrequently if they have low slip rates (Sibson, 2002).

Paleoseismic data provide an incomplete record of earthquakes and fault rupture. Not all earthquakes leave recognizable evidence in the geologic record. Tectonic deformation due to fault slip may be difficult to recognize if rupture does not reach the surface, especially in intraplate regions (e.g., Lettis *et al.*, 1997). Earthquakes on blind faults may induce subtle surface deformations over broad areas (Stein and Yeats, 1989), subsurface sediment deformation (Dolan *et al.*, 2000), or distributed evidence of shaking in disrupted sediments (Reiter, 1995). The existence of an active blind fault can be detected by geomorphic analysis (e.g., Bullard and Lettis, 1993; Keller and Pinter, 1996), by mapping Quaternary sediments and identifying areas of uplift (e.g., Grant *et al.*, 1999, 2002), or by recognizing effects of shaking that are not associated with known surficial faults (Reiter, 1995; Brune, 2002).

Effects of tectonic deformation can be enhanced or obscured by anthropogenic or natural processes such as weathering and erosion. If tectonic changes cannot be readily distinguished from changes induced by other mechanisms, then the results of paleoseismic studies may be biased toward higher or lower estimates of tectonic displacement. Such problems have been noted in a variety of tectonic and human environments (e.g., Ricci, 1995; Obermeier, 1996; Nur, 2000; Zachariasen *et al.*, 1999).

4.19.2.1.2 Application of geochronology

Geochronology is dating of earth materials, surfaces, and processes. Geochronology is essential for paleoseismology because it constrains dates of paleoearthquakes and average rates of fault displacement. The most useful geochronologic methods for paleoseismic investigations yield high-resolution ages for common, Late Quaternary materials such as soils or buried flora and fauna (Lettis and Kelson, 2000). **Table 1** (from Noller *et al.* (2000a)) summarizes geochronologic methods for dating Quarternary materials in fault zones. A detailed discussion of this complex topic is beyond the scope of this chapter. A comprehensive compilation and summary of Quaternary geochronologic methods and applications is provided by Noller *et al.* (2000b). The most commonly used methods are also described by McCalpin (1996) and Yeats *et al.* (1997). Recent developments in cosmogenic surface exposure dating methods are beginning to be widely applied in paleoseismology to determine the age of offset features for slip-rate measurements (e.g., van der Woerd *et al.*, 2006).

Table 1 Classification of Quaternary geochronologic methods

Sidereal	Isotopic	Radiogenic	Chemical and biologic	Geomorphic	Correlation
Dendrochronology	Radiocarbon	Fission track	Amino-acid racemization	Soil profile development	Stratigraphy
	Uranium series	Thermoluminescence	Rock-varnish cation ratio	Rock-varnish development	Paleomagnetism
Varve chronology	^{210}Pb	Optically stimulated luminescence	Obsidian and tephra hydration	Scarp morphology and landform modification	Tephrochronology
	U–Pb Th–Pb				Paleontology
Historical records		Electron-spin resonance	Soil chemistry	Rate of deformation	Tectites and microtectites
	K–Ar and ^{39}Ar–^{40}Ar		^{10}Be accumulation in soils	Rate of deposition	Climate correlation
Sclerochronology and growth rings	Cosmogenic isotopes	Infrared stimulated luminescence	Lichenometry	Rock and mineral weathering	Astronomical correlation
					Stable isotopes
				Geomorphic position	Archeology

Source: Adapted from Noller JS, Sowers JM, Colman SM, and Pierce KL (2000a) Introduction to quaternary geochronology. In: Noller JS, Sowers JM, and Lettis WR (eds.) *AGU Reference Shelf Series 4*: *Quaternary Geochronology: Methods and Applications*, pp. 1–10. Washington, DC: American Geophysical Union.

Radiocarbon dating is the most widely used method for dating Holocene and latest Pleistocene earthquakes. The half-life of radioactive ^{14}C (5730 years) limits the application of radiocarbon dating to organic matter formed from carbon fixed within the last 50 000 to 60 000 years (Trumbore, 2000). The amount of ^{14}C in atmospheric CO_2 has varied in the past, particularly in the last few centuries due to anthropogenic emissions. To compensate for this variation, radiocarbon ages are calibrated to correspond to calendar ages (absolute ages). Calibrations based on tree rings and glacial varves extend back to the Early Holocene. Calibration curves are not linear. Plateaus in the calibration curves limit the precision of radiocarbon dating (Trumbore, 2000). This problem is acute for the last few centuries. For example, Yeats and Prentice (1996) note that the two largest historic ruptures of the San Andreas Fault in California, which occurred in 1857 and 1906, are indistinguishable using radiocarbon dating.

All methods listed in **Table 1** have limitations and uncertainties. Uncertainty in geochronologic methods is a major source of uncertainty in paleoseismic data (Lettis and Kelson, 2000). In addition to uncertainty in calibration and accuracy of analysis, errors may be introduced in the selection, collection, and interpretation of field samples. Therefore, most paleoseismic studies employ multiple methods of dating to reduce uncertainty and cross-check results. Methods such as radiocarbon dating that yield accurate, high-precision ages of common or widely distributed materials are preferred. Recent improvements in geochronology methods and in the statistical treatment of dates have reduced uncertainties in previously published ages (e.g., Sieh *et al.*, 1989; Biasi and Weldon, 1994; Biasi *et al.*, 2002).

4.19.2.2 Chronologies of Paleoearthquakes

Large earthquakes create features that may be preserved in the stratigraphic record and recognized by geological, geomorphic, environmental, or archeological analysis. Major effects of earthquakes include surface rupture, ground deformation, and ground failures due to shaking. Chronologies of earthquakes are developed by identifying and dating evidence of multiple paleoearthquakes along a specific fault, or by documenting evidence of shaking or tectonic deformation in a seismically active region. The longest chronology of paleoearthquakes spans 50 000 years, with a mean recurrence time of ~1600 years (Marco *et al.*, 1996).

Nonseismic processes can create features that appear very similar to those generated by earthquakes (Obermeier 1996; Ricci, 1995). This is especially true in coastal areas where vertical tectonic motion is difficult to discriminate from changes in sea level (Zachariasen *et al.*, 1999, 2000). The most commonly reported and least ambiguous paleoseismic data are derived from observations of surface ruptures, such as those from the North Anatolian Fault in Turkey and the San Andreas Fault in California, USA.

For most Quaternary faults, the dates of prehistoric earthquakes are not known. Well-studied faults such as the San Andreas Fault (Sieh *et al.*, 1989; Grant and Lettis, 2002; Weldon *et al.*, 2004, 2005; Grant *et al.*, 2005), and the North Anatolian Fault (e.g., Hartleb *et al.*, 2003, 2006) have chronologies of only a dozen or so paleoearthquakes. The North Anatolian Fault ruptured most recently in 1999, culminating a westward propagating sequence of eight $M \geq 7$ earthquakes since AD 1939 (Hartleb *et al.*, 2003). The historic record of earthquakes in Turkey spans approximately 2000 years in the western part of the country (Hartleb *et al.*, 2003; Klinger *et al.*, 2003) near the 1999 rupture. A 2500-year-long record of paleoearthquakes (Hartleb *et al.*, 2006) in north-central Turkey reveals evidence for the five most recent historic earthquakes and at least two prehistoric paleoearthquakes. Compilation of historic and paleoseismic data suggests that the North Anatolian Fault typically ruptures in large, infrequent earthquakes with inter-event times that range from 200 to 900 years, and that the entire length of the fault ruptured at least twice historically, in sequences similar to the 1939–1999 rupture pattern (Hartleb *et al.*, 2006). Repeated, similar ruptures along the western segment of the North Anatolian Fault support the concept of characteristic slip (Klinger *et al.*, 2003).

The San Andreas Fault in California, USA, has perhaps the best-documented record of paleoearthquakes of any fault in the world (Grant and Lettis, 2002). Descriptions of large-magnitude historic ruptures in AD 1906 and 1857 are supplemented by the record of paleoearthquakes at many sites along the ~1000-km-long fault. Dates of paleoearthquakes at selected sites on the fault are shown in **Table 2**. The longest and most detailed paleoearthquake chronology is from the Wrightwood site in southern California (Weldon *et al.*, 2005, 2004; Fumal *et al.*, 2002). The Wrightwood site has ruptured, on

Table 2 San Andreas Fault paleoearthquake event dates

Reference	Site	Event dates, (Years AD)
Schwartz *et al.* (1998)	Point Arena	1906 1650–1812
Kelson *et al.* (2006), in Grant and Lettis (2002)	Fort Ross	1906 1660–1812 1220–1380 1040–1190
Schwartz *et al.* (1998) Niemi and Hall (1992) Fumal *et al.* (2006)	Vedanta/Olema	1906 1430–1906 1430–1505 1276–1450 1015–1105
Knudsen *et al.* (2002), in Grant and Lettis (2002)	Bodega Harbor Bolinas Lagoon	1906 After 1600 980–1300 1906 ~1250
Schwartz *et al.* (1998) Hall (1986)	Dogtown	1906 1524–1906
Lienkaemper *et al.* (2002), in Grant and Lettis (2002)	Tyson's Lagoon, Hayward Fault	1868 1730 (1650–1790) 1630 (1530–1740) 1470 (1360–1580)
Fumal *et al.* (2006)	Arano Flat	1906 1700–1850 1600–1700 1500–1600 1430–1505 1300–1400 1200–1300 1100–1200 1015–1105
Stone *et al.* (2002)	LY4	1877–1881 1857 1291–1857 1058–1857
Sims (1994)	Phelan Creeks	1857 1405–1605 1323–1399 1221–1251 983–1023 803–873
Grant *et al.* (2005) Akciz *et al.* (2006)	Bidart Fan	1857 1540–1840 1540–1840 1370–1425 1285–1340 850–1050 650–850 250 BC–150 AD 250 BC–150 AD 1150 BC–950 BC
Lindvall *et al.* (2002), in Grant and Lettis (2002)	Frazier Mountain	1857 1460–1600
Biasi *et al.* (2002), in Grant and Lettis (2002)	Pallett Creek (after Sieh *et al.*, 1989.)	1857 1812 (1758–1837)

(*Continued*)

Table 2 (Continued)

Reference	Site	Event dates, (Years AD)
		1496–1599
		1343–1370
		1046–1113
		1031–1096
		914–986
		803–868
		749–775
		614–666
Weldon *et al.* (2004); Fumal *et al.* (2002)	Wrightwood	1857
		1812
		1662–1700
		1518–1542
		1463–1502
		1343–1370
		1230–1286
		1070–1152
		981–1039
		825–864
		758–794
		706–729
		676–708
		602–658
		464–594
Seitz *et al.* (2000)	Pitman Canyon	1812
		1690–1790
		1473–1573
		1428–1528
		1202–1302
		1040–1140
McGill *et al.* (2002), in Grant and Lettis (2002)	Plunge Creek	1510–1730
		~1450
Fumal *et al.* (2002), in Grant and Lettis (2002)	Thousand Palms	>1520–1680
		1450–1555
		1170–1290
		840–1150
		770–890
Sieh (1986)	Indio	1640–1720
		1300–1600
		1210–1390
		1000–1040
Williams and Seitz (2006)	Salt Creek	1640–1720
		1400–1700
		1100–1400
		800–1100

average, approximately once per century since AD 500, although the time interval between earthquakes has varied from approximately 31 to 165 years (Weldon *et al.*, 2004). Slip per event has varied from 0.7 to 6.6 m, in a complex pattern of strain accumulation and release (Weldon *et al.*, 2004). For the southern half of the San Andreas Fault, the existing paleoearthquake record can be fit by models of repeated long ruptures, randomly occurring ruptures of variable length, or alternating long and short ruptures (Weldon *et al.*, 2005).

4.19.2.2.1 Surface ruptures

Evidence of surface rupture can be observed directly in natural exposures or, more commonly, in trench excavations. Paleoseismologists look for trench sites with distinct evidence of multiple ruptures and abundant material for dating. Trenches are generally excavated

with a backhoe or similar excavation equipment. Shallow trenches in unconsolidated sediments may be dug by hand. Deep trenches must be stabilized with shoring, benching, or sloping. Stabilization methods are chosen to maximize wall exposure. The maximum practical depth for typical trench investigations is 5 m.

Trench exposures are cleaned by scraping and brushing to highlight stratigraphy and faults. Geologists map the exposures with the aid of a string grid or survey control. The maps are called 'trench log.' Trench logs are supplemented with descriptive notes of observations and objective records such as photographs. Samples of datable material are collected after their locations have been documented on the trench log. A good trench log documents observations of stratigraphy and structure as objectively as possible. However, it is impossible to make an entirely objective log because construction of a log is partly an act of interpretation.

4.19.2.2.1.(i) Recognition and dating of ruptures

The goal of logging is to recognize and record the number of surface ruptures and their stratigraphic position relative to dateble material so that each paleoearthquake can be analyzed and dated as accurately as possible. Evidence of surface rupture is commonly called a 'paleoseismic event'. The stratigraphic unit that was the ground surface at the time of the earthquake is referred to as the 'event horizon'. **Figure 1** shows typical stratigraphic and structural relationships that are used for recognizing and dating previous earthquakes (Lettis and Kelson, 2000). **Figure 2** shows an example of a trench log from the San Andreas Fault, with several paleoseismic events and event horizons marked with letters. The most useful criteria for identifying ruptures include multiple fault terminations at a single stratigraphic horizon, tilted and folded strata overlain by less-deformed strata, and colluvial wedges draped over fault scarps (Yeats and Prentice, 1996). Fault strands may die out and terminate below the surface, or be poorly expressed at the event horizon (Bonilla and Lienkaemper, 1990). Therefore, multiple exposures of a fault should be examined whenever possible to identify the event horizon with maximum confidence.

In some areas a fault cannot be examined with excavations because of a high water table, the presence of human structures, or burial by younger sediments. Subsurface investigation methods such as cone penetrometer testing (CPT) and large-diameter borings may be applied for these conditions (Grant *et al.*, 1997; Dolan *et al.*, 2000). These methods must be used in combination with other subsurface investigation methods such as seismic imaging, ground-penetrating radar, and standard borings. There are greater uncertainties in the identification and dating of paleoearthquakes from subsurface investigations than from surface exposures.

The chronology of surface ruptures revealed by paleoseismic methods is a subset of the actual earthquake history. The paleoseismic record is spatially and temporally incomplete because the conditions required for preservation of earthquakes are not present at all times on all active faults. The minimum earthquake magnitude associated with surface rupture is about M_L 5 (Bonilla, 1988), so $M_L < 5$ earthquakes are rarely recognized in trench exposures. Surface faulting is commonly associated with $M > 6$ earthquakes (Wells and Coppersmith, 1994). Therefore, most paleoseismic events recognized in trenches represent $M \geq 6$ earthquakes. Nor have all faults been studied by paleoseismologists. Early paleoseismic studies focused on examining evidence of larger earthquakes ($M > 7$) because the data were relatively easy to interpret. Therefore, the published paleoseismic record of earthquakes contains less data on the occurrence of smaller ($M < 7$) earthquakes (McCalpin, 1996).

4.19.2.2.2 Regional coseismic deformation

Earthquakes that generate regional deformation may be recognized and cataloged. Vertical deformation can induce changes in local rates of deposition and erosion that provide evidence of a paleoearthquake, particularly in fluvial (Schumm *et al.*, 2000) and coastal environments (Carver and McCalpin, 1996). Indicators of coseismic uplift in coastal areas include shorelines, platforms, or flights of terraces above mean sea level, and unusual growth or die-off patterns in coral microattols (Zachariasen *et al.*, 1999, 2000). Formation of elevated terraces or submerged shorelines have been reported following historic earthquakes in coastal regions (e.g., Plafker and Rubin, 1978; Matsuda *et al.*, 1978, Pirazzoli, 1991) such as the Great Ache-Andaman earthquake of 2004 off the coast of Sumatra. Sequences of similar terraces and shorelines at higher elevations provide evidence of multiple episodes of tectonic uplift (**Figure 3**). In shallow tropical waters, such as the coastal areas of Sumatra, the growth of corals responds to uplift or subsidence relative to sea level and may preserve a record of vertical coseismic motion (e.g., Zachariasen *et al.*, 2000; Natawidjaja *et al.*, 2006) and aseismic tectonic deformation (Natawidjaja *et al.*, 2004).

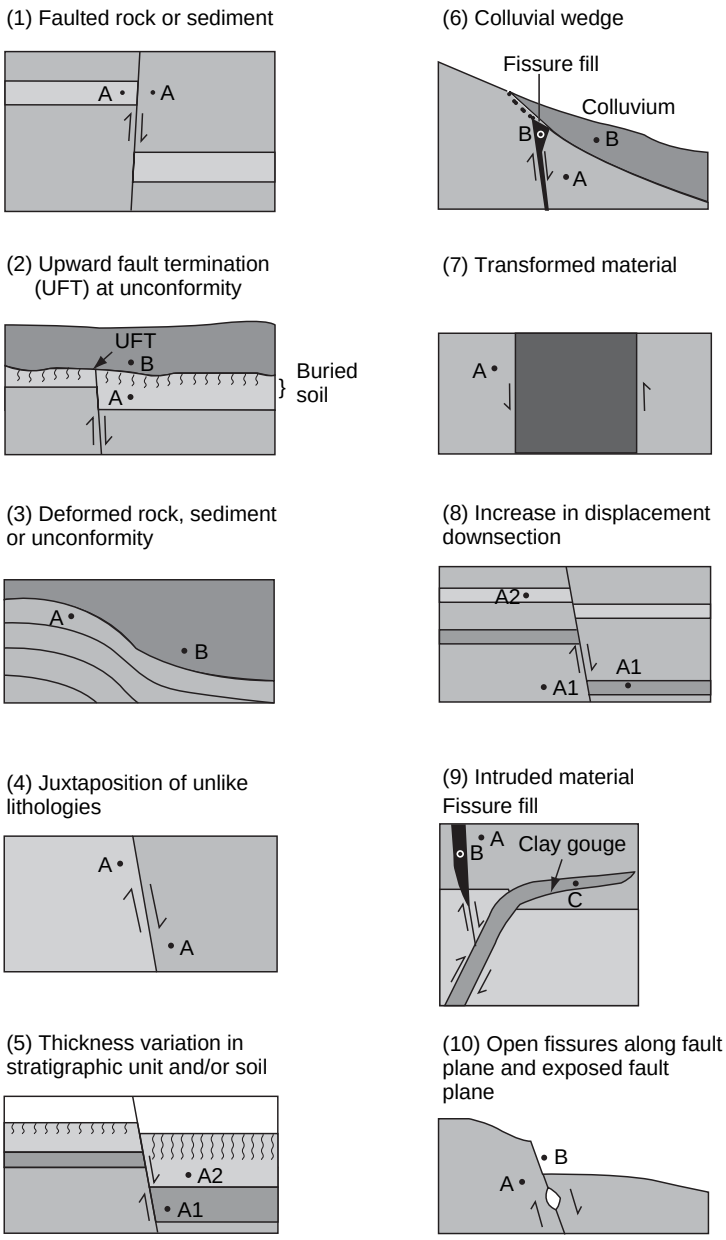


Figure 1 Diagrams illustrating stratigraphic and structural criteria used to identify the occurrence and timing of paleoearthquakes. Stratigraphy, faults, and soils are shown schematically. Dated samples at locations marked 'A' predate the earthquake. Dated samples at locations 'B' postdate the earthquake. Dated samples at 'C' are not helpful for deciphering the chronology of past earthquakes because the sample may predate or postdate the earthquake. Where evidence of multiple earthquakes is present, locations A1 predate the earlier event and locations A2 postdate the earlier event and predate the later event. From Lettis WR and Kelson KI (2000) Applying geochronology in paleoseismology. In: Noller JS, Sowers JM, and Lettis WR (eds.) *AGU Reference Shelf, Series 4*: *Quaternary Geochronology: Methods and Applications*, pp. 479–495. Washington, DC: American Geophysical Union.

Tectonic subsidence induces sedimentation that can record and preserve evidence of an earthquake (Carver and McCalpin, 1996). Buried flora or fauna may provide material for dating the earthquake. For example, in the Cascadia region of northwestern US, tsunami deposits, drowned forests, and grasslands along the coastline provided material for high-precision dating of subduction zone earthquakes (e.g., Atwater *et al.*, 1995). The date of the most recent earthquake (AD 1700) was resolved from radiocarbon dating of

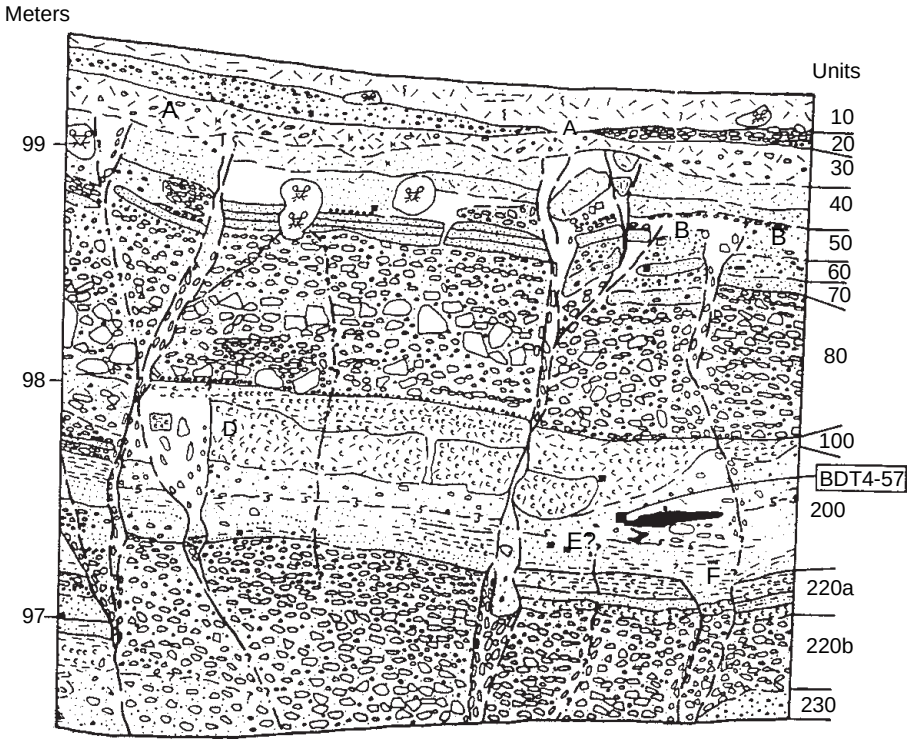


Figure 2 Example of a portion of a trench log from the San Andreas Fault at Bidart Ranch in the Carrizo Plain, California. Letters indicate the approximate location of evidence for several paleoearthquakes (A,B,D,E?,F). Query indicates uncertainty. From Grant LB and Sieh K (1994) Paleoseismic evidence of clustered earthquakes on the San Andreas Fault in the Carrizo Plain, California. *Journal of Geophysical Research* 99(B4): 6819–6841.

submerged deposits, analysis of tree rings (dendroseismology), and historic accounts of tsunami recorded in Japan (Nelson *et al.*, 1995; Satake *et al.*, 1996).

Measurement and interpretation of vertical deformation in coastal regions must be based on knowledge of sea level at the time of uplift. Sea level was 100–150 m below modern levels during the last glacial maximum 15 000–20 000 years BP (Lajoie, 1986). After rising rapidly in the Early Holocene, sea level stabilized in the mid-Holocene. However, sea level has fluctuated slightly (±2 m) during the Late Holocene (Lajoie, 1986) and is currently at different levels in different parts of the world (Pirazzoli, 1991). Fluctuations and regional variations have been sufficiently large to complicate interpretations of coseismic uplift or subsidence in areas with low vertical deformation rates.

4.19.2.2.3 Ground shaking and secondary effects

Surface rupture and regional deformation are considered primary effects of earthquakes because they can be directly associated with tectonic movement of specific faults. Secondary effects of earthquakes such as liquefaction, ground failure, slope failure, tsunami, and seismic seiche are induced by shaking. The locations and characteristics of precariously balanced rocks provide constraints on maximum ground motions over the time period in which the rocks remained precarious (e.g., Brune, 2002). The maximum amplitude acceleration ground motion required to topple a precariously balanced rock can be calculated from field measurements of the rock height and width. Thus, precariously balanced rocks apparently exist only in places where the ground motions have been relatively low. A history of seismic shaking is an indirect history of regional fault activity and is useful for seismic hazard assessment. If secondary effects disturb the environment, and if they are preserved, then these effects can be cataloged to develop a chronology of paleoearthquakes. The dates of shaking events may be constrained by geochronology methods.

Evidence of liquefaction, inundation by tsunami, and seismically induced ground failures may be observed in natural or man-made exposures of young sediments, or in sediment cores. Liquefaction

Figure 3 Photo of Holocene shorelines and wave erosion platforms along rocky coastline of San Joaquin Hills, California. L. Ballenger is standing on the active erosion platform. An elevated platform intersects the sea cliff at the paleo-shoreline. An older paleo-shoreline forms a notch in the sea cliff near the top of the photo. The paleo-shorelines were formed by tectonic uplift of the San Joaquin Hills (Grant and Ballenger, 1999).

is a loss of bearing strength that occurs when saturated cohesionless sediments are subjected to strong shaking or cyclical loading. Fluidized sediment may mobilize during liquefaction to form distinctive sedimentary or geomorphic structures such as sand blows, sand dikes, and craters (Obermeier, 1996; Kuhn, 2005). Ground failures such as lateral spreading, slumping, lurching, and cracking may occur in soft sediment. In submarine or lacustrine environments, deposition of turbidites (Obermeier, 1996) or mixed layer clays (Marco *et al.*, 1996; Doig, 1998) can provide evidence of a paleoearthquake. Nonseismic events can create structures that are virtually indistinguishable from seismically deformed sediments, or seismites (Ricci, 1995). Therefore, paleoseismologists must correlate candidate seismites over regions and rule out nontectonic origins before concluding that an earthquake occurred. For example, the widespread presence of features attributed to liquefaction and ground failure provide strong evidence of the AD 1700 Cascadia earthquake (Obermeier and Dickinson, 2000; Adams, 1996).

Earthquake ground motions often trigger slope failures. Mountainous regions with steep slopes or unstable slope materials are particularly susceptible to seismically induced slope failures. Slope failures have been responsible for many fatalities in historic earthquakes (Bolt, 1999). In seismically active regions, slope failures over large areas may be linked to paleoearthquakes. Several methods have been developed to date paleoearthquakes by measuring the age of seismically induced landslide deposits. For example, lichenometry has been applied to date seismically induced rockfalls along the Alpine Fault Zone, New Zealand (Bull, 1996). The criteria for determining seismic origin, and methods for dating are summarized by Jibson (1996).

4.19.2.3 Measurement of Slip

Measurement of slip, or surface displacement, across a fault yields information about the magnitude of paleoearthquakes and the average rate of deformation. The slip rate of a fault provides an upper bound for the rate of seismic moment (Youngs and Coppersmith, 1985; WGCEP, 1995). Slip from paleoearthquakes can be used in combination with slip rate to estimate the recurrence intervals between earthquakes and to constrain the magnitude of past ruptures.

4.19.2.3.1 Slip rates

The slip rate of a fault, V, is obtained by measuring the displacement, D, across a fault zone during time interval, T, using eqn [1]:

$$V = D/T \qquad [1]$$

To measure a slip rate, a geologist must find a feature that crosses the fault, is well defined, and is datable. Such features are called 'piercing lines'. The points where they intersect a fault zone are called 'piercing points'. Fault slip is measured from the displacement of piercing points and deformation of piercing lines. Geological piercing lines include streams, gullies, and linear sedimentary features. Anthropogenic piercing lines include walls, row crops, rice paddy boundaries, fences, and roads. The ideal piercing line crosses the fault at a right angle, is perfectly linear, has the same trend on either side of the fault zone, and is of precisely known age. Such conditions rarely exist. Therefore, most slip rates have considerable measurement uncertainty. For example, a compilation of fault slip rates in California (Peterson and Wesnousky, 1994) shows significant uncertainty in slip rate measurements.

Uncertainties in slip rates are caused by uncertainties in measurements of displacement and in estimates of the time interval over which the displacement occurred (e.g., Keller *et al.*, 1982; van der Woerd *et al.*, 2006). Aperture of measurement affects slip rates because deformation may occur beyond the main fault zone (McGill and Rubin, 1999; Rockwell *et al*; 2002) and have only subtle effects on piercing

lines (Grant and Sieh, 1993; Grant and Donnellan, 1994; Sieh, 1996). Therefore, geologically measured slip rates may provide only minimum values if measured over short apertures. For blind faults, the displacement on the fault plane must be inferred from deformation far outside the fault zone, and the resulting slip rates have even larger uncertainties.

The length and variation of time intervals between surface ruptures may affect measurements of slip rate or slip per event. For faults that do not creep, surface displacements are generated episodically by earthquakes. If the slip rate of a fault is low and the average time between surface ruptures is long – the case for most faults – measurements of slip rate may be significantly greater or less than the average rate over multiple rupture cycles. It is generally assumed that slip rate is constant over the interval of measurement, but for most faults there are insufficient data to test this assumption. Geomorphic analysis of normal fault scarps in the Basin and Range Province, USA, reveals significant temporal variation in slip rates for several faults (Slemmons, 1995), and irregular earthquake recurrence along the Aksu Thrust in Tianshan China led to irregular strain release and slip rate (Hubert-Ferrari *et al.*, 2005).

4.19.2.3.2 Slip per event

The amount of slip per event can be used to estimate the magnitude of previous earthquakes and the time between earthquakes (if the average slip rate is known). The distribution of slip may also reveal the location of segment boundaries, and areas of characteristic slip, if any are present (Ward, 1997).

To measure slip per paleoearthquake, it is necessary to identify a piercing line that has been displaced by a known number of earthquakes. Ideally, this is done by measuring the displacement of a piercing line of known age in an area where the dates of individual surface ruptures are well constrained (e.g., Liu *et al.*, 2004; Liu-Zeng *et al.*, 2006). For normal and reverse faults, the vertical component of displacement can be measured directly by excavating a trench across the scarp (**Figure 4**) because material deposited by collapse or decay of a vertical scarp

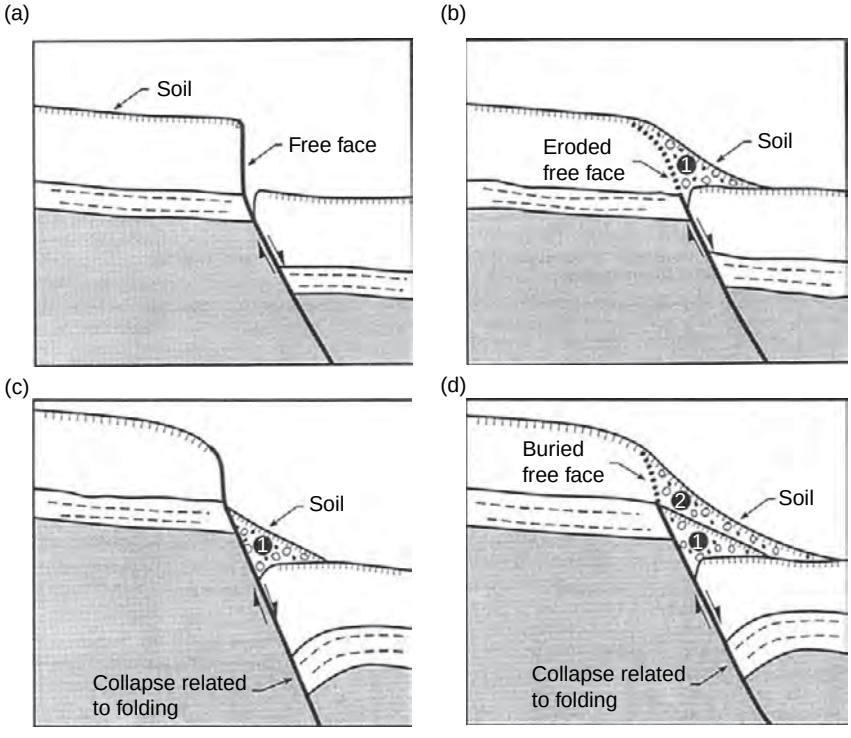


Figure 4 Idealized sequence of two normal faulting events and formation of a colluvial wedge, as exposed in trench excavations. (a) First surface faulting event; (b) first colluvial wedge 1 and soil formation; (c) Second surface faulting event; (d) second colluvial wedge 2 and soil formation. From Lettis WR and Kelson KI (2000) Applying geochronology in paleoseismology. In: Noller JS, Sowers JM, and Lettis WR (eds.) *AGU Reference Shelf, Series 4*: *Quaternary Geochronology: Methods and Applications*, pp. 479–495. Washington, DC: American Geophysical Union.

typically buries and preserves evidence of the rupture. Changes in slope or height of a scarp may indicate multiple ruptures. Variation in the amount of lateral displacement of piercing lines along a strike-slip fault can indicate multiple surface ruptures. In coastal areas, the vertical component of displacement can be measured from elevated marine shorelines (Matsuda *et al.*, 1978; Plafker and Rubin, 1978). For strike-slip faults, the best measurements are obtained from three-dimensional excavation using multiple trenches (Weldon *et al.*, 1996, 2002; Liu-Zeng *et al.*, 2006). An alternate but less reliable method is to measure displacement of surficial features such as laterally offset channels. The total number of earthquakes is then inferred from peaks in the frequency distribution of measurements (**Figure 5**), or by the assumption of constant (or 'characteristic') slip per event (see Section 4.19.3.2). Uncertainty in the resulting slip-per-event measurements is unconstrained.

4.19.2.4 Magnitude of Paleoearthquakes

In general, the effects of an earthquake are proportional to its size (Slemmons and dePolo, 1986). Therefore, the amount of slip, tectonic deformation, or secondary effects preserved in the paleoseismic record helps constrain the size of a paleoseismic event. The magnitude of paleoearthquakes can be estimated from measurements of slip per event or from rupture length inferred by correlating dates of paleoseismic events at different sites.

Empirical regression relationships and other statistical relationships between historic earthquake magnitude and rupture length, average slip, or maximum slip (Wells and Coppersmith, 1994; Biasi and Weldon, 2006) can be applied to estimate the magnitude of paleoearthquakes (**Table 3**). Hemphill-Haley and Weldon (1999) developed empirical relationships specifically for estimating magnitude from paleoseismic measurements of slip. Their method provides uncertainties based on the number of measurements and percentage of rupture length sampled.

For faults with chronologies of earthquakes from multiple paleoseismic investigations, surface rupture lengths can be estimated by assuming that ruptures with overlapping time windows occurred at the same time (see Section 4.19.3.4). However, there are large uncertainties in correlating ruptures between paleoseismic investigation sites, so the inferred rupture length and resulting magnitudes are poorly constrained. Segmentation models (see Section 4.19.3.1) have also been applied to estimate paleoearthquake rupture length and magnitude.

4.19.3 Models

Paleoseismic data provide information about earthquakes over scales of time and magnitude that are useful for seismic hazard assessment and essential for

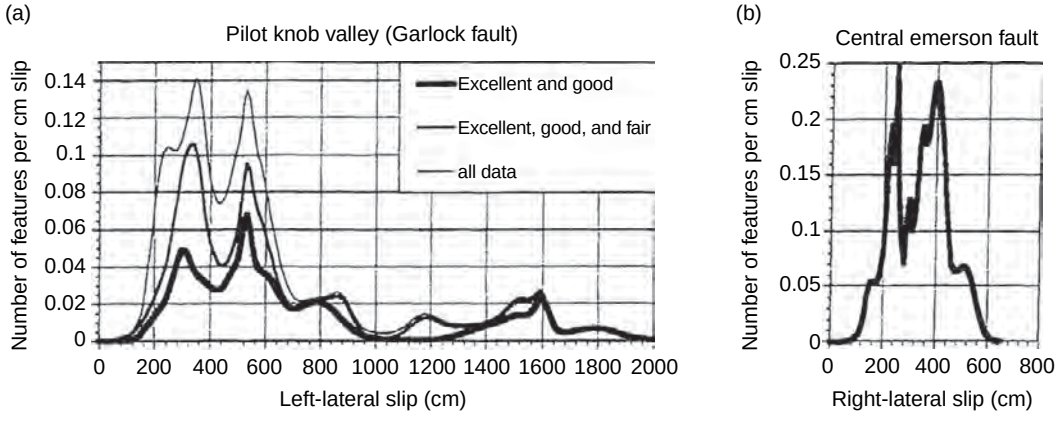


Figure 5 Summed Gaussian probability density functions for measured surficial offsets along the Garlock Fault (from McGill and Sieh, 1991), and from the 1992 M 7.3 Landers earthquake on the Emerson Fault, California. Each peak in the Garlock distribution was originally interpreted to be the result of displacement from a separate surface rupture. Measurements from the Landers earthquake show that surficial offset from a single earthquake can have a multimodal frequency distribution. From McGill SF and Rubin CM (1999) Surficial slip distribution on the central Emerson fault during the June 28, 1992, Landers earthquake, California. *Journal of Geophysical Research* 104(B3): 4811–4833.

Table 3 Selected empirical relationships between moment magnitude (*M*), average displacement (AD), maximum displacement (MD), and surface rupture length (SRL)[a]

Equation[b]	Slip type[c]	a (sa)[d]	b (sa)[d]	Standard deviation	Correlation coefficient
$M = a + b^*\log(AD)$	SS	7.04(0.05)	0.89(0.09)	0.28	0.89
	R[e]	6.64(0.16)	0.13(0.36)	0.50	0.10
	N	6.78(0.12)	0.65(0.25)	0.33	0.64
	All	6.93(0.05)	0.82(0.10)	0.39	0.75
$M = a + b^*\log(MD)$	SS	6.81(0.05)	0.78(0.06)	0.29	0.90
	R[e]	6.52(0.11)	0.44(0.26)	0.52	0.36
	N	6.61(0.09)	0.71(0.15)	0.34	0.80
	All	6.69(0.04)	0.74(0.07)	0.40	0.78
$M = a + b^*\log(SRL)$	SS	5.16(0.13)	1.12(0.08)	0.28	0.91
	R[e]	5.00(0.22)	1.22(0.16)	0.28	0.88
	N	4.86(0.34)	1.32(0.26)	0.34	0.81
	All	5.08(0.10)	1.16(0.07)	0.28	0.89

[a]Compiled from Wells DL Coppersmith KJ (1994). *New empirical relationships among magnitude, rupture length, rupture area, and surface displacement. *Bulletin of the Seismological Society of America* 84: 974–1002
[b]Displacement in meters. Surface rupture length in kilometers.
[c]SS, strike-slip; R, reverse; N, normal; All, all fault types.
[d]Coefficients and standard errors.
[e]Relationship is not significant at the 95% confidence level.

understanding the long-term rupture patterns of faults. Therefore, paleoseismic data have been influential in the development and testing of models that describe fault behavior over multiple rupture cycles (e.g., Smith and Sandwell, 2006; Rundle *et al.*, 2006), and in forecasting earthquakes (e.g., Rundle *et al.*, 2005). This section describes models that depend critically on observations of paleoearthquakes, and discusses uncertainty in paleoseismic data.

4.19.3.1 Segmentation

Segmentation models assume that faults are divided into discrete, identifiable sections that behave distinctively over multiple rupture cycles (e.g., Schwartz and Coppersmith, 1986; Slemmons, 1995). Segment boundaries are thought to control the termination and initiation of fault ruptures and therefore limit the magnitude and rupture pattern of an earthquake. Fault segments are defined based on structural discontinuities such as step-overs that are large enough to terminate ruptures (Wesnousky, 2006), changes in strike and rheology (Allen, 1968) as well as by paleoseismic data (Sibson, 2002; Schwartz and Coppersmith, 1984; Sieh and Jahns, 1984; Schwartz and Sibson, 1989; WGCEP, 1988, 1990).

Segmentation models are attractive because they simplify fault behavior. For example, in seismic hazard assessment it is desirable to estimate the size of the largest earthquake that can occur on a fault (Yeats *et al.*, 1997; Mualchin, 2005). The maximum magnitude earthquake can be estimated from empirical relationships (**Table 3**) and the estimated length of maximum rupture (Wells and Coppersmith, 1994). This would be either the maximum length of the fault, or of the segment. Similarly, the size of a paleoearthquake can be estimated by assuming that the entire segment ruptured.

Testing of segment models requires defining segments and then observing multiple ruptures. This would take decades to centuries for the fastest slipping faults, and thousands of years for most faults. Therefore, few segment models have been tested against rupture patterns of historic earthquakes (Sieh, 1996). Results are mixed. For example, two large earthquakes on strike-slip faults in the western US (M 7 1989 Loma Prieta and M 7.3 1992 Landers) propagated through previously identified segment or fault boundaries (WGCEP, 1990; Sieh *et al.*, 1993). However, several historic ruptures of normal faults in the Great Basin have terminated at or near fault discontinuities (Zhang *et al.*, 1999). More observations are needed to test segmentation hypotheses and their utility for seismic hazard assessment.

4.19.3.2 Characteristic Earthquakes

The characteristic earthquake model (Schwartz and Coppersmith, 1984) is probably the most influential model of fault rupture and earthquake recurrence

developed from paleoseismic data. The basic tenet of the model is that most surface slip on a fault occurs in 'characteristic earthquakes'. Characteristic earthquakes are the result of 'characteristic slip'. **Figure 6(c)** illustrates the concept of characteristic slip: at a specific location along a fault, the displacement (slip) is the same in successive characteristic earthquakes. This implies that characteristic earthquakes have similar rupture patterns and that a fault can be divided into segments that behave characteristically. Each segment would have a distinctive or 'characteristic' rupture pattern and magnitude. The characteristic slip model implies that different amounts of displacement at different locations along a fault are related to differences in slip rate.

If characteristic slip occurs on a fault, then most seismic moment is released by repetition of characteristic earthquakes of approximately the same magnitude. Characteristic slip causes a kink in the frequency magnitude relationship, known as 'characteristic rcurrence' (**Figure 7**), due to the dominance of relatively large-magnitude characteristic earthquakes (Schwartz and Coppersmith, 1984). The characteristic recurrence curve appears to fit the frequency–magnitude distribution for some faults (Wesnousky, 1994) but does not fit global seismicity rates as well as other models (Kagan, 1993).

The characteristic earthquake model is convenient for seismic hazard assessment because it assumes that faults can be divided into identifiable segments that rupture with characteristic slip and recurrence. Many studies of active faults have focused on identifying characteristic rupture segments of faults and their properties (e.g., Wesnousky, 1986; Peterson and Wesnousky, 1994; WGCEP 1988, 1990, 1995). For example, earthquake forecasts for the San

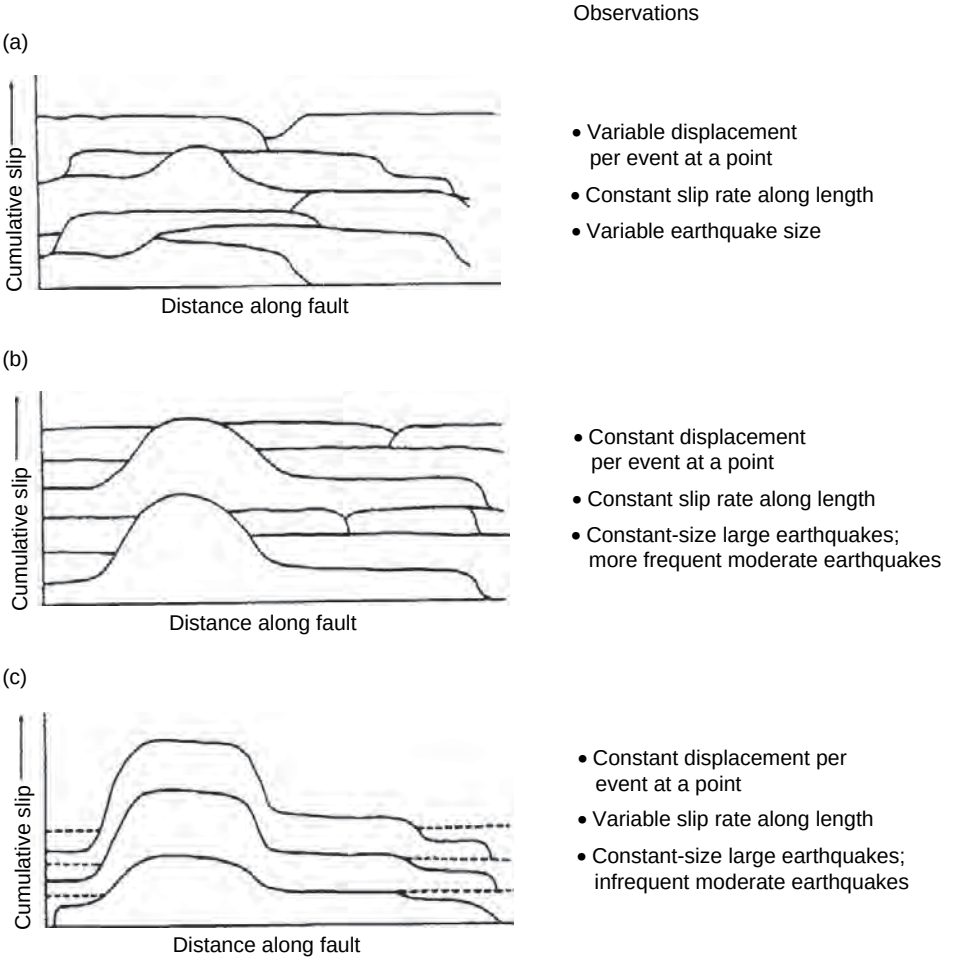


Figure 6 (*Continued*)

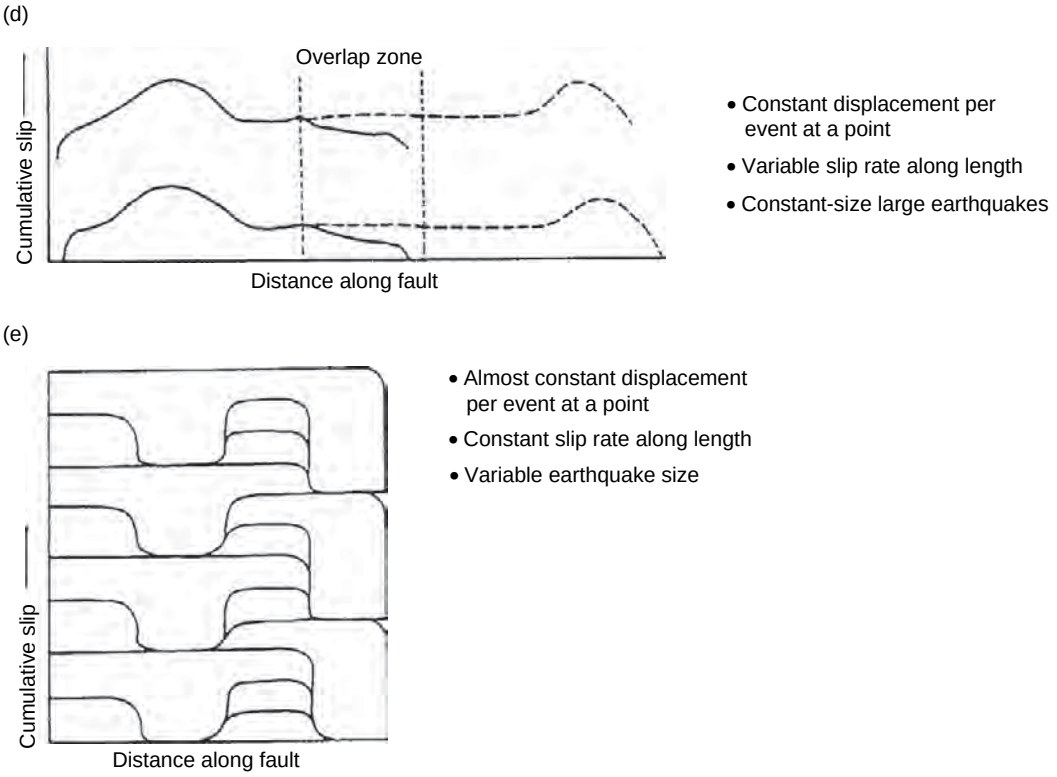


Figure 6 Diagrammatic representations of fault slip models showing the spatial distributions of surface slip and cumulative offset along a fault from multiple ruptures. Each rupture is represented by a line, with the most recent rupture showing the lowest 'cumulative slip'. Features that have been displaced by multiple earthquakes are offset by greater amounts due to cumulative slip. At any location along the fault, the slip rate is a function of repeated slip. The diagrams illustrate models of: (a) variable slip, (b) uniform slip, (c) characteristic slip, (d) overlap, and (e) coupled slip. (From Schwartz and Coppersmith, 1984; Berryman and Beanland, 1991). Reprinted from McCalpin JP (ed.) (1996) '*Paleoseismology*'. San Diego, CA: Academic Press.

Andreas Fault have been based on a combination of segmentation, time-predictable, and characteristic earthquake models. The Parkfield segment of the San Andreas Fault was predicted to break before 1993 with repetition of characteristic earthquake (Bakun and Lindh, 1985; WGCEP, 1988) but did not occur until 2004. Several aspects of the rupture were similar to previous Parkfield earthquakes, but other features were unexpected (Harris and Arrowsmith, 2006; Toke *et al.*, 2006). The Working Group on California Earthquake Probabilities (WGCEP, 1988, 1990, 1995) divided the San Andreas Fault into characteristic segments and issued probabilistic rupture forecasts. In 1989, the M7 Loma Prieta earthquake occurred on portions of two segments defined prior to the earthquake (WGCEP, 1990) rather than on a single segment. Hecker and Schwartz (1994) analyzed paleoseismic data for many faults and concluded that most faults exhibit characteristic slip. Characteristic earthquakes have also been reported from Greece (Papadopoulos *et al.*, 2003) and Turkey (Klinger *et al.*, 2003). However, recent paleoseismic studies of the San Andreas Fault reveal complexity in spatial and temporal rupture patterns that was not evident when the characteristic earthquake model was first proposed (Grant, 1996). Debate about the applicability of the characteristic earthquake model for forecasting earthquakes is likely to continue (Yeats *et al.*, 1997; Yeats, 2001).

4.19.3.3 Time Predictable and Recurrence Models

Recurrence time, T_r, is the time interval between successive ruptures of the same fault. T_r is useful for describing the seismic parameters of an active fault (Somerville and Moriwaki, 2002). Paleoseismic investigations are the main source of data for measuring or estimating T_r. Individual recurrence times or average recurrence intervals (the average time

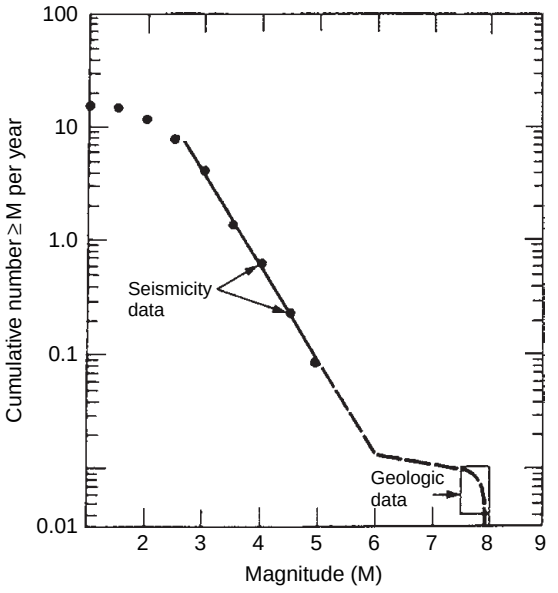


Figure 7 Characteristic recurrence model for cumulative frequency–magnitude distribution of seismicity on a specific fault. From Schwartz DP and Coppersmith KJ (1984) Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones. *Journal of Geophysical Research* 89: 5681–5698.

interval between ruptures) can be measured directly by dating successive surface ruptures. For well-studied faults such as the San Andreas, average recurrence intervals and their variance are available for several locations along the fault (e.g., WGCEP, 1995, 1999). For most faults, there is insufficient information to describe variability in recurrence times. Tests of the recurrence models against the paleoearthquake chronology of the San Andreas Fault at Wrightwood show that the rupture record cannot be simply described by either a time-predictable or slip-predictable model (Weldon *et al.*, 2004).

If dates of paleoearthquakes are not available, recurrence time may be estimated for a fault using the relationship

$$T_r = D/V \quad [2]$$

where V is the slip rate and D is average displacement (slip). Assuming the slip rate is constant over the period of observation and there is no creep, then the recurrence time is a linear function of displacement. If displacement occurs in constant-size (characteristic) earthquakes, then the recurrence time can be 'predicted'. Thus eqn [2] is referred to as the time-predictable model, after Shimazaki and Nakata (1980) who developed it from observations of uplifted coastal terraces in Japan. If the characteristic earthquake model and time-predictable model are combined, the result is a periodic recurrence model.

Analysis of historic earthquakes and paleoseismic data from several regions shows that occurrence of large earthquakes is irregular (Goes, 1996). Surface ruptures of the San Andreas Fault have been proposed to occur irregularly (Jacoby *et al.*, 1988; Sieh *et al.*, 1989) or in clusters (Grant and Sieh, 1994; Rockwell *et al.*, 2002). Temporal and/or spatial clustering of surface ruptures has also been proposed to describe surface ruptures in the Basin and Range (Wallace, 1981; Slemmons, 1995; McCalpin, 1996) and eastern California shear zone, USA (Rockwell *et al.*, 2002). Historic ruptures of the North Anatolian Fault Zone in Turkey (Hartleb *et al.*, 2003, 2006) and patterns of Late Holocene seismicity in Iran suggest that seismicity may be triggered by stress shadows from previous ruptures to form temporal clusters or sequences of damaging earthquakes (Stein *et al.*, 1997; Berberian and Yeats, 1999). For most faults, the number of documented paleoearthquakes and the precision of their dates are inadequate to test hypotheses of clustered or triggered earthquakes (Biasi *et al.*, 2002). Larger paleoseismic data sets are needed to conduct statistically significant tests of recurrence models.

4.19.3.4 Rupture Patterns

Patterns of fault rupture in space and time yield insights about the physics of faulting (Ward, 1997; Weldon *et al.*, 2004, 2005) and provide templates for estimating the location and size of future earthquakes. Several models of fault behavior shown in **Figure 6** (see caption for explanation of diagram) are based on multiple paleoseismic measurements of slip at different sites along a fault. The variable slip model (**Figure 6(a)**) assumes a constant slip rate along the fault with variable displacement per event at a location, and variable earthquake size. The uniform slip model (**Figure 6(b)**) is similar to the characteristic slip model (**Figure 6(c)**) because both assume constant-size displacement at a location. The slip rate is constant along the fault in the uniform slip model. Therefore, a range of earthquake sizes is required to account for sections of fault with large or small slip per event. The overlap model (**Figure 6(d)**) assumes constant displacement per event at a point (characteristic slip), variable slip rate along the fault, and repetition of constant-size large earthquakes. Finally, the coupled model (**Figure 6(e)**) is a uniform slip model coupled with a segmentation model.

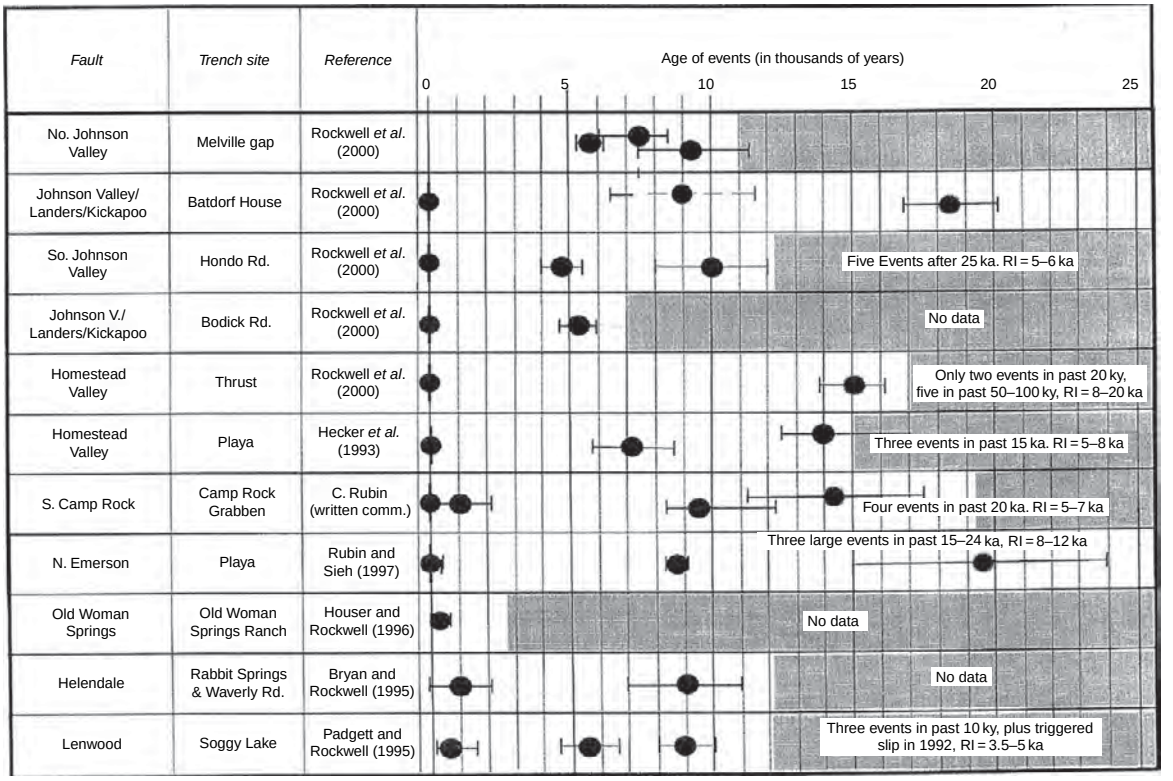


Figure 8 Timing of moderate to large surface rupturing events on Mojave Block faults. Diagram displays data on the rupture history of faults in the Mojave desert, California, USA, that ruptured in the 1992 Landers earthquake. The dates of paleoearthquakes are represented by black dots with horizontal bars to indicate uncertainty. The dates of many paleoearthquakes overlap, suggesting that these faults have ruptured together prior to the Landers earthquake, or in relatively brief temporal 'clusters' of earthquakes. (Rockwell TK, Lindvall S, Herzberg M, Murbach D, Dawson T, and Berger G (2000) Paleoseismology of the Johnson Valley, Kickapoo, and Homestead Valley faults: Clustering of earthquakes in the eastern California shear zone. *Bulletin of the Seismological Society of America* 90(5): 1200–1236.

Spatial and temporal patterns of fault rupture can be analyzed by plotting the dates of paleoearthquakes on a space–time diagram. **Figures 8** and **9** show examples from a group of faults that ruptured in the 1992 Landers earthquake (**Figure 8**), and the southern San Andreas Fault (**Figure 9**). Uncertainty in dates ranges from decades to more than a thousand years. Paleoearthquakes at adjacent sites with overlapping dates could have been the same rupture, or separate ruptures that are indistinguishable within the uncertainty. Additional dates of paleoearthquakes would provide better constraints on the locations and magnitudes of past ruptures.

4.19.3.5 Uncertainty in Paleoseismic Data

The quality and reliability of paleoseismic data are affected by the investigation method, the characteristics of the study site. and the perspective of the investigator. Many paleoseismic studies are destructive to the study site, so the data cannot be reproduced or compared with a control. Therefore, each paleoseismic publication becomes a repository of data that will be interpreted by scientists who will not be able to evaluate it firsthand. Collection of paleoseismic data requires recognition and interpretation of complex patterns in the geologic record and landscape. Therefore, some interpretation is automatically convolved with the data, and no deconvolution algorithm can be applied after the data are published. Paleoseismologists have developed several methods to reduce subjectivity and standardize uncertainty in their data. Fault exposures are generally documented with photographs as well as logs. Photos are objective records, devoid of interpretation. 'Trench parties' are held to allow reviewers to examine primary field data prior to publication. When possible, care is taken to fill excavations so that they can be reopened at a later date. Uncertainty in data collection and field

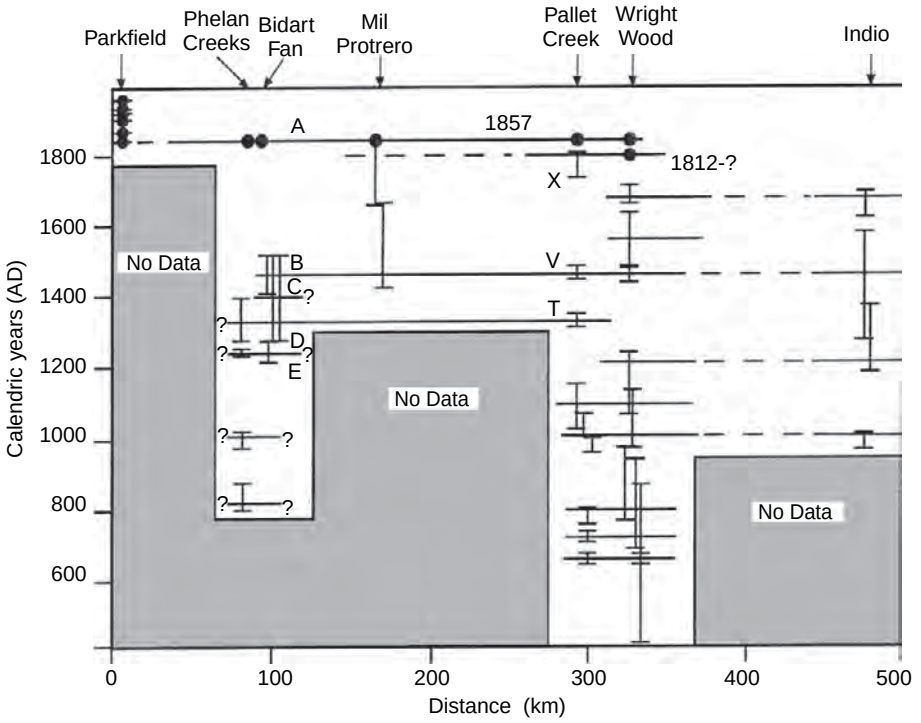


Figure 9 Diagram displays dates of paleoearthquake ruptures at different sites along the southern San Andreas Fault, California, USA. Vertical lines show age limits for paleoearthquakes at each site. Horizontal lines show possible ruptures through multiple sites at the same time. This is one of many possible interpretations of rupture history based on the paleoseismic data available in 1997. (From Arrowsmith *et al.*, 1997). Dates of more paleoearthquakes are now available (see **Table 2**), and newer rupture models have been proposed to fit the expanded data set (Weldon *et al.*, 2004, 2005), but multiple interpretations are still allowed by uncertainties in the dates of individual paleoearthquakes. Modified from Grant and Sieh (1994), Sims (1994), Biasi and Weldon (1994), and Grant (1996).

interpretations are often reported qualitatively. For example, Sieh (1978) and McGill and Sieh (1991) reported the apparent 'quality' of their measurements of offset streams. Paleoseismologists may report high-quality measurements with 'high confidence'. Such assessments are qualitative, and should not be confused with statistical confidence intervals.

4.19.4 Applications and Future Directions

Paleoseismic data are used for seismic-source characterization in hazard assessment and mitigation (Somerville and Moriwaki, 2002). Probabilistic seismic hazard assessment incorporates information on fault location and geometry, slip rates, recurrence intervals, dates of previous ruptures, and maximum earthquakes. Maximum earthquakes are estimated from segmentation models, slip per event (Biasi and Weldon, 2006), and lengths of previous ruptures. Deterministic hazard assessments use paleoseismic data to develop earthquake scenarios including rupture location, magnitude, and surface displacement. Earthquake scenarios are used to estimate expected damage for emergency response planning. Surface rupture hazard can be mitigated by identifying fault zones that are most likely to rupture and restricting land uses within such zones. Paleoseismic investigations provide data on the recency and frequency of fault rupture, as well as the type and width of surface rupture. This information can be used to classify fault zones according to rupture potential and expected displacement in future earthquakes. Critical structures can be sited to avoid rupture hazard zones (Mualchin, 2005), or designed to accommodate expected displacement.

Paleoseismology is a young science that focuses on understanding the long-term patterns of large earthquakes. Applications of new or improved geochronological methods (e.g., van der Woerd *et al.*, 2006; Akciz *et al.*, in review) are leading to significant

reductions in uncertainty in paleoseismic data. Paleoseismic data can be applied toward understanding the behavior of faults, estimating the potential for large earthquakes to occur in populated areas, and for mitigating seismic hazard. As populations increase in seismically active areas, many paleoseismologists are moving toward predictive or applied work in seismic hazard assessment. Establishment of digital databases (Grant *et al*., 2005) and standard formats for paleoseismic data (Grant and Gould, 2004) allows better integration with more quantitative fields of seismology, seismic hazard assessment, and earthquake engineering. As paleoseismology matures, it is likely to play an increasingly important role in seismology and society.

References

Adams J (1996) Paleoseismicity of the Cascadia subduction zone: Evidence from turbidites off the Oregon–Washington margin. *Tectonics* 9: 569–583.

Akciz SO, Grant LB, Arrowsmith JR, *et al.* (2006) Re-dating Bidart earthquakes: Confirmation of short recurrence time (<150 years) for ruptures along the San Andreas fault in the Carrizo Plain: Proceedings and abstracts, 2006 SCEC Annual Meeting, vol. 16, p. 65.

Allen CR (1968) The tectonic environments of seismically active and inactive areas along the San Andreas fault system. In: Dickinson WR and Grantz A (eds.) *Stanford University Publications in the Geological Sciences, vol. 11: Proceedings of Conference on Geologic Problems of San Andreas Fault System*, 70–82. Stanford, CA: Stanford University.

Arrowsmith JR, McNally K, and Davis J (1997) Potential for earthquake rupture and M7 earthquakes along the Parkfield, Cholame and Carrizo segments of the San Andreas fault. *Sesmological Research Letters* 68: 6902–6916.

Atwater BF, Nelson AR, Clague JJ, *et al.* (1995) Summary of coastal geologic evidence for past great earthquakes at the Cascadia subduction zone. *Earthquake Spectra* 11: 1–18.

Bakun WH and Lindh AG (1985) The Parkfield, California earthquake prediction experiment. *Science* 229: 619–624.

Berberian M and Yeats RS (1999) Patterns of historical earthquake rupture in the Iranian plateau. *Bulletin of the Seismological Society of America* 89: 1120–139.

Berryman KR and Beanland S (1991) Variation in fault behavior in different tectonic provinces of New Zealand. *Journal of Structural Geology* 13: 177–189.

Biasi G and Weldon RJ, II (1994) Quantitative refinement of calibrated ^{14}C distributions. *Quaternary Research* 41: 1–18.

Biasi G and Weldon RJ, II (2006) Estimating surface rupture length and magnitude of paleoearthquakes from point measurements of rupture displacement. *Bulletin of the Seismological Society of America* 96(5): 1612–1623.

Biasi GP, Weldon RJ, II, Fumal TE, and Seitz GG (2002) Paleoseismic event dating and the conditional probability of earthquakes on the southern San Andreas fault, California. *Bulletin of the Seismological Society of America* 92(7): 2761–2781.

Bolt BA (1999) *Earthquakes,* 4th edn. New York: W. H. Freeman.

Bonilla MG (1988) Minimum earthquake magnitude associated with coseismic surface faulting. *Bulletin of the Association of Engineering Geologists* XXV: 17–29.

Bonilla MG and Lienkaemper JJ (1990) Visibility of fault strands in exploratory trenches and timing of rupture events. *Geology* 18: 153–156.

Brune JN (2002) Precarious rock constraints on ground motion from historic and recent earthquakes in Southern California. *Bulletin of the Seismological Society of America* 92(7): 2602–2611.

Bull WB (1996) Prehistorical earthquakes on the Alpine fault, New Zealand. *Journal of Geophysical Research* 101: 6037–6050.

Bullard TF and Lettis WR (1993) Quaternary fold deformation associated with blind thrusting, Los Angeles Basin, California. *Journal of Geophysical Research* 98: 8349–8369.

Carver GA and McCalpin JP (1996) Paleoseismology of compressional tectonic environments. In: McCalpin JP (ed.) *Paleoseismology*, pp. 183–270. San Diego, CA: Academic Press.

Crone AJ and Omdahl EM (eds.) (1987) Directions in Paleoseismology. *Proceedings of Conference* XXXIX, *US Geological Survey Open File Report 87–673*, pp. 1–456.

Doig R (1998) 3000-yr Paleoseismological record from the region of the 1988 Saguenay, Quebec, earthquake. *Bulletin of the Seismological Society of America* 88: 51198–51203.

Dolan JF, Stevens D, and Rockwell TK (2000) Paleoseismologic evidence for an early to Mid-Holocene Age of the most recent surface rupture on the Hollywood fault, Los Angeles, California. *Bulletin of the Seismological Society of America* 90: 334–344.

Fumal T, Niemi T, and Zhang H (2006) High resolution paleoseismic records at three sites on the northern San Andreas Fault. *Seismological Research Letters* 77: 269.

Fumal TE, Weldon RJ, II, Biasi GP, *et al.* (2002) Evidence for large earthquakes on the San Andreas fault at the Wrightwood, California, paleoseismic site: A.D. 500 to present. *Bulletin of the Seismological Society of America* 92: 72726–72760.

Goes SDB (1996) Irregular recurrence of large earthquakes: An analysis of historic and paleoseismic catalogs. *Journal of Geophysical Research* 101: 5739–5749.

Goldfinger C and Nelson CH (2000) Holocene seismicity of the northern San Andreas fault based on the turbidite record. '*3rd Conference on Tectonic Problems of the San Andreas Fault System*: *Program and Abstracts*', Stanford, CA.

Grant LB (1996) Uncharacteristic earthquakes on the San Andreas fault. *Science* 272: 826–827.

Grant LB (2002) Paleoseismology. In: Lee WH, Kanmori H, and Jennings (eds.) *IASPEI International Handbook of Earthquake and Engineering Seismology*, ch. 10, vol. 81A, pp. 475–489. Colorado, CA: International Association of Seismology and Physics of the Earth's Interior.

Grant LB, Arrowsmith JR, and Akciz S (2005) A composite chronology of earthquakes from the Bidart fan Paleoseismic Site, San Andreas Fault, California. *EOS Transactions of American Geophysical Union,* Fall Meeting Supplement 86(52): U42A-07.

Grant LB and Ballenger LJ (1999) Paleoseismic evidence of a historic coastal earthquake and uplift of the San Joaquin Hills, Southern California. Eos, *Transactions of the American Geophysical Union* 80: F736.

Grant LB, Ballenger LJ, and Runnerstrom EE (2002) Coastal uplift of the San Joaquin Hills, Southern Los Angeles basin, California, by a large earthquake since 1635 A.D. *Bulletin of the Seismological Society of America* 92: 590–599.

Grant LB and Donnellan A (1994) 1855 and 1991 surveys of the San Andreas fault: Implications for fault mechanics. *Bulletin of the Seismological Society of America* 84: 241–246.

Grant LB and Gould MM (2004) Assimilation of paleoseismic data for earthquake simulation. *Pure and Applied Geophysics* 161(11/12): 2295–2306.

Grant LB, Gould MM, Donnellan A, *et al.* (2005) Web-service based universal approach to heterogeneous fault databases. *Computing in Science and Engineering* 7: 51–57.

Grant LB and Lettis WR (eds.) (2002) Paleoseismology of the San Andreas fault system. *Special Issue: Bulletin Seismological Society of America*. 92(7): 2551–2554.

Grant LB, Mueller KJ, Gath EM, *et al.* (1999) Late quaternary uplift and earthquake potential of the San Joaquin Hills, Southern Los Angeles basin, California. *Geology* 27: 1031–1034.

Grant LB and Sieh K (1993) Stratigraphic evidence for 7 meters of dextral slip on the San Andreas fault during the great 1857 earthquake in the Carrizo Plain. *Bulletin of the Seismological Society of America* 83: 619–635.

Grant LB and Sieh K (1994) Paleoseismic evidence of clustered earthquakes on the San Andreas fault in the Carrizo Plain, California. *Journal of Geophysical Research* 99(B4): 6819–6841.

Grant LB, Waggoner JT, von Stein C, and Rockwell TK (1997) Paleoseismicity of the north branch of the Newport–Inglewood fault zone in Huntington Beach, California, from cone penetrometer test data. *Bulletin of the Seismological Society of America* 87: 277–293.

Hall NT (1986) Holocene slip rates on the San Andreas fault in Northern, California. Paper presented at the *The National Earthquake Prediction Evaluation Council and The San Francisco Bay Region Special Study Areas Workshop*, 26 Febraury–1 March, 1986, Menlo Park, California.

Harris RA and Arrowsmith JR (2006) Introduction to the special issue on the 2004 Parkfield earthquake and the Parkfield earthquake prediction experiment. *Special issue: Special Issue on the 2004 Parkfield Earthquake and Parkfield Experiment*. *Bulletin of the Seismological Society of America* 96(4B): S1–S10 (doi:10.1785/0120050831).

Hartleb RD, Dolan JF, Akyuz HS, and Yerli B (2003) A 2000-year-long paleoseismologic record of earthquakes along the central North Anatolian fault, from trenches at Alayurt, Turkey. *Bulletin of the Seismological Society of America* 93(5): 1935–1954.

Hartleb RD, Dolan JF, Kozaci O, Akyuz HS, and Seitz GG (2006) A 2500-yr-long paleoseismologic record of large, infrequent earthquakes on the North Anatolian fault and Cukurcimen, Turkey. *GSA Bulletin* 118: 823–840.

Hecker S and Schwartz DP (1994) The characteristic earthquake revisited: Geological evidence of the size and location of successive earthquakes on large faults. US Geological Survey Open File Report 94–568: pp. 79–80.

Hemphill-Haley MA and Weldon RJ, III (1999) Estimating prehistoric earthquake magnitude from point measurements of surface rupture. *Bulletin of the Seismological Society of America* 89(5): 1264–1279.

Hubert-Ferrari A, Suppe J, Wang X, and Lu HF (2005) Irregular earthquake cycle along the southern Tianshan front, Aksu area, China. *Journal of Geophysical Research Solid Earth* 110(B6): B06402 (Jun 11).

Jacoby G, Sheppard P, and Sieh K (1988) Irregular recurrence of large earthquakes along the San Andreas fault – Evidence from trees. *Science* 241: 196–199.

Jibson RW (1996) Using landslides for paleoseismic analysis. In: McCalpin (ed.) *Paleoseismology*, pp. 397–438. San Diego, CA: Academic Press.

Kagan Y (1993) Statistics of characteristic earthquakes. *Bulletin of the Seismological Society of America* 83: 7–24.

Keller EA, Bonkowski MS, Korsch RJ, and Shlemon RJ (1982) Tectonic geomorphology of the San Andreas fault zone in the southern Indio Hills, Coachella Valley, California. *Geological Society of America, Bulletin* 93(1): 46–56.

Keller EA and Pinter N (1996) *Active Tectonics: Earthquakes, Uplift, and Landscape*. Upper Saddle River, NJ: Prentice Hall.

Kelson KI, Streig AR, Koehler RD, and Kang KH (2006) Timing of late Holocene paleoearthquakes on the northern San Andreas Fault at the Fort Ross Orchard Site, Sonoma County, California. *Bulletin of the Seismological Society of America* 96(3): 1012–1028.

Klinger Y, Sieh K, Altunel E, *et al.* (2003) Paleoseismic evidence of characteristic slip on the western segment of the North Anatolian fault, Turkey. *Bulletin of the Seismological Society of America* 93: 2317–2332.

Knudsen KL, Witter RC, Garrison-Laney CE, Baldwin JN, and Carver GA (2002) Past earthquake-induced rapid subsidence along the northern San Andreas Fault: A paleoseismological method for investigating strike-slip faults. *Bulletin of the Seismological Society of America* 92(7): 2612–2636.

Kuhn GG (2005) Paleoseismic features as indicators of earthquake hazards in North Coastal San Diego County, California, USA. *Engineering Geology* 80(1–2): 115–150.

Lajoie KR (1986) Coastal tectonics. In: Wallace RE (chairman) *Active Tectonics: Studies in Geophysics*, 95–124. Washington, DC: National Academic Press.

Lee WH, Kanamori H, and Jennings PC (eds.) (2002) *International Handbook of Earthquake and Engineering Seismology*. Oxford, UK: International Association of Seismology and Physics of the Earth's Interior, Committee on Education.

Lettis WR and Kelson KI (2000) Applying geochronology in paleoseismology. In: Noller JS, Sowers JM, and Lettis WR (eds.) *AGU Reference Shelf, Series 4: Quaternary Geochronology: Methods and Applications*, pp. 479–495. Washington, DC: American Geophysical Union.

Lettis WR, Wells DL, and Baldwin JN (1997) Empirical observations regarding reverse earthquakes, blind thrust faults, and Quaternary deformation; are blind thrust faults truly blind? *Bulletin of the Seismological Society of America* 87: 1171–1198.

Lienkaemper JJ, Dawson TE, Personius SF, Seitz GG, Reidy LM, and Schwartz DP (2002) A record of large earthquakes on the southern Hayward Fault for the past 500 years. *Bulletin of the Seismological Society of America* 92(7): 2637–2658.

Liu J, Klinger Y, Sieh K, and Rubin C (2004) Six similar sequential ruptures of the San Andreas fault, Carrizo Plain, California. *Geology* 32: 649–652.

Liu-Zeng J, Klinger Y, Sieh K, Rubin C, and Seitz G (2006) Serial ruptures of the San Andreas fault, Carrizo Plain, California, revealed by three-dimensional excavations. *Journal of Geophysical Research* 111: B02306 (doi:10.1029/2004JB003601).

Marco S, Stein M, Agnon A, and Ron H (1996) Long term earthquake clustering: A 50,000 year paleoseismic record in the Dead Sea graben. *Journal of Geophysical Research* 101: 6179–6191.

Matsuda T, Ota Y, Ando M, and Yonekura N (1978) Fault mechanism and recurrence time of major earthquakes in southern Kanto district, Japan, as deduced from coastal terrace data. *Geological Society of America, Bulletin* 89: 1610–1618.

McCalpin JP (*ed*.) (1996) *'Paleoseismology'*. San Diego, CA: Academic Press.

McCalpin JP and Nelson AR (1996) Introduction to paleoseismology. In: McCalpin JP (ed.) *Paleoseismology*, pp. 1–32. San Diego, CA: Academic Press.

McGill S, Dergham S, Barton K, *et al.* (2002) Paleoseismology of the San Andreas Fault at Plunge Creek, near San Bernardino, southern California. *Bulletin of the Seismological Society of America* 92(7): 2803–2840.

McGill SF and Sieh K (1991) Surficial offsets on the central and eastern Garlock Fault associated with prehistoric

earthquakes. *Journal of Geophysical Research* 96: 21597–21621.
McGill SF and Rubin CM (1999) Surficial slip distribution on the central Emerson fault during the June 28, 1992, Landers earthquake, California. *Journal of Geophysical Research* 104(B3): 4811–4833.
Mualchin L (2005) Seismic hazard analysis for critical infrastructure in California. *Engineering Geology* 79(3–4): 177–184.
Natawidjaja DH, Sieh K, Chlieh M, *et al.* (2006) Source parameters of the great Sumatran megathrust earthquakes of 1797 and 1833 inferred from coral microatolls. *Journal of Geophysical Research Solid Earth* 111(B6): B06403.
Natawidjaja DH, Sieh K, Ward SN, *et al.* (2004) Paleogeodetic records of seismic and aseismic subduction from central Sumatran microatolls, Indonesia. *Journal of Geophysical Research Solid Earth* 109(B4): B04306.
Nelson AR, Atwater BF, Bobrowsky PT, *et al.* (1995) Radiocarbon evidence for extensive plate-boundary rupture about 300 years ago at the Cascadia subduction zone. *Nature* 378: 371–374.
Niemi TM and Hall NT (1992) Late Holocene slip rate and recurrence of great earthquakes on the San Andreas fault in Northern California. *Geology* 20(3): 195–198.
Noller JS, Sowers JM, Colman SM, and Pierce KL (2000a) Introduction to quaternary geochronology. In: Noller JS, Sowers JM, and Lettis WR (eds.) *AGU Reference Shelf Series 4: Quaternary Geochronology: Methods and Applications*, pp. 1–10. Washington, DC: American Geophysical Union.
Noller JS, Sowers JM, and Lettis WR (2000b) *AGU Reference Shelf, Series 4: Quaternary Geochronology: Methods and Applications*. Washington, DC: American Geophysical Union.
Nur A (2000) Earthquakes, armageddon, and the Dead Sea scrolls. *Seismological Research Letters* 71: 261.
Obermeier SF (1996) Using liquefaction-induced features for paleoseismic analysis. In: McCalpin JP (ed.) *Paleoseismology*, pp. 331–396. San Diego, CA: Academic Press.
Obermeier SF and Dickinson SE (2000) Liquefaction evidence for the strength of ground motions resulting from late Holocene Cascadia subduction earthquakes, with emphasis on the event of 1700 A.D. *Bulletin of the Seismological Society of America* 90: 876–896.
Pantosti D and Yeats RS (1993) Paleoseismology of great earthquakes of the Late Holocene. *Annali di Geofisica* 36(3–4): 237–257.
Papadopoulos GA, Karastathis VK, Ganas A, Pavlides S, Fokaefs A, and Orfanogiannaki K (2003) The Lefkada, Ionian Sea (Greece), shock (Mw6.2) of 14 August 2003: Evidence for the characteristic earthquake from seismicity and ground failures. *Earth, Planets and Space* 55(11): 713–718.
Pavlides SB, Pantosti D, and Zhang P (1999) Earthquakes, paleoseismology and active tectonics. *Tectonophysics* Special Issue 308,1–2: vii–x.
Petersen MD and Wesnousky SG (1994) Review: Fault slip rates and earthquake histories for active faults in southern California. *Bulletin of the Seismological Society of America* 84(5): 1608–1649.
Pirazzoli PA (1991) *Elsevier Oceanography Series 58: World Atlas of Holocene Sea-Level Changes.* Amsterdam: Elsevier.
Plafker G and Rubin M (1978) Uplift history and earthquake recurrence as deduced from marine terraces on Middleton Island, Alaska. *US Geological Survey Open File Report* 78–943: 687–722.
Prentice CS, Schwartz DP, and Yeats RS (conveners) (1994). Proceedings of the Workshop on Paleoseismology, 18–22 September 1994, Marshall, California. *US Geological Survey Open File Report* 94–568: 1–210.
Reiter L (1991) *Earthquake Hazard Analysis: Issues and Insights*. New York: Columbia University Press.
Reiter L (1995) Paleoseismology – A user's perspective. In: Serva L and Slemmons DB (eds.) *Association of Engineering Geologist Special Publication, vol. 6: Perspection in Paleoseismology*, pp. 7–17. Denver, CO: Peanut Butter
Ricci Lucchi F (1995) Sedimentological indicators of paleoseismicity. In: Serva L and Slemmons DB (eds.) *Association of Engineering Geologist Special Publication, vol. 6: Perspectives in Paleoseismology*, pp. 7–17. Denver, CO: Peanut Butter
Rockwell TK, Lindvall S, Dawson T, Langridge R, and Lettis W (2002) Lateral offsets on surveyed cultural features resulting from the 1999 Izmit and Duzce earthquakes, Turkey. *Bulletin of the Seismological Society of America* 90: 1200–1236.
Rockwell TK, Lindvall S, Herzberg M, Murbach D, Dawson T, and Berger G (2000) Paleoseismology of the Johnson Valley, Kickapoo and Homestead Valley faults: Clustering of earthquakes in the eastern California shear zone. *Bulletin of the Seismological Society of America* 90(5): 1200–1236.
Rundle JB, Rundle PB, Donnellan A, *et al.* (2006) Stress transfer in earthquakes and forecasting: Inferences from numerical simulations. *Tectonophysics* 413: 109–125.
Rundle JB, Rundle PB, Donnellan A, *et al.* (2005) A simulation-based approach to forecasting the next great San Francisco earthquake. *Proeedings of the National Academy of Sciences* 102: 15363–15367.
Satake K, Shimazaki K, Tsuji Y, and Ueda Y (1996) Time and site of a giant earthquake in Cascadia inferred from Japanese tsunami records of January 1700. *Nature* 379: 246–249.
Schumm SA, Dumont JF, and Holbrook JM (2000) *Active Tectonics and Alluvial Rivers*. Cambridge: Cambridge University Press.
Schwartz DP and Coppersmith KJ (1984) Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones. *Journal of Geophysical Research* 89: 5681–5698.
Schwartz DP and Coppersmith KJ (1986) Seismic hazards: New trends in analysis using geologic data. In: Wallace RE (chairman) *Active Tectonics: Studies in Geophysics*, pp. 215–230. Washington, DC: National Academy Press.
Schwartz DP and Sibson RH (eds.) (1989) Fault Segmentation and Controls of Rupture Initiation and Termination. *US Geological Survey Open File Report 89–315*: pp. 1–447.
Schwartz DP, Pantosti D, Okumura K, Powers TJ, and Hamilton JC (1998) Paleoseismic investigations in the Santa Cruz Mountains, California: Implications for recurrence of large magnitude earthquakes on the San Andreas fault. *Journal of Geophysical Research* 103(B8): 17985–18001.
Seitz GG, Biasi GP, and Weldon RJ, II (2000) An improved paleoseismic record of the San Andreas Fault at Pitman Canyon. *AGU Reference Shelf* 4: 563–566.
Serva L and Slemmons DB (eds.) (1995) *Association of Engineering Geologist Special Publication 6 Perspectives in Paleoseismology*. Denver, CO: Peanut butter publication.
Shimazaki K and Nakata T (1980) Time-predictable recurrence model for large earthquakes. *Geophysical Research Letters* 7: 279–282.
Sibson RH (2002) Geology of the crustal earthquake source. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *Handbook of Earthquake and Engineering Seismology*, pp. 455–475. San Diego, CA: Academic Press.
Sieh KE (1978) Slip along the San Anreas fault associated with the great 1857 earthquake. *Bulletin of the Seismological Society of America* 68: 1421–1448.

Sieh K (1986) Slip rate Across the San Andreas fault and prehistoric earthquakes at Indo, California. *EOS Transactions* 67(44): 1200.
Sieh K (1996) The repetition of large-earthquake ruptures, 93: 3764–3771.
Sieh K, Jones L, Hauksson E, *et al.* (1993) Near-field investigations of the Landers earthquake sequence, April to July 1992. *Science* 260: 171–176.
Sieh KE (1981) A review of geological evidence for recurrence times for large earthquakes. In: Simpson DW and Richards PG (eds.) *Maurice Ewing Series 4: Earthquake Prediction: An International Review*, pp. 209–216. Washington, DC: American Geophysical Union.
Sieh KE and Jahns RH (1984) Holocene activity of the San Andreas fault at Wallace Creek, California. *Geological Society of American Bulletin* 95: 883–896.
Sieh KE, Stuiver M, and Brillinger D (1989) A more precise chronology of earthquakes produced by the San Andreas fault in Southern California. *Journal of Geophysical Research* 94: 603–623.
Sims JD (1994) Stream channel offset and abandonment and a 200-year average recurrence interval of earthquakes on the San Andreas fault at Phelan Creeks, Carrizo Plain, California. In *Proc. Of the Workshop on Paleoseismology*, *18-22 Sept. 1994*, *Marshall*, *California* (Prentics CS, Schwart DP, and Yeats RS (eds.)), USGS. Open-File Report 94–568, 170–172.
Slemmons DB (1995) Complications in making paleoseismic evaluations in the basin and range province, Western United States. In: Serva L and Slemmons DB (eds.) *Association of Engineering Geologist Special Publication: Perspectives in Paleoseismology*, pp. 19–34. Washington, DC: Association of Engineering Geologists.
Slemmons DB and dePolo CM (1986) Evaluation of active faulting and related hazards. In: Wallace RE (chairman) *Active Tectonics: Studies in Geophysics*, pp. 45–62. Washington, DC: National Academy Press.
Smith BR and Sandwell DT (2006) A model of the earthquake cycle along the San Andreas fault system for the past 1000 years. *Journal of Geophysical Research Solid Earth* 111(B1): B01405.
Somerville P and Moriwaki Y (2002) Seismic hazards, risk assessment, and building codes. In: Lee WHK, Kanamori H, and Jennings PC (eds.) *IASPEI International Handbook of Earthquake and Engineering Seismology*, pp. 1065–1080. San Diego, CA: Academic Press.
Stein RS and Yeats RS (1989) Hidden earthquakes. *Scientific American* 260(6): 48–57.
Stein RS, Barka AA, and Dieterich JH (1997) Proressive failure on the North Anatolian Fault since 1939 by earthquake stress triggering. *Geophysical Journal International* 128(3): 594–604.
Stein RS, Dieterich JH, and Barka AA (1996) Role of stress triggering in earthquake migration on the North Anatolian fault. *Physics and Chemistry of Earth* 21: 225–230.
Stone EM, Grant LB, and Arrowsmith JR (2002) Recent rupture history of the San Andreas fault, southeast of Cholame in the northern Carrizo Plain, California. *Bulletin of the Seismological Society of America* 93(3): 983–997.
Toké NA, Arrowsmith JR, Young JJ, and Crosby C (2006) Paleoseismic and postseismic observations of surface slip along the Parkfield segment of the San Andreas Fault. *Special issue: Special issue on the 2004 Parkfield Earthquake and Parkfield Experiment, Bulletin of the Seismological Society of America,* 96(4b): S221–S238 (doi: 10.1785/0120050809).
Trumbore SE (2000) Radiocarbon geochronology. In: Noller JS, Sowers JM, and Lettis WR (eds.) *AGU Reference Shelf, Series 4: Quaternary Geochronology: Methods and Applications*, pp. 41–60. Washington, DC: American Geophysical Union.
Vittori E, Labini SS, and Serva L (1991) Palaeoseismology; review of the state-of-the-art. *Tectonophysics* 193: 9–32.
Wallace RE (1981) Active faults, paleoseismology, and earthquake hazards in the Western United States. In: Simpson DW and Richards PG (eds.) *Maurice Ewing Series 4: Earthquake Prediction: An International Review*, pp. 209–216. Washington, DC: American Geophysical Union.
Wallace RE (1986) *Active Tectonics: Studies in Geophysics*. Washington, DC: National Academy Press.
Ward SN (1997) Dogtails versus rainbows: Synthetic earthquake rupture models as an aid in interpreting geological data. *Bulletin of the Seismological Society of America* 87: 1422–1441.
Weldon RJ, II, Fumal TE, Powers TJ, Pezzopane SK, Scharer KM, and Hamilton JC (2002) Structure and earthquake offsets on the San Andreas Fault at the Wrightwood, California, paleoseismic site. *Bulletin of the Seismological Society of America* 92(7): 2704–2725.
Weldon RJ, III, McCalpin JP, and Rockwell TK (1996) Paleoseismology of strike-slip tectonic environments. In: McCalpin JP (ed.) *Paleoseismology*, pp. 271–329. San Diego, CA: Academic Press.
Weldon R, Scharer K, Fumal T, and Biasi G (2004) Wrightwood and the earthquake cycle: What a long recurrence record tells us about how faults work. *GSA Today* 14: 4–10.
Weldon RJ, Fumal TE, Biasi GP, and Scharer KM (2005) Past and future earthquakes on the San Andreas fault. *Science* 308: 966–967.
Wells DL and Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture area, and surface displacement. *Bulletin of the Seismological Society of America* 84: 974–1002.
Wesnousky SG (1986) Earthquakes, quaternary faults, and seismic hazard in California. *Journal of Geophysical Research* 91: 12587–12631.
Wesnousky SG (1994) Gutenberg–Richter or characteristic earthquake distribution: Which one is it? *Bulletin of the Seismological Society of America* 84: 1940–1959.
Wesnousky SG (2006) Predicting the endpoints of earthquake ruptures. *Nature* 444: 358–360.
Williams P and Seitz G (2006) New insights to earthquake behavior of the southernmost San Andreas fault. *Seismological Research Letters* 77: 270.
Working Group on California Earthquake Probabilities (WGCEP) (1988) Probabilities of large earthquakes occurring in California on the San Andreas fault. *US Geological Survey Open File Report* 88–398: pp. 1–62.
Working Group on California Earthquake Probabilities (WGCEP) (1990) Probabilities of large earthquakes in the San Francisco Bay region, California. US Geological Survey Circular 1053: 1–51.
Working Group on California Earthquake Probabilities (WGCEP) (1995) Seismic hazards in southern California – probable earthquakes, 1994–2024. *Bulletin of the Seismological Society of America* 85: 379–525.
Working Group on California Earthquake Probabilities (WGCEP) (1999) Earthquake probabilities in the San Francisco Bay Region: 2000 to 2030 – A summary of findings. *US Geological Survey Open File Report* 99–517.
Yeats RS (2001) *California Earthquakes: A Survivor's Guide*, pp. 406. Corvallis, OR: Oregon State University Press.
Yeats RS and Prentice CS (1996) Introduction to special section: Paleoseismology. *Journal of Geophysical Research* 101(B3): 5847–5853.
Yeats RS, Sieh KE, and Allen CR (1997) *The Geology of Earthquakes*. New York: Oxford University Press.

Youngs RR and Coppersmith KJ (1985) Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bulletin of the Seismological Society of America* 75: 939–964.

Van der Woerd J, Klinger Y, Sieh K, Tapponier P, Ryerson FJ, and Meriaux AS (2006) Long-term slip rate of the southern San Andreas Fault from Be-10-Al-26 surface exposure dating of an offset alluvial fan. *Journal of Geophical Research Solid Earth* 111(B4): B044407.

Zachariasen J, Sieh K, Taylor FW, Edwards RL, and Hantoro WS (1999) Submergence and uplift associated with the giant 1833 Sumatran subduction earthquake: Evidence from coral microatolls. *Journal of Geophysical Research Solid Earth* 104(B1): 895–919.

Zachariasen J, Sieh K, Taylor FW, and Hantoro WS (2000) Modern vertical deformation above the Sumatra subduction zone: Paleogeodetic insights from coral microatolls. *Bulletin of the Seismological Society of America* 90: 897–913.

Zhang PZ, Mao FY, and Slemmons DB (1999) Rupture terminations and size of segment boundaries from historical earthquake ruptures in the Basin and Range Province. *Tectonophysics* 308: 37–52.

Ziony JI and Yerkes RF (1985) Evaluating earthquake and surface faulting potential. In: Ziony JI (ed.) *US Geological Survey Proof Paper 1360: Earthquake Hazards in the Los Angeles Region*, pp. 43–91. Washington, DC: US Government Printing Office.

4.20 Historical Seismicity – Archeoseismology

A. Nur, Stanford University, Stanford, CA, USA

4.20.1 Overview

4.20.1.1 Frequency–Magnitude Relation and the Historical Observation Window

Earthquakes recur. One of the fundamental observations of modern earthquake science is that earthquakes occur on faults, and that the seismicity on any fault follows a pattern of recurrence over time, called the frequency–magnitude relation.

If we sort the many small earthquakes recorded by seismographs for a given fault, including ones too small to be felt without sensitive instruments, we see a pattern emerge. There are many more small quakes than large ones. In fact, a magnitude-3 earthquake is 10 times more likely to occur than a magnitude 4, which is 10 times more likely than a 5, and so on. At first glance, this seems to imply that very large earthquakes do not occur. On the contrary, the distribution actually gives us a specific empirical relationship between earthquake magnitude and recurrence interval on a particular fault, for all quake magnitudes, up to the largest that can occur on that fault. This relationship is called the frequency–magnitude relation. If 10 magnitude-6 quakes occurred in 100 years of recorded seismicity, and one magnitude 7 occurred, we would expect 0.1 magnitude-8 quakes to occur in that same period, provided that the fault is long enough to generate an earthquake that large. In other words, one magnitude-8 earthquake should occur every 1000 years or so. It is not surprising that we have not caught one of these large quakes on our instruments, since we have only been using them for 1/10 of that time interval. However, a city that has stood near that fault for 3000 years should have experienced three such large events during that time, along with many smaller events.

As the period for which we have been able to measure earthquake epicenters and magnitudes gets longer (it now stretches more than a century), we are beginning to see larger patterns emerge in the timing of large earthquakes, at least in some regions of the world. For instance, only 3 months after the enormous Sumatra earthquake of 26 December 2004, a very large earthquake, with a magnitude estimated at 8.7 as of this writing, occurred south of the first earthquake, along a section of the plate boundary that did not slip in the first event. Ironically, some of the first detailed scientific analyses of the Great Sumatran earthquake were just appearing in journals when the second earthquake occurred, and one article, published by Seth Stein and Emile Okal in the British journal *Nature*, warned of the continued danger on this southern section of the boundary (Stein and Okal, 2005). This is because we now understand that some earthquakes can actually increase the stress on sections of the same fault, or nearby faults, that did not slip initially (King *et al.*, 1994). Thus, one earthquake can increase the chances of another earthquake nearby.

In fact, on a large fault, accumulated stress can be released in a series of large earthquakes, each one triggering another on an adjacent section of the fault a few months or years apart, so that the fault 'unzips' in steps until the stress has been released along its entire length. The whole process of gradual accumulation and subsequent release of stress then repeats, leading eventually to another series of earthquakes after centuries of quiescence. As a result, whole regions can be at significant earthquake risk, even in areas where modern seismicity is not significant.

Geophysicists have not reached a consensus on what to call this phenomenon, describing modern examples as 'earthquake sequences' (e.g., Ambraseys, 1970), 'earthquake migrations' (Mogi, 1968; Roth, 1988), 'progressive failures' (Stein *et al.*, 1997), or, in places where many intersecting faults are involved, 'earthquake storms' (Nur and Cline, 2000).

An example of such an 'earthquake sequence' took place on the North Anatolian Fault during the twentieth century (e.g., Ambraseys, 1970; Allen, 1975; Wood, 1996; Stein *et al.*, 1997). This 30-year 'sequence', which is well known and studied among geophysicists, consisted of a series of earthquakes, all with magnitudes greater than 5.6, in 1939, 1942, 1943, 1944, 1951, 1957, and 1967. The earthquakes progressed westward along the 1000-km-long fault zone, releasing strain that had accumulated over the previous 200 or more years (**Figure 1**); this 'sequence' might now be extended to 60 years in length, in light of the August and November 1999 earthquakes that continued the earlier rupture pattern. The region affected by this sequence is home to many significant archeological sites, including many cities of the Hittite empire, as well as the ruins of ancient Troy.

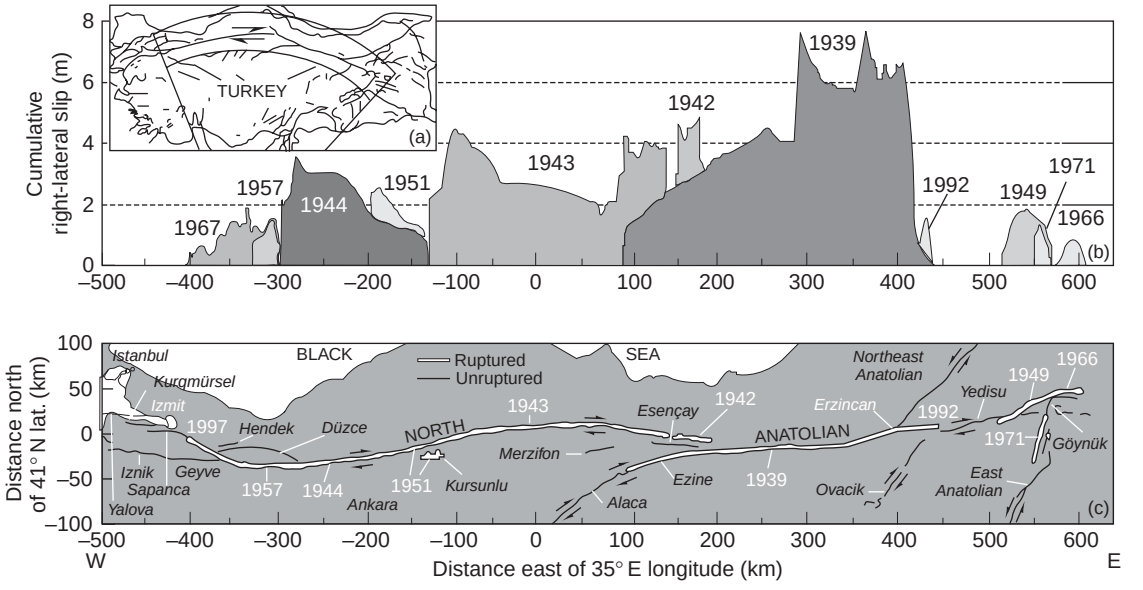


Figure 1 This map shows the locations of earthquakes that occurred in a sequence from 1937 to 1967 on the North Anatolian Fault. Reprinted from Stein R, Barka A, and Dietrich JH (1997) Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysics Journal International* 128(3): 594–604.

4.20.1.2 Limited Usefulness of History

While the precise timing, locations, and magnitudes of earthquakes cannot be predicted, and may follow complicated patterns, we can, given a sufficient history of instrumentally recorded earthquakes, estimate the mean recurrence interval for earthquakes of a given size on a particular fault. In practice, though, this technique has serious limitations, the greatest of which is the short time over which earthquakes have been instrumentally recorded. Seismometers have been in widespread use only since the early 1900s, and in many remote regions, the record is even more limited than that. Since the frequency–magnitude relation implies that the most devastating earthquakes will also be the least frequent, many regions of significant seismic hazards have not seen a major earthquake since the birth of modern seismology.

Thus, seismic hazard must often be based on crude historical data that do not include seismograms. Written records of historical earthquakes often include lists of cities where shaking was felt, or where significant damage had to be repaired. By piecing together different accounts of such damage, we can crudely estimate the size and location of past earthquakes (Guidoboni *et al.*, 1994). But written records are also a relatively recent thing, and the reliability of the oldest of these is suspect. Lists of affected areas are biased toward cities of political or religious importance, and may entirely omit areas that were less populated or more isolated. Written records are also unlikely to survive for smaller earthquakes, which can be very destructive locally.

Another problem with historical earthquake records is that it can be difficult to determine which fault was responsible for an ancient earthquake. Estimates of damage were very subjective, and were also influenced by politics and population density. Some earthquakes may well have happened on faults that have not yet been mapped, and may represent completely unanticipated seismic hazard zones.

How then, besides waiting for more earthquakes to occur, can we extend our observation window to gain a more complete understanding of seismic risk? One of the most promising avenues for collaborative research is the relatively new field of archeoseismology, the study of ancient earthquake damage using archeological methods. This chapter will discuss the current state of this fledgling field, and explore some of the obstacles scientists face in achieving a true synthesis of these two very different disciplines.

4.20.1.3 Effects of Tectonics on Sites of Human Occupation, Resonant Effects, and Vulnerability of Tells

Many sites of ancient human habitation are located in areas of significant seismic risk. The Middle East, Turkey, Greece, and Rome are renowned for their early cultures and often-magnificent ruins, and all are areas where earthquakes are common. The Americas, particularly south of the United States, had their own early urban cultures, many of which were in seismically active areas.

In fact, some ancient cities owe their locations in part to the effects of earthquakes. For example, the ancient walls of Mycenae in Greece were built upon a fault scarp (Maroukian *et al.*, 1996), which afforded a steep, slick, natural foundation for the city's defenses (**Figure 2**). The builders of Mycenae were almost certainly unaware that this landform indicated a significant earthquake hazard, since the idea that earthquakes occur on faults is a recent one.

Figure 2 The defensive walls of ancient Mycenae were built atop a fault scarp, visible in this picture as the steeply sloped, light-gray base of the wall.

The locations of other ancient cities are more subtly related to earthquakes, because the strategic value of earthquake-generated topography is a major factor that determines routes of trade and migration; cities that spring up along those routes may be subject to hazards from the same earthquakes that are responsible for their strategic importance. An example of this is Megiddo, in Israel (Aharoni, 1993; Kempinski, 1993). This ancient city has been a strategic battle ground for millenniums, and, as recently as 1918, during the British mandate years. Accumulated motion on the Carmel–Gilboa fault system in this area resulted in a natural bottleneck in trade routes between Syria and the Mediterranean; Megiddo guards the major mountain pass through the Carmel Mountains, and thus has maintained political importance from the fifteenth century BC through the twentieth century AD. Multiple layers of destruction and rebuilding at this one site have left an artificial hill, called a tell by archeologists. Some of the destruction events preserved in this tell were certainly the results of military action. However, the tectonic environment of the site makes it almost inescapable that some of the destruction is from ancient earthquakes (Nur and Ron, 1997).

Tells are diagnostic of important ancient sites in the Mediterranean region, and although the names and significance of many ancient cities have been lost to time, their mounds still stand, waiting to be excavated, preserving archeological and geophysical secrets for future scientists to discover. The very nature of these tells makes them interesting to geophysicists for other reasons. The structure of the tells, with successive layers of construction built upon the unconsolidated rubble of previous layers, makes them particularly susceptible to earthquake damage, because of the weakness of unconsolidated fill when shaken. This is one of the effects that caused damage in the 1989 Loma Prieta, California, earthquake to be much greater in the San Francisco Marina district than in regions much closer to the epicenter. The damage in the Marina District was exacerbated by a high water table, resulting in saturated soil that could easily liquefy. This is less common in the arid Mediterranean region, but can occur in coastal or riverside areas, or in areas with natural springs.

Another phenomenon that tells exhibit during earthquakes is topography-related resonance when exposed to seismic waves, an effect that has been noted in modern earthquakes (Spudich *et al.*, 1996). For example, during the Northridge, California, earthquake of 1994, 21 three-component geophones recorded the pattern of ground motion atop an elongated hill in Tarzana, California (**Figure 3**). Ground motion was generally much higher at the top of the hill than at the bottom of the hill and in the surrounding countryside. This pattern suggests that certain wavelengths of seismic energy can be focused and concentrated by the shape of the topography, creating a resonant effect that greatly increases the seismic hazard at the top of the hill. Thus, the resonant effects of the artificial topography, combined with the earthquake susceptibility of loose rubble foundations, may cause considerably more damage in tells than in the unaltered surrounding countryside, particularly when the tells are noncircular, as is true in most cases. This makes archeoseismology potentially even more useful than paleoseismology in natural terrain.

4.20.1.4 Communication Issues between Disciplines

The differences between the damage levels in tells and in unaltered terrain environments are particularly important, because many currently inhabited cities in the Mediterranean region are built the same way the tells were – upon the rubble of past civilizations. Thus, modern cities in ancient regions share this extreme vulnerability to earthquakes. Unfortunately, many archeologists who work in seismic hazard zones are completely unaware of these facts.

This is not surprising, for as far as I know, earthquake seismology is not a requirement of any archeology degree program. Thus the level of earthquake awareness among archeologists is not much higher than that of the general educated public. Archeology and geology share many characteristics, including the use of relative dating based on stratigraphy. However, archeologists are not trained to recognize earthquake damage, or even to know what sorts of damage can occur during an earthquake. Those who have never personally experienced an earthquake can have particularly unrealistic ideas about what earthquakes leave behind, expecting that they should find ground fissures or displaced faults cutting through the rubble if there has been any earthquake damage. When excavating, some archeologists do not consider the possibility of earthquakes at all, and most of the rest consider earthquake damage to be an explanation of last resort when determining the cause of destruction in an ancient city.

Such reluctance is illogical, since we know that earthquakes remain an ever-present threat in any

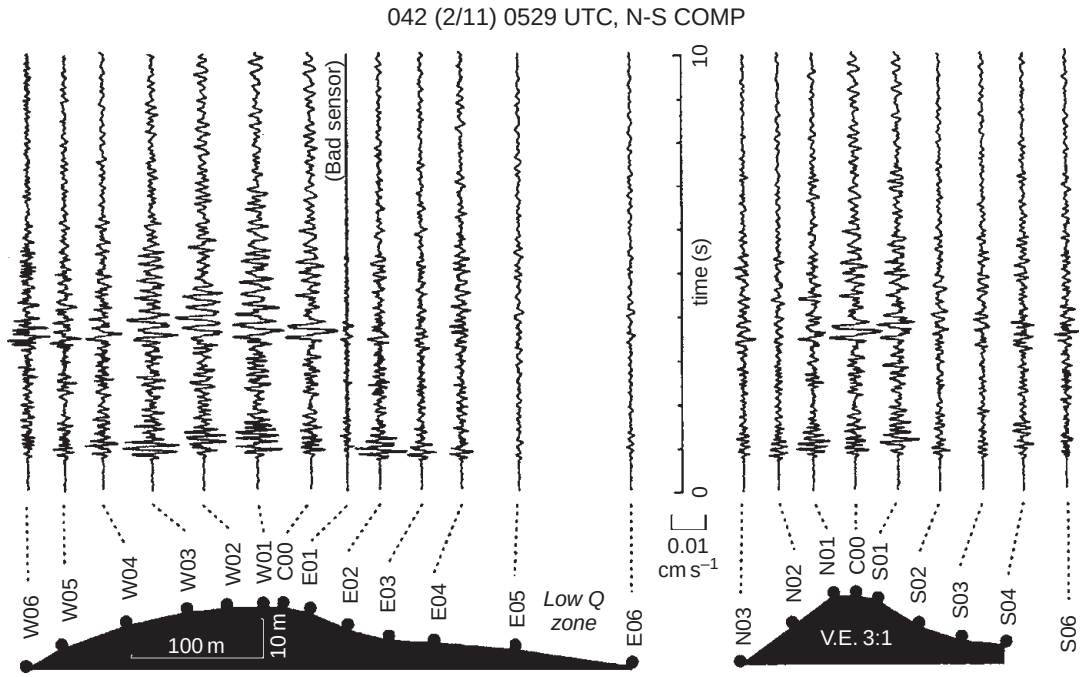


Figure 3 This is a plot of ground motion recorded on Tarzana hill by three-component geophones. Note that the shaking is much more severe atop the hill than on lower ground. This indicates a topographic resonance effect that is likely to contribute to increased damage atop tells. Reprinted from Spudich P, Hellweg M, and Lee WHK (1996) Directional topographic site response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake: Implications for mainshock motions. *Bulletin of Seismological Society of America* 86(1B): S193–S208.

seismic hazard zone; in fact, if a city has existed in an earthquake-hazard zone for a significant period of prehistory, it is nearly inescapable that it should have experienced significant shaking. Earthquakes should be one of the primary suspects whenever a demolished city is found, unless there is specific evidence of human agents in the destruction. Even when evidence is found of human conflict amid the ruins, the possibility should be considered that enemies attacked opportunistically after an earthquake.

On the other hand, earthquake scientists are often ignorant of the conventions of archeology. They may be unaware of the complex social and military interactions that have dominated a region, and since they see earthquakes as a scientific inevitability, they are certainly biased toward earthquake damage as an explanation for destruction. But perhaps a greater difficulty is that the jargon of seismology and the jargon of archeology are different languages that often give different meanings to the same words. For instance, the word 'epicenter', used in seismology, means the precise spot on the earth's surface that is directly above the underground focus of an earthquake. However, the word is sometimes used loosely by archeologists to mean the center of a region of destruction, with multiple epicenters sometimes being cited for the same earthquake. Archeoseismologists need to learn some of the jargon of archeology, so that they can communicate the science behind earthquakes in a less ambiguous way. Then it will be more possible to collaborate with archeologists in uncovering traces of ancient earthquakes.

But what signs do earthquakes leave behind? The sections that follow examine some of the features of earthquake damage in archeological remains, which archeologists can document with traditional methods. They also explore special research methods that can complement traditional archeology, and which have had success in recent years resolving the prehistoric earthquake record in the Middle East.

4.20.2 Earthquake Traces in Archeological Sites

It often can be very difficult to distinguish earthquake damage in archeological sites from damage caused by slow decay over time, or damage caused

by human hands. This section will describe the various traces earthquakes leave behind. Documenting many of these types of damage from a single site during a single habitation level may implicate an earthquake as the most likely cause of destruction, even in the absence of historical data.

4.20.2.1 Effects of Earthquakes on Landforms

Perhaps the only unequivocal evidence for earthquakes is the presence of large-scale land deformation, associated either with faulting or wholesale rotation or displacement of the Earth's surface. There are several cases where such evidence has been uncovered in the archeological record, but such cases are rare. However, this does not indicate, as some archeologists have asserted, that earthquake damage is rare; most damage in earthquakes is associated with ground shaking, which may occur quite far from faults, and may leave no evidence (beyond collapsed structures) after the shaking stops. Nevertheless, the dramatic nature of such land deformation makes it excellent evidence for earthquakes, when it is found.

4.20.2.1.1 Fault displacement

It is well known that earthquakes are the result of sudden displacement on faults in the Earth's crust. When earthquakes are large enough and shallow enough, that fault displacement may break the earth's surface, resulting in ground displacement, which may cut through any structures that straddle the fault trace. The fault scarp described beneath the Mycenaean walls is unmistakable evidence that large earthquakes have occurred repeatedly in Mycenae. However, the fact that the wall was built atop the scarp only tells us that the earthquakes that formed the scarp predate the wall, and that later earthquakes that activated the fault were insufficient to topple that wall (or the wall was subsequently repaired).

In tells, it is sometimes possible to bracket an earthquake between two successive layers of occupation. If a fault cuts and displaces the remains of one wall, but an undeformed wall has been built over the fault in the next layer, we know that the earthquake happened some time between the building of the first wall and the building of the second.

An example of this is from the Crusader fortification of Ateret, by the banks of the Jordan River in northern Israel. The foundations of the fortress' walls have been displaced horizontally by motion on a fault that cuts through the walls; thus, the displacement must have occurred after the walls' construction in 1178 (and destruction by Saladin a year later). An early-modern Muslim structure built within the ruins is also displaced, but whereas the Crusader walls are displaced by 2.1 m across the fault, the walls of the Muslim structure (dated somewhere during the Turkish Ottoman period of 1517–1917) are displaced by only 0.5 m. Trenches dug across the zone of deformation indicate that the slip occurred during several earthquakes (Ellenblum *et al.*, 1998). By combining trench data with current geodetic measurements to rule out aseismic slip, Ellenblum *et al.* (1998) were able to determine that this fault is currently locked, and that the displacement that has accumulated in the ruins has been accommodated by at least two large earthquakes. Thus their research has contributed information toward the seismic risk estimates in the region, a region that includes heavily populated areas of Jordan, Israel, and Syria.

4.20.2.1.2 Elevated or submerged coastal land surfaces

While the example above was an exciting and valuable contribution, most ruins are not directly bisected by active fault traces. Horizontal displacement is only detectable when a continuous wall or other feature can be traced across the fault surface. However, when a town or structure was built on an ancient shoreline, any vertical displacement of the structures can be measured, even if they are far from the fault trace itself. There are many cases where archeology has posited that a port or coastal fort has been destroyed during an earthquake. For example, the town of Dioscura on the shore of the Black Sea shows evidence that it was destroyed by earthquakes in the first century AD (Nikonov, 1996). The external walls of the city have been found slightly tilted, partially collapsed, and submerged in Suchumi Bay. Archeological evidence was used to argue against gradual sea-level rise being responsible for this apparent subsidence, and geologic stratigraphy of the bay-bottom sediments was used to discount the possibility of a large landslide that moved the city into the bay. This is an excellent example of the use of interdisciplinary collaboration to implicate a major earthquake as the cause of this destruction, and it is particularly important in this region. Because local seismicity has been relatively quiet in modern times, the seismic hazard in this region has been estimated to be low to moderate. However, if Nikonov is

correct, a very large earthquake, with Mercalli intensities of IX to X, has occurred in this region, and therefore is likely to occur again. Additional paleoseismic and archeoseismic work in the area is necessary to confirm or discount this hypothesis, one that has important consequences for the region's current inhabitants.

Submerged and uplifted coastal settlements have been more thoroughly studied in Greece, where a prolific ancient building tradition and a known history of earthquakes has made an ideal workshop for archeoseismology. There have been many studies of apparent changes in coastal elevation in this region (e.g., Pirazzoli *et al.*, 1992; Papageorgiou and Stiros, 1996; Pirazzoli, 1996), and the employment of geostratigraphy has been particularly rigorous. By using fossil evidence, especially from fragile organisms unlikely to survive a significant period of wave erosion, researchers can rule out mechanisms of gradual deformation or sea-level change in favor of sudden uplift or subsidence by earthquakes.

4.20.2.1.3 Tilting of horizontal surfaces

Besides displacing surfaces horizontally or vertically, fault motion can act to rotate surfaces, with the result that previously horizontal layers in the ground acquire an increasing dip over time. An example of this is seen in a small pull-apart basin that is part of the Dead Sea fault system in Israel. Located in this basin is Ubeidiya, thought by many to be the oldest known hominid construction on earth (Bar Yosef, 1993). The archeological structure in this case is the simplest possible (a floor made of pebbles), put there by the appropriately named 'Pebble Man', thought to be a member of the species *Homo erectus*. Some 1.5 million years ago, this floor was made by pressing pebbles, parts of stone artifacts, and even bones of now-extinct mammals into what was then the muddy bank of the proto-Jordan River.

When it was first excavated, it was not immediately clear what the surface was, since it is now inclined 60° from the horizontal. Geologists and archeologists collaborating eventually realized that the surface had been horizontal, or nearly so, when it was built. The present tilt of the floor is due to its rotation about a horizontal axis during the 1.5 My since it was created, rotation due to the growth of the pull-apart basin, which in turn was caused by motion on the Dead Sea Transform and associated faults. Since modern data suggest that large earthquakes recur on average every 400 years or so in this area, some 3000–4000 such events must have occurred, each tilting the floor on average by the miniscule amount of just two-hundredths of a degree.

Rotation can also tilt vertical structures out of true or change the azimuths of walls. However, these two cases would be harder to detect in archeological settings, since in the first case, nonvertical walls tend to collapse, and in the second, the original alignment of a wall is generally unknown.

4.20.2.2 Architectural Damage

The features described in the previous section are extreme, exciting, and generally strong indicators of earthquake damage. However, such ground deformation occurs only in the largest earthquakes, and usually only relatively close to the responsible fault. Most damage in earthquake occurs because of the transient shaking of the ground, which leaves permanent evidence only on the above-ground structures, not in the ground itself. Thus, architectural damage, without associated fault traces or ground deformation, is what we should expect to find most often in archeological excavations of earthquake-damaged cities. The challenge is to differentiate between earthquake destruction and either slow decay over time or sudden destruction by human action. There are several types of architectural damage that are suggestive of earthquake damage. When several of these signs are present in a single excavation, without contraindications that would implicate human destruction, earthquake should be a primary candidate for the destructive agent.

The archeological tradition of carefully noting the location of every artifact found goes a long way toward making archeological data useful to seismologists. However, when archeologists are aware of the possibility of earthquakes, and know the importance of directional alignment, angle, and regional patterns of collapse in interpreting earthquake damage, there is less chance that crucial data will be lost as the excavated fragments are removed for cataloging or restored to their original positions.

4.20.2.2.1 Collapsed walls with preferential axes of destruction

When walls are destroyed by invading armies, they tend to be destroyed haphazardly. Earthquakes, particularly smaller ones, can also cause this type of chaotic wall collapse, if the walls crumble slowly during an extended period of shaking. However, strong earthquakes often topple walls with the first arrival of the damaging shear waves, in which case

Figure 4 This view of the reconstructed storerooms of Masada shows the long, unsupported walls that are susceptible to earthquake damage.

Figure 5 This picture, taken before the reconstruction of Masada, shows how the walls of the parallel storerooms collapsed into the adjacent rooms, leaving rows of bricks that are diagnostic of earthquake collapse. From Masada: Herod's Fortress and Zealot's Last Stand by Yigael Yadin, © 1966 by Yigael Yadin. Used by permission of Random House, Inc.

the ruins will show a preferential direction of collapse. We find such damage in many different sites, including some layers at Troy, Jericho, and Mycenae. The ruins of storerooms at Masada in Israel illustrate this effect particularly well.

At the end of the Jewish war against the Romans in AD 70, the stronghold of Masada, which towers over the Dead Sea in the Judean desert, became the last free Jewish community in Judea (Yadin, 1966). It was eventually conquered by the Roman army, but its buildings were not destroyed until much later. There is strong evidence that this destruction happened in an earthquake, perhaps one that destroyed other sites in the area in AD 363. One of the things that made it possible for Masada's people to withstand the Roman siege for as long as they did was their elaborate system of storerooms (**Figure 4**). The storehouses were designed as a series of long, narrow rooms, all parallel to each other, with walls of unreinforced stone 11 ft high. This type of construction is particularly vulnerable to shaking perpendicular to the long direction of the storerooms, since the long walls have no bracing along most of their length. The shorter walls are braced at every intersection with the long walls, at least until the long walls collapse.

The ruins at Masada show that the storehouse walls apparently collapsed as a single unit in one direction, with each wall falling into the storeroom next to it. The rows of stones seen in the aerial photo in **Figure 5**, which was taken before excavation (and restoration of the walls) began, are the kind of marker that we have come to look for as evidence of an earthquake.

4.20.2.2.2 Aligned fallen columns

Not all building styles incorporate columns, but those that do are goldmines for archeoseismology. Perched on their narrow bases, columns are very vulnerable to earthquake damage. In Greek designs, these columns were often supported on top only by the next course of stone, and in Roman construction only by light wooden structures. Thus, when such columns fell in strong earthquakes, they often fell together in a very distinctive way, as in the ruins of Susita in northern Israel, shown in **Figure 6**. The most striking feature here is that all the columns are aligned in the same direction, indicating that they fell in response to single impulse. The impulse here was the AD 749 earthquake that hit the Sea of Galilee region. Since the bases of the columns are not tilted, which could indicate a slow tilting followed by collapse, this damage almost certainly happened during

Figure 6 These columns were discovered in the ruins of Susita, overlooking the Sea of Galilee. Note how the columns all fell in approximately the same direction, yet they are no longer aligned with their bases. These are signs that the columns collapsed in an earthquake, probably one that occurred in AD 749.

an earthquake. Had the columns been destroyed by human action, it would have been very difficult for the invaders to achieve such precise alignment. Examples of such systematic column collapse are widespread, including Knidos in southwestern Turkey, Jerash, Petra, and Pella in Jordan, Betshean in Israel, Chichén Itzá in Mexico, and Selinunte in Sicily, where the association with an earthquake is historically documented.

Another, more subtle feature of **Figure 6** is the fact that the bases of the columns do not line up with the direction of collapse. Instead, the lines that connect the bases to the columns that were perched on them are all parallel to one another in a slightly different direction. There are two possible explanations for this: either the columns were bucked sideways off their bases and then fell during subsequent shaking, or they fell and then rolled out of alignment with the bases during subsequent shaking.

4.20.2.2.3 Slipped keystones

Sometimes an earthquake does not topple a wall, but only partly damages it. In this case, the remains of the wall preserves telltale signs of earthquake action. Arches are particularly good places to look for this evidence.

Arches are used to allow large openings in stone walls while still supporting the weight of stone above. The key to the strength of this structure, which could be built without mortar or pins, was allowing the weight of the stones to compress the arch, pushing each stone firmly against its neighbor. Unless a stone completely disintegrated under the pressure, a well-designed arch would stand indefinitely under its own weight. However, if a single of the wedge-shaped stones, or 'voussoirs', were removed from the perimeter of the arch, the entire structure could collapse.

In ordinary circumstances, this poses no problem. The immense weight of a masonry structure holds the stones firmly in place. However, in the extraordinary circumstance of an earthquake, the earth's motion wields the structure's own weight against it, effectively throwing the whole edifice around. Moderate and large earthquakes can result in very high ground acceleration, occasionally even exceeding the acceleration of gravity, particularly in areas where the local topography or geology amplifies ground shaking (as in tells). In other words, an earthquake can effectively cancel gravity, which would obviously have disastrous effects for gravity-dependent structures. However, even when the motion is less severe, structures are subjected to both strong horizontal forces and vertical forces acting against gravity.

The key to the weakness of arch-based architecture is the ease with which the base of the arch can slide horizontally, especially if partially deprived of gravitational compression. Typically, even in the second half of the twentieth century, the bases of arches were not pinned in any way to the structures beneath them. This means that, under the enormous horizontal forces generated in an earthquake, the base of the arch can simply slip sideways on its foundation, relieving the compression

Figure 7 These arches are from Bet Shean, 45 km south of Susita, shown in **Figure 6**. These arches collapsed out into the street, probably in the same earthquake as destroyed Susita in AD 749.

that holds the arch together, and causing total collapse. Because there was no concept of seismic hazard in ancient times, that is, no idea that earthquakes could and should be expected to repeat, and no way to grasp the huge forces involved in earthquakes, no thought was given to pinning the structures against horizontal motion as an antiseismic precaution. In fact, when exactly such earthquake damage occurred at Stanford University during the 1906 San Francisco Earthquake, the displacement was simply corrected without pinning the arches' bases; as a result, the same arches were damaged again in the 1989 Loma Prieta quake.

Most arches that were hit by sufficiently large past earthquakes simply collapsed, and all we find in the excavated rubble is a heap of stones that made up the arch. Unless the whole wall toppled neatly outward, like in **Figure 7**, the chaotic remains of collapsed arches are not very useful indicators of earthquake damage. They are too difficult to distinguish from collapse due to slower earth movements or destruction by violence. However, earthquakes sometimes stop short of collapsing the building, and one or more of the voussoirs in an arch drops down until it is again pressed tight against its neighbors.

Figure 8 shows an example of this from Kala'at Namrud, an Arab and Crusader fortification on the main road from Galilee to Damascus. On the left side, indicated by the arrows in the photograph, the whole base of the doorway slipped sideways, loosening the stones in the arch. Subsequently, several of the voussoirs in the arch dropped down, and the arch was again under compression, leaving the doorway a bit wider and shorter than the original architect had intended.

Another example can still be seen in Rome, in the Colosseum, which was partly destroyed by an earthquake in 1349. **Figure 9** shows some of the arches that remain. The arches that would once have stood to the left of these collapsed in the destructive earthquake, but the leftmost one in this photo did not. The movement was arrested just in time, as we can see in the close-up. The voussoirs slipped only partially, and the arch survived.

4.20.2.3 Skeletons of Trapped Humans and Domestic Animals

One of the challenges in distinguishing earthquake damage from other sorts of damage is the difficulty of showing that the collapse was sudden and unexpected. It is a sure sign of this when skeletons of humans or domestic animals are found trapped beneath collapsed walls, especially when they are found in defensive postures. Trapped skeletons have been found in archeological sites around the world, including Mycenae, Tiryns, Midea, Thebes, Crete, and the Menelaion in Greece (Mylonas, 1966; Mylonas-Shear, 1969, 1987; Kilian, 1996; Åström and Demakopoulou, 1996; Sampson, 1996; Sakellarakis and Sapouna-Sakellaraki, 1981; Catling 1981), and Dor, Petra, Jerash, and Pella in the Middle East (Stewart, 1993; Stucky, 1990; Amiran *et al.*, 1994; Browning, 1982).

Figure 8 This is an example of a slipped voussoir in the Crusader fortification of Kala'at Namrud. Note how base of one side of the arch slipped outward, allowing the voussoir to drop.

The oldest known examples of skeletal remains of earthquake victims come from Shanidar Cave in Iraq. In that cave, there are multiple layers in which Neanderthal skeletons are crushed under boulders that fell from the ceiling of the cave, presumably in earthquakes. Interestingly, the excavator of the cave, Ralph Solecki, experienced an earthquake inside the cave while excavating (Solecki, 1971).

The ancient port of Kourion, located on the southern coast of Cyprus, has revealed compelling evidence of earthquake victims. In 1934, archeologist J. F. Daniel unearthed two crushed human skeletons dubbed 'Romeo and Juliet' beneath fallen rubble of their house, which also held coins dated up to AD 365 and crushed wares (Soren, 1988a). The early 1980s excavation by David Soren found several human skeletons crushed by falling walls and roofs of separate buildings. Some skeletons were curled into fetal postures. A room revealed a man in his thirties, crushed under rubble with his hands spread to cover his head. Another room, apparently a stable, held the bones of a mule still tethered to a stone trough. Intertwined with the animal's skeleton is that of a 13-year-old girl, caught perhaps trying to calm her frightened beast. One remarkable find is of a young family of three (**Figure 10**), trapped under the ruins of their bedroom (Soren, 1988a, 1988b). These archeological findings are associated with the historical earthquake of AD 365.

4.20.2.4 Fire

Fire is one of the types of damage that archeologists generally attribute to invaders. This is because archeologists and the general public have a misconception about the firestorms that often follow severe earthquakes. In modern times, such fires are usually caused by ruptured natural gas lines and downed

Figure 9 These arches are from the outer wall of the Colosseum in Rome. Once, there were more arches to the left of the leftmost one in this photo, but they collapsed in an earthquake. Note the chaotic state of the keystones in the leftmost remaining arch.

Figure 10 Kourion, Cyprus: A young couple and their 18-month-old child crushed under blocks from the ceiling of their bedroom, some weighing as much as 300 pounds. Soren D (1988b) The day the world ended at Kourion: Reconstructing an ancient earthquake. *National Geographic* 174(1): 30–53, reprinted with permission. Credit: Martha Cooper/National Geographic Image Collection.

electrical wires, so the popular view is that such fires would not have occurred in antiquity, before the advent of such modern hazards. However, in antiquity, heating, cooking, and lighting were all accomplished by open flame, which in itself is a significant hazard. In an earthquake, fires burning in a hearth or lamp can be upset or scattered into a dwelling, where combustible materials can ignite.

Another argument against fires following ancient earthquakes is that the building materials used in ancient buildings (stone, mud brick, and tile) were not combustible. This argument is illogical; if the buildings were so fireproof that an earthquake could not ignite them, then how did the enemies manage to do so? Did enemies haul in all the fuel necessary to burn an entire settlement?

While it is true that the remains we find in archeological sites are generally incombustible, it is not true that combustible materials were not used in ancient buildings; it is just that combustible materials do not generally survive milleniums of burial. Roof support timbers were often used in all but the grandest of stone structures, and furnishings were generally made of wood and cloth. And the fact is, there are plenty of known cases of fires following earthquakes that predate modern gas and electric utilities. The fire that destroyed Lisbon, Portugal, in 1755 (British Historical Society of Lisbon, 1990) was one of the most dramatic examples; it burned for 6 days after the earthquake, and was responsible for the loss of uncounted treasures of art and history in that city.

Therefore, evidence of widespread fire in an archeological site is just as likely to be due to an earthquake as to human action, and certainly should not be used as an argument against earthquake destruction.

4.20.3 Techniques in Archeoseismology

The field of archeoseismology is a relatively new one, and it is by nature highly collaborative. Despite the difficulties in cross-discipline communication, scientists from the fields of geology, geophysics, archeology, paleontology, and even forensic anthropology have made great strides toward integrating their varied perspectives into a coherent approach to ancient earthquakes. There are several techniques that have been employed to that end.

4.20.3.1 Informing Archeological Methods with a Seismology Perspective

The first and most important of archeoseismological techniques is simply one of disseminating knowledge and changing perspectives. The standard techniques of archeology are ideally suited to reconstructing an earthquake scenario, as long as the archeologist is aware of the types of data that seismologists need. Not only are locations of artifacts important, but also their orientations, and the structure of the soil in which they are buried. These are data that archeologists can record, and often do, in the course of a normal excavation. However, there are instances where specialized expertise is required, which may warrant the involvement of an outside expert.

4.20.3.2 Intersection of Archeology and Paleoseismology

One case where additional expertise may be necessary is when there is evidence of a fault structure cutting through an archeological site. In this case, the involvement of paleoseismology can make an excavation much more informative.

Paleoseismologists specialize in recognizing traces of ancient earthquakes in subsurface deposits, using trenching methods to examine layers and how they have been displaced relative to each other. They are familiar with how earthquakes disturb sediments, and may be able to pinpoint relative dates for earthquakes more accurately than an archeologist who is not familiar with examining earthquake traces. Paleoseismologists are also more likely to make use of microscopic biological or paleontological data than are archeologists, who are more likely to use pottery and other cultural indicators, and thus may be able to produce an independent relative date for comparison with the archeological date.

Another paleoseismological method that has shown promise for informing archeological digs is the analysis of sea- or lake-bottom sediments for evidence and dating of earthquakes. When an earthquake shakes the water-saturated, unconsolidated sediments on a sea or lake floor, it creates layers of chaotic, mixed sediment called seismites. In an analysis of Dead Sea seismites (Ken-Tor *et al.*, 2001), radiocarbon dating of organic matter in the disturbed layers allowed seismites to be correlated to eight historical earthquakes, dating from 64 BC to AD 1927. In the Dead Sea region, where the historical record is very long, such correlation can help scientists estimate the size of these ancient earthquakes and determine what archeological sites might have been affected around the Dead Sea; it can give clues to possible earthquake epicenters as well. In a region without such long records, seismite information might reveal entirely unknown seismic hazards.

4.20.3.3 Forensic Anthropology

Another case where outside expertise is indicated is in determining the cause of death when a skeleton is found in the ruins. In the case of skeletons found under collapsed walls, it may be crucial to determine whether the victim was trapped by falling debris in an earthquake, or died by other means and was left, with its skeleton eventually being buried in the slow decay of the building. The field of forensic anthropology is uniquely equipped to deal with these sorts of questions, using medical knowledge of how fresh bone fractures differently than dry bone, and how the presence of soft tissues around and inside bone can change how it responds to sudden force, burial, or fire.

The discovery of a female skeleton buried beneath by a wall in the coastal Israeli town of Dor prompted archeologist Andrew Stewart (Stewart, 1993) to consult an expert in bone fracture analysis to determine whether the falling wall could have killed her, or whether she might have been already dead when the wall fell. While the distinction cannot be made with that precision, Dr. Patricia Smith at Hadassah Medical Center in Jerusalem found evidence of a particular kind of breakage, a spiral fracture, that only occurs in fresh bone. This made it clear that she was either alive, or fairly freshly dead when the falling wall broke her bones. Further evidence can be obtained by reconstructing the victim's position to determine whether the fractures occurred *in situ*, or while the victim was standing erect. These are generally determinations that an archeologist or geologist is not qualified to make.

Another instance where forensic knowledge was able to extract more data from a site than archeology or geology could do alone is the excavation of a Minoan temple from about 1700 BC in Arkhanes, Crete. Excavators of this site found several skeletons within the collapsed building, but in this case, they uncovered a particularly shocking scene. As the main room of the temple was excavated, archeologists discovered a priest beside an altar, with his sacrificial victim still bound before him. It had been thought that altars from this period were used for the sacrifice of bulls, not of young men as was the case in this scene.

From the archeological evidence, either during the ceremony, or immediately before or after the sacrificial act, an earthquake struck, collapsing the building. The earthquake was followed by an intense fire, which charred the entire site, including the bones of the people trapped within. Forensic study of the bones on the altar was able to answer the timing question that the archeology alone could not resolve.

According to the forensic anthropologist, bones that are burned in the presence of blood turn black, while if the corpse is drained of blood, the bones would remain white. In the sacrificial victim, who was lying on his side, the bones in the uppermost

side of his body were white, while the lower bones were black. This indicates that much of his blood had been drained before his body burned (Sakellarakis and Sapouna-Sakellaraki, 1981).

4.20.3.4 Numerical Simulation

One valuable piece of information that scientists hope to gain from archeoseismology is a magnitude estimate for the earthquake that destroyed the site in question. There are many factors that determine the damage caused by an earthquake, including the distance from the epicenter, the depth of the earthquake's focus, the duration of shaking, the structure of the subsurface, and the type of architecture. All of these factors combine to create a seismic intensity, which is simply a numerical description of the damage level.

Given an archeological site with, say, huge collapsed columns, determining the amount of shaking required to topple the columns is not an archeological or seismological problem, but rather an engineering problem. Numerically simulating the effects of shaking on engineering structures is a complex problem, but it can yield some information about possible earthquake magnitudes (e.g., Papastamatiou and Psycharis, 1996).

Other simulations can reveal how the structure of subsurface sediments can intensify shaking during an earthquake. This technique was used to show why the damage to Rome's Colosseum in the 1349 earthquake was so much more severe on its southern edge than in the rest of the structure. The entire southern half of the exterior collapsed, while the rest remains standing to this day (**Figure 11**). Studies of the ground beneath the Colosseum show that the southern portion of the structure is built on a buried river channel, which is filled with unconsolidated, water-saturated sediments, while the rest of the structure is underlain by more stable layers. A numerical simulation of the response of the structure beneath the stadium confirmed that resonance of earthquake waves within the buried river channel would greatly amplify the shaking of the ground above (Moczo *et al.*, 1995). In terms of seismic hazard estimates, this discovery highlights the importance of using architectural remnants of ancient buildings to inform scientists of possible hidden seismic hazards in the subsurface. Besides the damage pattern in the Colosseum, there is no visible evidence at ground level of this buried river channel; thus, the pattern of ancient earthquake damage alerted modern seismologists of a seismically vulnerable structure that also threatens more modern buildings in the same area.

Figure 11 An aerial view of the Colosseum in Rome.

4.20.4 Mutual Benefits of Archeoseismology

4.20.4.1 Benefits to Seismology

Seismology can gain a great deal by incorporating data from archeology (Nur, 2002). The most notable benefit is the ability to extend the observation window for seismic events to include prehistoric events. This can of course be done with paleoseismology as well, but the ongoing excavations of so many archeological sites around the world is a potential wellspring of seismological information that now largely goes untapped.

In addition to extending our observation window in time, archeology can also give us better spatial coverage, since many places for which we have few historical records have archeological sites that could yield seismic information. This is particularly true in the Americas, where our historical record is particularly poor prior to European contact.

Finally, as shown by the numerical studies of the Colosseum, archeology can fine-tune our seismic hazard maps, revealing seismically weak ground structure that may be invisible from the surface. Areas that suffered extreme damage in ancient times can indicate places where we should carry out subsurface seismic studies, to determine

whether resonance or other complicating effects make the seismic hazard higher than in the surrounding region.

4.20.4.2 Benefits to Archeology

The obvious benefit of archeoseismology to archeology is a clearer understanding both of what earthquake destruction looks like after the fact, and of where earthquake damage is most likely. There are certainly many instances where earthquake damage has been ascribed to human action, not because there is evidence of warfare or intentional destruction, but because that is the usual assumption. By being aware of earthquake hazards and being trained to recognize earthquake damage, archeologists can incorporate seismological causes into their hypotheses without the feeling that they are resorting to a *deus ex machina* in the absence of any other ideas.

Another potential benefit of archeoseismology stems from earthquakes' randomness and independence from human factors. Seismic waves travel much faster than human migrations or cultural diffusion. Thus, if the region affected by a particular ancient earthquake can be established with reasonable certainty, that earthquake could be used as a correlation layer. This would allow relative dating with unprecedented precision among all the sites in the earthquake-affected region, regardless of the dominant cultures of the sites, and regardless of whether the sites had any commerce with one another.

Finally, although ancient earthquakes were a tragedy to those they affected, they are a boon for archeologists. The excavation of tombs teaches us much about how ancient peoples died; the excavation of abandoned cities teaches us much about how they built; the excavation of ancient battlegrounds teaches us much about how they fought. But only sudden, unexpected catastrophes like earthquakes can preserve snapshots of how they actually lived their day-to-day lives.

References

Aharoni Y (1993) Megiddo. In: Stern E (ed.) *New Encyclopedia of Archaeological Excavations in the Holy Land*, Vol. 3, pp. 1003–1012. New York: Simon and Schuster.

Allen CR (1975) Geological criteria for evaluating seismicity. *Geological Society of America Bulletin* 86: 1041–1057.

Ambraseys NN (1970) Some characteristic features of the Anatolian fault zone. *Tectonophysics* 9: 143–165.

Amiran DHK, Arieh E, and Turcotte T (1994) Earthquakes in Israel and adjacent area: Macroseismic observations since 100 BCE. *Israel Exploration Journal* 44: 260–305.

Åström P and Demakopoulou K (1996) Signs of an earthquake at Midea? In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 37–40. Athens: The British School at Athens.

Bar Yosef O (1993) Ubeidiya. In: Stern E (ed.) *The New Encyclopedia of Archaeological Excavations in the Holy Land*, pp. 1487–1488. Carta, Jerusalem: The Israel Exploration Society.

British Historical Society of Lisbon (1990) Lisbon recalled: All Saints' Day, 1 November 1755. Terramoto de 1755. Testemunhos Britanicos, British account Lisboa, 670–672.

Browning I (1982) *Jerash and the Decapolis*. London: Chatto & Windus Ltd.

Catling HW (1981) Archaeology in Greece, 1980–81. *Archaeological Reports for 1980–81*, 3–48. London: The Society for the Promotion of Hellenic Studies.

Ellenblum R, Marco S, Agnon A, Rockwell T, and Boas A (1998) Crusader castle torn apart by earthquake at dawn, 20 May 1202. *Geology* 26(4): 303–306.

Guidoboni E, Comastri A, and Traina G (1994) *Catalogue of Ancient Earthquakes in the Mediterranean Area up to the 10th Century*. Phillips, B (trans.) Rome: Insituto Nazionale di Giofisica.

Kempinski A (1993) *Megiddo: A City-State and Royal Centre in North Israel*. Munich, Germany: Verlag C. H. Beck.

Ken-Tor R, Agnon A, Enzel Y, Stein M, Marco S, and Negendank JFW (2001) High-resolution geological record of historic earthquakes in the Dead Sea basin. *Journal of Geophysical Research* 106(B2): 2221–2234.

Kilian K (1996) Earthquakes and archaeological context at 13th century BC Tiryns. In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 63–68. Athens: The British School at Athens.

King GCP, Stein RS, and Lin J (1994) Static stress changes and the triggering of earthquakes. *Bulletin of Seismological Society of America* 84: 935–953.

Maroukian H, Gaki-Papanastassiou K, and Papanastassiou D (1996) Geomorphologic–seismotectonic observations in relation to the catastrophes at Mycenae. In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 189–194. Athens: The British School at Athens.

Moczo P, Rovelli A, Labák P, and Malagnini L (1995) Seismic response of the geologic structure underlying the Roman Colosseum and a 2-D resonance of a sediment valley. *Annali di Geofisica* 38(5–6): 939–956.

Mogi K (1968) Migration of seismic activity. *Bulletin of the Earthquake Research Institute (Tokyo)* 46: 53–74.

Mylonas G (1966) *Mycenae and the Mycenaean Age*. Princeton: Princeton University Press.

Mylonas-Shear I (1969) *Mycenaean Domestic Architecture*. Ph.D. Thesis. Ann Arbor, University of Michigan.

Mylonas-Shear I (1987) *The Panagia Houses at Mycenae*. Philadelphia: University Museum Press.

Nikonov A (1996) The Disappearance of the ancient towns of Disocuria and Sebastopolis in Colchis on the Black Sea: A problem in engineering geology and palaeoseismology. In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 195–204. Athens: The British School at Athens.

Nur A (2002) Earthquakes and archaeology. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake and Engineering Seismology*, 81A, pp. 765–774. London: Academic Press.

Nur A and Cline EH (2000) Poseidon's horses: Plate tectonics and earthquake storms in the Late Bronze Age Aegean and

Eastern Mediterranean. *Journal of Archaeological Science* 27(1): 43–63.
Nur A and Ron H (1997) Armageddon's earthquakes. *International Geology Review* 39: 532–541.
Papageorgiou S and Stiros SC (1996) The harbour of Aigeira (North Peloponnese, Greece): An uplifted ancient harbour. In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 211–214. Athens: The British School at Athens.
Papastamatiou D and Psycharis I (1996) Numerical simulation of the seismic response of megalithic monuments: Preliminary investigations related to the Apollo temple at Vassai. In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 225–236. Athens: The British School at Athens.
Pirazzoli PA (1996) Uplift of ancient Greek coastal sites: Study, methods and results. In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 237–244. Athens: The British School at Athens.
Pirazzoli PA, Ausseil-Badie J, Giresse P, Hadjidaki E, and Arnold M (1992) Historical environmental changes at Phalasarna harbor, West Crete. *Geoarchaeology* 7: 371–392.
Roth F (1988) Modeling of stress patterns along the western part of the North Anatolian fault zone. *Tectonophysics* 152: 215–226.
Sakellarakis Y and Sapouna-Sakellaraki E (1981) Drama of death in a Minoan temple. *National Geographic* 174: 205–222.
Sampson A (1996) Cases of earthquakes at Mycenaean and Pre-Mycenaean Thebes. In: Stiros S and Jones RE (eds.) *Archaeoseismology*, Fitch Laboratory Occasional Paper 7, pp. 113–118. Athens: The British School at Athens.
Solecki RS (1971) *Shanidar: The First Flower People*. New York: Knopf.
Soren D (1988a) *Kourion: The Search for a Lost Roman City*. New York: Anchor Books.
Soren D (1988b) The day the world ended at Kourion: Reconstructing an ancient earthquake. *National Geographic* 174(1): 30–53.
Spudich P, Hellweg M, and Lee WHK (1996) Directional topographic site response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake: Implications for mainshock motions. *Bulletin of Seismological Society of America* 86(1B): S193–S208.
Stein R, Barka A, and Dietrich JH (1997) Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysics Journal International* 128(3): 594–604.
Stein S and Okal EA (2005) Speed and size of the Sumatra earthquake. *Nature* 434: 581–582.
Stewart A (1993) A death at Dor. *Biblical Archaeology Review* 19: 30–36 84.
Stucky RA (1990) Schweizer Ausgrabungen in ez Zantur, Petra: Vorbericht der Kampagne 1988. *Annual of the Department of Antiquities of Jordan* XXXIV: 249–283.
Wood M (1996) *In Search of the Trojan War,* 2nd ed. Berkeley: University of California Press.
Yadin Y (1966) *Masada: Herod's Fortress and the Zealot's Last Stand*. New York: Random House, Inc.

4.21 Earthquake Hazard Mitigation: New Directions and Opportunities

R. M. Allen, University of California Berkeley, Berkeley, CA, USA

4.21.1 Introduction

Few natural events can have the catastrophic consequences of earthquakes, yet evidence abounds for repeating disasters in the same location. Archeological studies point to the recurring destruction of Troy, Jericho, and Megiddo in the Mediterranean and the Middle East, and, in the New World, debris from the 1906 San Francisco earthquake was found beneath the destruction caused by the 1989 Loma Prieta earthquake in San Francisco's Marina District.

Historical examples illustrate the sociopolitical impact of earthquakes. In 464 BCE, a powerful earthquake beneath the ancient Greek city Sparta led to the rebellion of Spartan slaves. According to Aristotle's Politics (1269a37-b5), these slaves were "like an enemy constantly sitting in wait for the disasters of the Spartans". The devastation that Sparta suffered from the earthquake offered them the perfect opportunity. The rebellion, which lasted for 10 years, limited Sparta's ability to check the growth of Athenian power in Greece and also led to the dissolution of the Spartan–Athenian alliance formed some 30 years earlier in the face of the Persian threat.

More recently, another natural disaster destroyed much of New Orleans and the Gulf coast of Louisiana. Few believed a natural hazard could be so devastating to a modern wealthy city, yet Hurricane Katrina flooded 80% of the city, much of which is below sea level, and destroyed over 300 000 housing units in August 2005. Despite a warning of the impending hurricane several days in advance, over 1800 people were killed. One year later, the population of the city is less than 50% of its previous level and it is clear that many will not return.

The challenge of natural hazard reduction generally, and earthquake hazard mitigation in particular, is the long return interval of these events. The infrequency of large seismic events provides only a limited data set for the study of earthquake impacts on modern cities, and the uncertainty as to when the next event will occur often places earthquake mitigation low on the priority list. The fields of seismology and earthquake engineering are also relatively juvenile, having only developed out of large destructive earthquakes at the end of the nineteenth and beginning of the twentieth centuries. Still, there has been considerable progress. Your chance of being killed in an earthquake is a factor of 3 less than it was in 1900.

Yet, earthquakes account for 60% of natural hazard fatalities today (Shedlock and Tanner, 1999). The number of people killed in earthquakes continues to rise in poorer nations, and the cost of earthquakes continues to rise for rich nations. The global population distribution is changing rapidly as underdeveloped nations continue to grow most rapidly in cities that are preferentially located in seismically hazardous regions. There has not yet been a large earthquake directly beneath one of these megacities, but when such an event occurs the number of fatalities could exceed 1 million (Bilham, 2004).

This chapter considers seismic hazard mitigation. First, we evaluate the hazard and risk around the globe to identify where mitigation is necessary. Next, we consider the topic of earthquake prediction which is often called upon by the public as the solution to earthquake hazard. Instead, effective earthquake mitigation strategies fall into two groups, long- and short-term. We address long-term methods first, focusing on the use of earthquake-resistant buildings. In the past, their development has been largely reactive and driven by observed failures in most recent earthquakes. As testing of building performance in future earthquakes has become more viable, there is a potential for more rapid improvements to structural design. At the same time, however, the challenges of implementation will persist, leading to a widening implementation gap between the rich and poor nations. Short-term mitigation is the topic of the final section. Over recent years, modern seismic networks have facilitated the development of rapid earthquake information systems capable of providing hazard information in the minutes after an earthquake. These systems are now beginning to provide the same information in the seconds to tens of seconds prior to ground shaking. We consider possible future applications around the world.

4.21.2 Recognizing and Quantifying the Problem

4.21.2.1 Forecasting Earthquakes at Different Spatial and Temporal Scales

The first step in seismic hazard mitigation is identification and quantification of where the hazard exists. Today, plate tectonics provides the theoretical framework for identifying and characterizing seismic source regions: where earthquakes have occurred in the past, earthquakes will occur in the future. But before the development of plate tectonic theory in the late 1960s, the same concept was in use to forecast future earthquakes. In a letter to the *Salt Lake City Tribune* in 1883, G. K. Gilbert reported the findings of his field work along the Wasatch Front. He noted that the fault scarps were continuous along the base of the Wasatch with the exception of the segment adjacent to Salt Lake City where a scarp was missing. He concluded that there had been no recent earthquake on the section adjacent to the city, and this section was therefore closer to failure. In his study of deformation associated with the 1906 San Francisco earthquake, H. F. Reid built on Gilbert's model to develop elastic rebound theory which remains the

basis of our understanding of the earthquake cycle today (Reid, 1910). In the elastic rebound model, the relative motion between two adjacent tectonic plates is accommodated by elastic deformation in a wide swath across the plate boundary. Once the stress on the plate boundary fault exceeds the strength of the fault, rupture occurs and the accumulated deformation across the plate boundary collapses onto the fault plane.

This cyclicity to earthquake rupture is the basis of the seismic gap method of earthquake forecasting. If a fault segment fails in a quasi-periodic series of characteristic earthquakes, then the recurrence interval between events can be estimated either from the dates of past earthquakes or calculated by taking the characteristic slip during an earthquake and dividing by the long-term slip rate of the fault. Reported successes of seismic-gap theory include the deadly 1923 Kanto earthquake and the great Nankaido earthquakes of 1944 and 1946 (Aki, 1980; Nishenko, 1989). In 1965, Fedotov published a map showing where large-magnitude earthquakes should be expected, and his predictions were promptly satisfied by the 1968 Tokachi-Oki, 1969 Kuriles, and 1971 central Kamchatka earthquakes (Fedotov, 1965; Mogi, 1985). In the 1970s, the approach was applied around the globe. The estimates of relative plate motions provided by the new plate tectonic theory could be translated into slip rates across major faults. Once combined with data on the recent occurrence of large earthquakes, maps were generated identifying plate boundary segments with high, medium, and low seismic potential (Kelleher *et al.*, 1973; McCann *et al.*, 1979).

However, the utility of the seismic gap method for earthquake forecasting remains a topic of debate today (e.g., Nishenko, 1989; Kagan and Jackson, 1991; Nishenko, 1991; Jackson and Kagan, 1993; Nishenko and Sykes, 1993; Kagan and Jackson, 1995). Challenges to its practical application include the incomplete historic record of earthquakes making it difficult to estimate recurrence intervals, and difficulty in identifying the characteristic earthquake for a given fault segment. Earthquakes are also observed to cluster in space and time. Mogi (1985) proposed that plate boundary segments go through alternating periods of high and low activity, and the earthquake catalog suggests alternating periods of subduction versus strike-slip earthquake activity (Romanowicz, 1993). Laboratory experiments of stick-slip behavior show that rupture occurs at irregular intervals with variable stress drops. This implies that the state of stress before and/or after each earthquake is also variable. In Reid's original development of elastic rebound theory, he forecast that the next earthquake should be expected when "the surface has been strained through an angle of 1/2000" (Reid, 1910). However, he also points out that this assumes a complete stress drop, that is, release of all accumulated strain, by the 1906 earthquake.

The Parkfield prediction experiment is one of the more famous applications of seismic gap theory. Three M 6 earthquakes located close to Parkfield in central California were instrumentally recorded in 1922, 1934, and 1966. Other data suggest an additional three events in 1857, 1881, and 1901 with a similar size and location. The similar recurrence interval of 22 years for the six events, the similar waveforms for the 1922, 1934, and 1966 events, and similar foreshock patterns prior to 1934 and 1966 make this one of the strongest cases for a characteristic earthquake (Bakun and McEvilly, 1979). Based on this evidence, Bakun and Lindh (1985) predicted that the next earthquake was due in 1988 with a 95% confidence that it would occur before 1993. An M 6.0 earthquake did occur on the Parkfield segment of the San Andreas, but not until September 28, 2004. While it had the same magnitude as previous events, the characteristics of its rupture were different (e.g., Langbein *et al.*, 2005).

These examples show that while the concept of recurring seismicity is useful for forecasting future seismic hazard, the application of a recurrence interval to predict the timing of the next earthquake is fraught with uncertainties. When viewed as a stationary series, past earthquake history can be used to forecast the probability of an earthquake over long time periods (hundreds of years), and this forms the basis of the probabilistic seismic hazard analysis discussed in the next section. However, as the spatial and temporal scales for the forecast become smaller, the uncertainties in those forecasts become greater. The challenge is to provide forecasts that are considered relevant by society, a society which at best plans for time periods of years to decades.

4.21.2.2 Global Seismic Hazard

The United Nations designated the 1990s the International Decade of Natural Disaster Reduction. The Global Seismic Hazard Assessment Program (GSHAP) was part of this effort and had the goal of improving global standards in seismic hazard assessment (Giardini, 1999). From 1992 to 1998, an

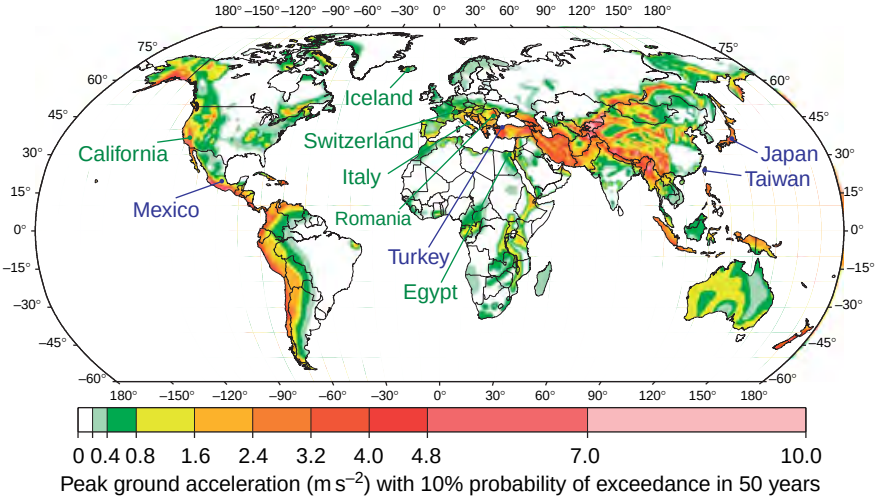


Figure 1 The global seismic hazard map developed by the Global Seismic Hazard Assessment Program (Giardini, 1999). The map depicts PGA with a 10% probability of exceedance in 50 years, corresponding to a return interval of 475 years. The cooler colors represent lower hazard while the warmer colors are high hazard: white and green correspond to low hazard (0–0.08 m s^{-2}); yellow and light orange correspond to moderate hazard (0.08–0.24 m s^{-2}); darker orange corresponds to high hazard (0.24–0.40 m s^{-2}); and red and pink correspond to very high hazard (> 0.40 m s^{-2}).

international collaboration of scientists conducted coordinated probabilistic seismic hazard analyses on a regional basis and combined them into the uniform global seismic hazard map shown in **Figure 1** (Giardini *et al.*, 1999). The maps present the levels of peak ground acceleration (PGA) with a 10% probability of exceedance (90% probability of nonexceedance) within 50 years, corresponding to a return period of 475 years. For more information on GSHAP, visit http://www.seismo.ethz.ch/GSHAP/.

Probabilistic seismic hazard analysis (PSHA) was first introduced by Cornell (1968). PSHA provides the relationship between some ground motion parameter, such as PGA, and its average return interval. There are three elements to the methodology. First, the seismic sources in a region must be characterized. It is necessary to determine where earthquakes occur, how often they occur, and how large they can be. Seismicity catalogs, both instrumental and preinstrumental, form the basis of this assessment. But these catalogs are inevitably incomplete with respect to geologic timescales. Additional geodetic and geologic data are therefore included when available. Second, the expected distribution of ground shaking for all possible earthquakes is estimated. This is usually achieved using attenuation relations which describe the level of ground shaking as a function of magnitude, distance, fault type, and local site conditions. The attenuation relations are determined by regression of peak ground shaking observations for past earthquakes in the region. The quality of the attenuation relations is therefore data-limited, as we do not have observations of all possible earthquakes, particularly the larger infrequent events. For this reason, theoretical modeling of waveform propagation is now being used to improve our understanding of likely ground motions for the largest earthquakes. Finally, the probability of ground shaking at various levels is calculated by determining the annual frequency of exceedance.

To illustrate PSHA, consider the historic parameter method (Veneziano *et al.*, 1984; McGuire, 1993). A uniform earthquake catalog is developed for the region, and attenuation functions are identified. The expected ground motion for each earthquake is then determined at every site across the region. Return periods for exceedance of ground shaking at various levels can then be tabulated and plotted to generate a hazard curve. The curve provides ground shaking level versus recurrence interval, or, equivalently, probability of exceedance within some time window. The choice of ground shaking parameter varies. PGA is a short period ground motion parameter that is proportional to force and is the most commonly mapped as the seismic provisions of current building codes specify the horizontal force a building should withstand during an earthquake. It is also the most appropriate measure for the most common building type, one- and two-story buildings, as they have short natural

periods of typically 0.1–0.2 s. Other parameters that are used include peak ground velocity (PGV), which is more sensitive to longer periods and therefore appropriate for taller buildings (the natural period of buildings is typically 0.1 s per floor), and spectral response ordinates at various periods (0.3 s, 0.5 s, 1.0 s, 2.0 s, etc.), which are also related to the lateral forces that damage taller, longer period, buildings.

The GSHAP applied PSHA around the globe. While every effort was made to apply a uniform analysis, the differences in available data set inevitably result in some differences in the analyses for different regions (see Grunthal *et al.*, 1999; McCue, 1999; Shedlock and Tanner, 1999; Zhang *et al.*, 1999). Hazard curves were generated for all locations, and **Figure 1** shows the PGA with a 10% probability of exceedance within 50 years. The greatest hazard is adjacent to the major transform and subduction plate boundaries: around the Pacific rim, and through the broad east–west belt running from the Italian Alps, through Turkey, the Zagros Mountains of Iran, the Hindu Kush and Tian Shan, and then broadening to a wider belt including the region from the Himalaya to Siberia. High seismic hazard also wraps around the coastlines of the northeast Indian Ocean, where the 2004 Sumatra–Andaman earthquake and tsunami was responsible for an estimated quarter of a million deaths. The largest recorded earthquakes are all subduction zone events; the largest three events during the last century were the 1960 Chile (M_w 9.5), 1964 Alaska (M_w 9.2), and 2004 Sumatra–Andaman (M_w 9.1) earthquakes. But the seismic hazard associated with major transform boundaries is just as large despite typically generating smaller earthquakes. This is due to the greater depth of large subduction zone earthquakes (tens of kilometers) and distance offshore, allowing attenuation of the seismic waves before they reach the land surface. By comparison, strike-slip faults rupture the shallow continental crust such as along the San Andreas Fault and the North Anatolian Fault.

4.21.2.3 Changing Seismic Risk

4.21.2.3.1 Earthquake fatalities since 1900

The new millennium has not started well in terms of earthquake impacts on society. As of October 2006, the twenty-first century has seen almost 400 000 deaths associated with earthquakes. This represents more than 20% of the estimated 1.8 million deaths during the entire twentieth century. There is no evidence of any increase in seismic hazard; the number of earthquakes is not increasing. But, is there an increase in seismic risk?

Seismic hazard analysis provides information about the likelihood of earthquakes and associated ground shaking (**Figure 1**). But the hazard is distinct from the 'seismic risk', which represents the anticipated losses in a region either for a given scenario earthquake or for all anticipated earthquakes. Determination of the risk involves a convolution of the seismic hazard with population density and the properties of the built environment, including the number of buildings and the type of construction. Fragility curves are used to describe the likely damage to a building, or construction type, given different levels of ground shaking. Frequent, large earthquakes in remote areas represent high seismic hazard but low seismic risk, while moderate earthquakes directly beneath a large urban center can represent low hazard but high risk.

Figure 2 shows the cumulative number of earthquake deaths since 1900. The data come from the Significant Earthquake Database (Dunbar *et al.*, 2006), edited to remove multiple entries and updated for the present paper. Statistical analysis of such data is notoriously difficult as it is dominated by infrequent high-fatality events. While there were 138 earthquakes with more than 1000 fatalities since 1900, the 10 events with the largest fatality rate caused over 60% of the deaths. The data show two trends, pre- and post- 1940 (**Figure 2**). Prior to 1940, fatalities occur at a rate of ~25 000 per year; after 1940, the character changes and is dominated by two large

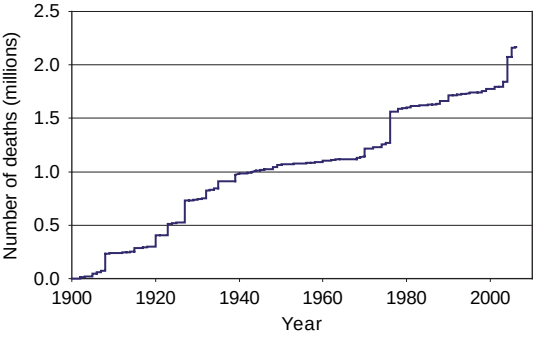


Figure 2 Cumulative number of earthquake fatalities since 1900. Note the change in character pre- and post-1940. The annual rates are ~25 000 per year pre-1940 and ~19 000 per year post-1940. Post-1940 fatalities are dominated by the Tangshan (1976) and Sumatra (2004) events with lower rates (~8000 and ~9000 per year) in between.

fatality events (**Figures 2** and **3(a)**) and lower fatality rates of ~8000 per year from 1940 to 1976, and ~9000 per year from 1976 to 2004. The 1976 Tangshan earthquake is the most recent in a series of earthquakes in China with very large numbers of fatalities. The 1920 Gansu and 1927 Tsinghai earthquakes both killed an estimated 200 000; another earthquake in Gansu Province killed 70 000 in 1932; and, finally, the Tangshan earthquake had an official death toll of 242 000 (as in **Figure 2**) but unofficial estimates as high as 655 000. The second major event in the post-1940 time series is the Sumatra earthquake and tsunami of 26 December 2004. The United Nations estimates 187 000 confirmed dead and an additional 43 000 missing. Most of these fatalities occurred in the Aceh Province of Indonesia, at the northern end of the island of Sumatra, and along the Nicobar and Andaman Islands extending to the north along the subduction zone. Sri Lanka to the west and Thailand to the east were also heavily affected. The fatalities from this event are therefore more distributed than the other major events since 1900 on account of the broader reach of tsunami hazard.

The total fatality rate from 1940 to 2006 is ~19 000 per year, lower than the ~25 000 per year rate from 1900 to 1940. However, it would be a mistake to conclude that the earthquake-related fatality rate is declining as the post-1940 rate is dominated by just two events. In fact, the last five centuries of earthquake fatalities show an increasing rate. Using a best-fit power law and data from the last five centuries, Bilham (2004) estimates that the annual rate of earthquake fatalities continues to increase. While the number of deaths is increasing,

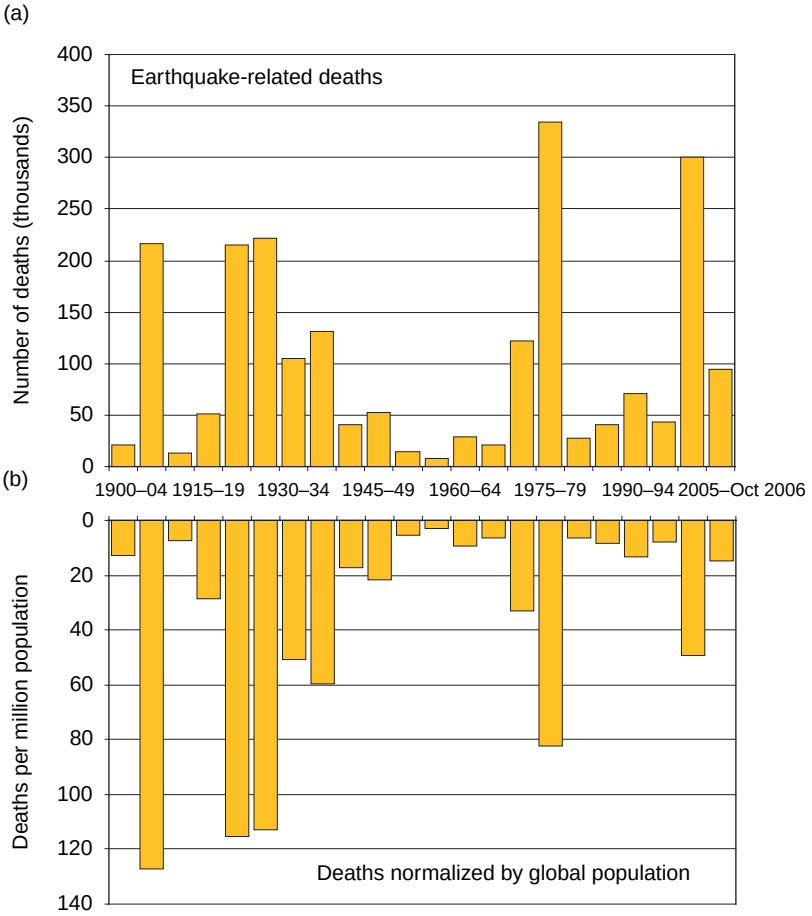


Figure 3 Earthquake-related deaths since 1900 in 5 year bins. (a) Total number of deaths. (b) Deaths per million global population. While the intervals including the 1976 Tangshan and 2004 Sumatra events have the highest number of fatalities, once normalized by global population it is the first part of the twentieth century which has the highest rates with over 100 per million in three intervals.

it is not increasing as quickly as global population. Normalizing Bilham's best-fit annual fatality rate by global population, an individual's risk of dying in an earthquake has reduced by a factor of 2 since 1950 and a factor of 3 since 1900. This can also be seen when considering the fatalities during 5 year intervals as shown in **Figure 3**. The largest number of fatalities in these 5 year intervals was due to the 1976 Tangshan and 2004 Sumatra events. During these intervals, there were over 300 000 deaths, but once normalized by the global population the highest fatality rates were during the first part of the twentieth century, when there were more than 100 deaths per million population during three 5 year intervals (**Figure 3(b)**).

So, are the advances in earthquake science and engineering paying off? Are we living in a more earthquake-resilient world? This conclusion would be premature for several reasons. First, the fatality rate is dominated by large impact events, and a few such events in the coming decades would reverse this trend. Second, the application of earthquake mitigation strategies is highly uneven around the globe, resulting in very different trends in regional earthquake fatality rates. Third, the distribution of global population is changing rapidly, on shorter timescales than the earthquake cycle. While the more-developed nations show zero growth, rapid growth continues in the less-developed nations, particularly in the cities. Finally, it would be irresponsible to declare success in global earthquake mitigation when the annual number of fatalities continues to increase.

4.21.2.3.2 Concentrations of risk

The high fatality rate earthquakes recur in a relatively small number of countries. Since 1900, the 12 earthquakes causing more than 50 000 fatalities have occurred in China, Pakistan, Iran, Indonesia, Japan, Italy, and Peru. Almost half of all earthquakes causing more than 1000 deaths have occurred in these seven countries. But the application of earthquake mitigation strategies is variable. In Japan, which has seen over 100 000 fatalities in the last century, most from the 1923 Tokyo earthquake and fire, stringent building codes are enforced, regular earthquake evacuation drills are carried out, and, most recently, an earthquake early-warning system was implemented. While in Iran, which experienced ~190 000 fatalities since 1900, the number of earthquake fatalities has tracked the population growth – one in 30 000 Iranians die in earthquakes – and the existence of earthquake building codes has had little or no effect (Berberian, 1990; Bilham, 2004).

The introduction of the medicinal control of contagious diseases at the beginning of the twentieth century finally allowed rapid growth of urban centers. Since 1950, 60% of global population growth has occurred in urban centers, almost 50% in the lesser-developed nations (United Nations, 2004). Today, the global rural population is almost flat and the number of urban dwellers will exceed rural dwellers in 2007 for the first time. This is causing a rapid redistribution of the global population. Most of the population growth is now occurring in the less-developed nations. Within each nation, the population is migrating to the urban centers, particularly in the less-developed nations. In a series of papers, Bilham (1988, 1996, 1998, 2004) has pointed to this trend and cautioned that much higher numbers of fatalities from single events might be expected when an earthquake strikes beneath one of the growing number of large urban agglomerations.

This migration of population to the cities results in concentrations of risk. As the number of cities grows, the likelihood that an earthquake will strike a city also grows. In addition to this trend, the global distribution of the world's largest urban centers is changing. The largest cities today are in locations with a greater seismic risk than the largest cities in 1950. **Figure 4** shows the seismic hazard for the world's 30 largest urban centers in 1950, 1975, 2000, and 2015. It shows that while only 10 were in regions of moderate to high hazard in 1950, this number had increased to 16 by 2000, and the trend is projected to continue. Most of the change occurred by adding new cities to the top 30 in regions of high hazard, while cities with a low hazard dropped off the list; the number of moderate hazard cities remains fairly constant. The geographic distribution of the 30 largest urban centers is shown in **Figure 5**. The reason for the changing hazard is clear. While the growth of cities in northern Europe and the northeastern United States has been relatively slow, rapid growth of cities in western South America and across Asia has propelled these cities with higher seismic hazard into the top 30 list.

It is tempting to associate the changing trend of global earthquake fatalities (**Figure 2**) with the growth of cities. Pre-1940 earthquake fatalities are more constant, while most fatalities post 1940 occurred in two events. One of the two events, 1976 Tangshan, was beneath a large city, but the fatalities in the 2004 Sumatra event were more distributed due

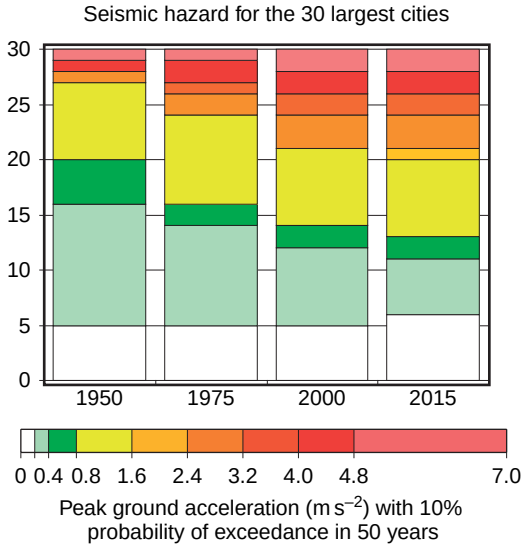


Figure 4 Seismic hazard for the 30 largest cities in 1950, 1975, 2000, and 2015 (projected). City population data from the United Nations. The seismic hazard at each city location is provided by the GSHAP map and represented as PGA with a 10% probability of exceedance in 50 years. The chart shows that cities in seismically safe regions are removed from the top 30 list as cities in hazardous regions grow more rapidly.

to the tsunami. The shortness of the time history makes it impossible to be certain of the cause, and there has not yet been a large earthquake beneath a megacity.

4.21.2.4 Local Hazard and Risk: The San Francisco Bay Area

4.21.2.4.1 The San Francisco Bay Area

All seismic hazard mitigation occurs on a local scale. For this reason, it is useful to consider a case example such as the San Francisco Bay Area (SFBA). The SFBA sits within the Pacific–North America plate boundary, which takes the form of multiple fault strands through the region (**Figure 6**). The interseismic displacement between the Pacific Plate and the western edge of the Central Valley is 38 mm yr^{-1}, representing approximately 80% of the motion between the Pacific and North American Plates (d'Alessio *et al.*, 2005). This narrow strip of land that forms the Coast Ranges of California is only ~100 km wide but has a population approaching 7 million concentrated around the bay. The SFBA has

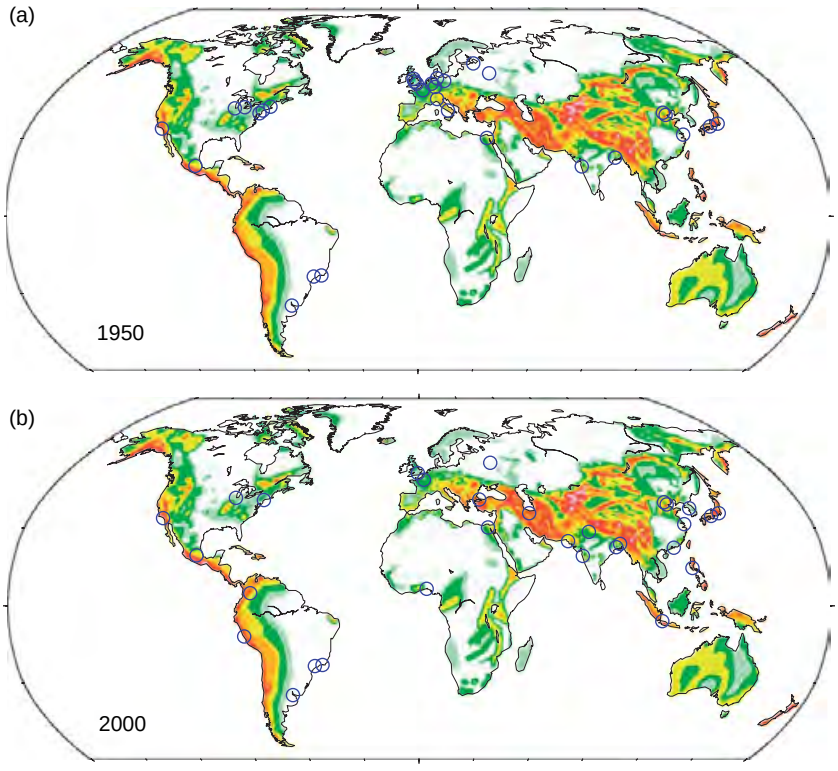


Figure 5 The locations of the 30 largest cities in (a) 1950 and (b) 2000 (blue circles) superimposed on the GSHAP hazard map. The increased seismic hazard for the largest cities is due to relatively slow growth of cities in the eastern US and northwest Europe while cities across Asia grow more rapidly.

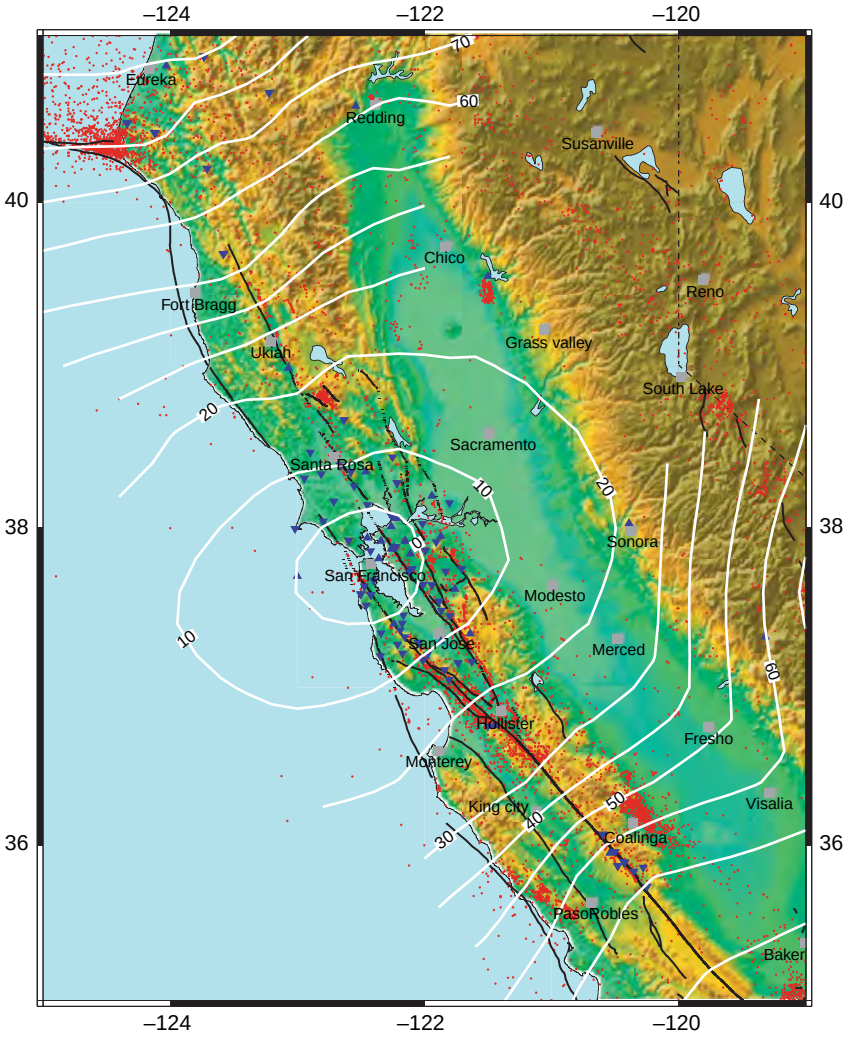


Figure 6 Map of northern California showing topography (color pallet), faults (black lines), earthquakes (red dots), seismic stations available for early warning (blue triangles and diamonds) and the warning time San Francisco could expect for earthquakes at all locations across the region (white contours, time in seconds). The warning time is estimated using ElarmS and the current seismic network and telemetry. There would be greatest warning for earthquake furthest away from the city. The existing seismic stations shown are operated by UC Berkeley and the US Geological Survey.

the highest density of active faults and the highest seismic moment rate per square kilometer of any urban area in the United States (WG02, 2003).

The historic earthquake record is short, believed to be complete for $M \geq 5.5$ since 1850, at which time the population exploded after gold was found in the Sierra foothills (Bakun, 1999). Some information is available back to 1776 when the first Spanish mission, Mission Delores, was founded. The record contains six $M \geq 6.5$ earthquakes in the SFBA in 1836, 1838, 1865, 1868, 1906, and 1989, four in the 70 years prior to 1906 and only one in the 100 years since. This change in the seismic energy release rate is believed to be due to the 'stress shadow' resulting from the 1906 earthquake (Harris and Simpson, 1998). The 1906 event ruptured the northernmost 450 km of the San Andreas Fault from San Juan Bautista to Cape Mendocino extending through the SFBA and destroying much of San Francisco and Santa Rosa to the north. As most faults in the SFBA share a subparallel, strike-slip geometry to the San Andreas Fault, they were relaxed by the 1906 rupture.

Mapping active faults in California is the responsibility of the California Geological Survey (CGS). Under the 1972 Alquist-Priolo Earthquake Fault Zoning Act, all faults that have ruptured within the last 11 000 years are considered active, and building

close to these known faults is tightly regulated to ensure that buildings are at least 50 feet from the fault trace. The CGS is now also in the process of mapping other seismic hazards including liquefaction and landslide hazards during earthquakes.

The Southern California Earthquake Center (SCEC) is a collaboration of earthquake scientists working with the goal of understanding the earthquake process and mitigating the associated hazards. While SCEC is focused on the earthquake problem in southern California, the methodologies developed by SCEC scientists to quantify earthquake probabilities and the shaking hazards associated with them are applicable everywhere, including in our chosen region of focus, the SFBA.

4.21.2.4.2 Earthquake probabilities

To evaluate the probability of future earthquakes and ground shaking in the region, the US Geological Survey established a Working Group on Earthquake Probabilities. In several incarnations starting in 1988, the group has collected data and applied the most up-to-date methodologies available to estimate long-term earthquake probabilities drawing on input from a broad cross section of the Earth science community. The most recent study (hereafter WG02) was completed in 2002 (WG02, 2003). In it, the probabilities of one or more earthquakes in the SFBA, on one of the seven identified fault systems, between 2002 and 2032 were estimated. The likely intensities of ground shaking were also combined to produce a probabilistic seismic hazard map for the region, similar to the GSHAP map discussed above. The WG02 results are shown in **Figure 7**.

The earthquake model used to estimate these probabilities has three elements. The first is a time-independent forecast of the average magnitudes and rates of occurrence of earthquakes on the major identified fault segments. It is derived from the past earthquake catalog. The second element includes four time-dependent models of the earthquake process to include the effects of the earthquake cycle and interactions between the fault systems. The concept of the earthquake cycle holds that after a major earthquake and associated aftershocks, another major rupture is not possible until the elastic strain has reaccumulated (Reid's elastic rebound theory). As time goes by, the probability of an earthquake therefore increases. A major earthquake also reduces the stress on any adjacent faults with a similar orientation, generating a stress shadow. This has been observed both in numerical models and in the reduced seismicity on faults adjacent to the San Andreas after the 1906 rupture. In the SFBA, both the 1906 event and the more recent 1989 earthquake cast stress shadows. The third element of the earthquake model characterizes the rates of background seismicity, that is, earthquakes that do not occur on the seven major fault systems. The 1989 Loma Prieta event was one such earthquake. These various earthquake models provide different estimates of earthquake probabilities. WG02 uses expert opinion to determine the relative weight for each probability estimate derived from each model.

Figure 7 shows the WG02 results. It is estimated that there is a 62% probability of one or more $M \geq 6.7$ earthquakes in the SFBA from 2002 to 2032. As shown in **Figure 7(a)**, the probability of one or more $M \geq 6.7$ events is greatest on the Hayward–Rodgers Creek and San Andreas Faults, which have probabilities of 27% and 21%, respectively. The estimated uncertainties in these numbers are substantial. For the SFBA the 95% confidence bounds are 37% to 87%. For the Hayward–Rodgers Creek and San Andreas Faults, the bounds are 10–58% and 2–45%, respectively. A critical source of this uncertainty is the extent to which the SFBA has emerged from the stress shadow of the 1906 earthquake. Simple elastic interaction models suggest that the region should have emerged from the stress shadow, while the low seismicity rates for the last century would suggest that the SFBA remains within the shadow. Rheological models of the crust and uppermost mantle, and perhaps the 1989 Loma Prieta earthquake, suggest that the region may just be emerging now. If the region is emerging, then it can expect an increase in the number of major events over the next few decades.

4.21.2.4.3 Future losses

Just as the world has not experienced a major earthquake beneath a megacity, the US has not experienced a major earthquake directly beneath one of its cities. The two most damaging earthquakes were the 1989 Loma Prieta earthquake (which was beneath the rugged mountains 100 km south of San Francisco and Oakland) and the 1994 Northridge earthquake (which, although centered beneath the populated San Fernando Valley, caused strongest ground shaking in the sparsely populated Santa Suzanna Mountains to the north). Each event caused ~60 deaths and the estimated damages were $10 and

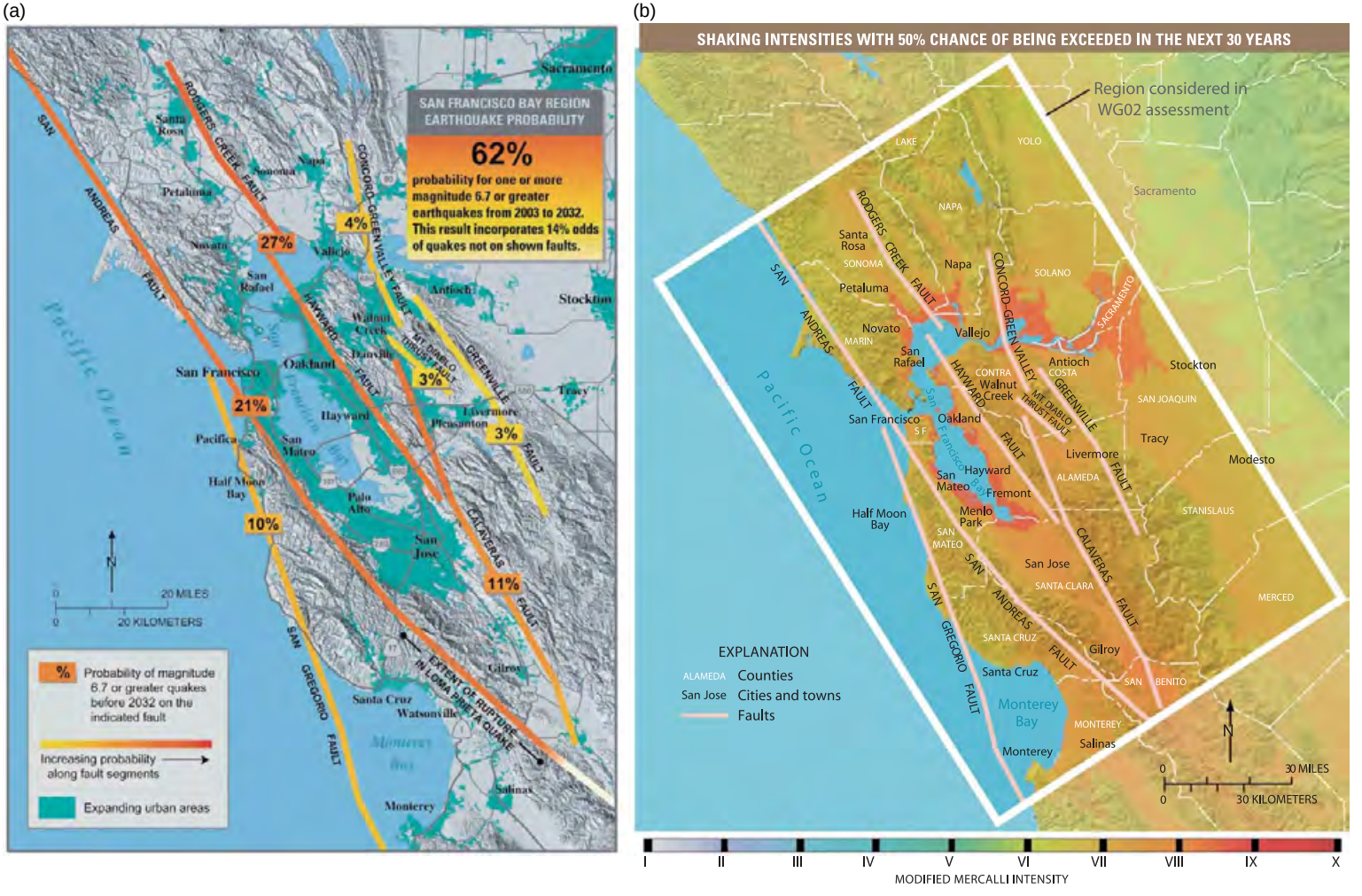


Figure 7 (a) Map of the San Francisco Bay Area (SFBA) showing the urban areas and the probabilities of $M \geq 6.7$ earthquakes by 2032. The probability of such an event in the SFBA is 62%, the probabilities of an $M \geq 6.7$ earthquake on each fault are also indicated. (b) Shaking intensities with a 50% probability of exceedance by 2032. The soft sediments and landfill around the bay and delta are where the shaking hazard is the greatest. Both figures are taken from USGS Fact Sheet 039-03 (2003).

$46 billion for Loma Prieta and Northridge, respectively (in 2000-dollars). While the impacts were significant, the events were relatively moderate in damage.

The ground shaking estimates, such as those that are part of WG02, provide the basis for loss estimation. Loss estimation methodologies use data on the locations and types of buildings, ground shaking maps for scenario or past earthquakes, and fragility curves relating the extent of damage to the ground shakng for each building type, to estimate the total damage from the event. The worst-case scenario considered for northern California is a repeat of the 1906 earthquake. The losses have been estimated at $170–225 billion for all related losses including secondary fires and toxic releases (RMS, 1995), a factor of 2 greater than the $90–120 billion loss estimate for property alone (Kircher *et al.*, 2006). It is estimated that the number of deaths could range from 800 to 3400 depending on the time of day, and 160 000–250 000 households will be displaced (Kircher *et al.*, 2006). An earthquake rupturing the length of the Hayward–Rodgers Creek Fault is estimated to cause $40 billion in damage to buildings alone (Rowshandel, 2006).

These estimates of seismic hazard and risk provide a quantitative basis for earthquake hazard mitigation in the region. The choice of a relatively short, 30-year time window by WG02 has the advantage that it is a similar timescale to that of property ownership. But, as pointed out above and by WG02, the reliability of PSHA analysis decreases as the temporal and spatial scales decrease. Our observations of large ($M > 6.5$) earthquakes in California are limited. Many of the recent damaging eartthquakes occurred on faults that had not been recognized, including the two most damaging earthquakes, the 1989 Loma Prieta and 1994 Northridge events. While 'background seismicity' is included in the seismic hazard estimates, these events are a reminder of the limitations to our current understanding of the earthquake hazard. These hazard and risk estimates are therefore most appropriately used to motivate broad efforts to mitigate seismic hazard across the entire SFBA rather than efforts along a specific fault segment. The limitations in our observational data set also caution against becoming too 'tuned' in mitigation strategy. The use of multiple mitigation strategies will prevent over-reliance on a single, and possibly limited, model of future earthquake effects.

4.21.3 The 'Holy Grail' of Seismology: Earthquake Prediction

"When is the big one?" is the first question asked by every member of the public or press when they visit the Berkeley Seismological Laboratory. Answering this question, predicting an earthquake, is often referred to as the Holy Grail of seismology. In this context, a prediction means anticipating the time, place, and magnitude of a large earthquake within a narrow window and with a high enough probability that preparations for its effects can be undertaken (Allen, 1976). For the general public, answering this question is the primary responsibility of the seismological community.

The public considers earthquake prediction important because it would allow evacuation of cities and prevention of injury and loss of life in damaged and collapsed buildings. However, the seismology and engineering communities have already developed a strategy to prevent building collapse by identifying the likely levels of ground shaking and designing earthquake-resistant buildings that are unlikely to collapse. Once building codes for earthquake-resistant buildings are fully implemented, earthquake prediction would not be as important. But even before full implementation of building codes, earthquake prediction would only be partially successful as it would be capable of mitigating immediate and not long-term impacts of earthquakes. A prediction would allow for evacuations, but the ensuing earthquake would leave the urban area uninhabitable and only a fraction of the prior occupants would likely return.

It is possible to make high-probability short-term predictions for hurricanes as was done in the case of Hurricane Katrina in August 2005. Still, an estimated 1800 people were killed when New Orleans and other areas of Louisiana and Mississippi were inundated by flood waters. In New Orleans, 80% of the city was flooded, destroying much of the housing and infrastructure, and it is not yet clear what proportion will be replaced. One year later, the population of New Orleans was less than half its pre-Katrina level and roughly equivalent to what it was in 1880. If the built environment was designed to withstand a hurricane of Katrina's strength, these lives would not have been lost and New Orleans would still be thriving.

For the scientific community, earthquake prediction has a much broader meaning, encompassing

the physics of the earthquake process at all timescales. The long-term probabilistic forecasts described in the previous section are predictions, but they have low probabilities of occurrence over large time windows. There is currently no approach that has consistently predicted large-magnitude earthquakes and most seismologists do not expect such short-term predictions in the foreseeable future. While many advances have been made in understanding crustal deformation, stress accumulation, rupture dynamics, friction and constitutive relations, fault interactions, and linear dynamics, a lack of understanding of the underlying physics and difficulty in making detailed field observations mapping the spatial and temporal variations in structure, strain, and fault properties makes accurate short-term predictions difficult.

In addition to these observational constraints, earthquakes are part of a complex process in which distinct structures such as faults interact with the diffuse heterogeneity of the Earth's crust and mantle at all scales. Even simple mechanical models of the earthquake process show chaotic behavior (Burridge and Knopoff, 1967; Otsuka, 1972; Turcotte, 1992), suggesting it will be difficult to predict earthquakes in a deterministic way. Instead, it may only be possible to make predictions in a statistical sense with considerable uncertainty (Turcotte, 1992). Kanamori (2003) details the important sources of uncertainty: (1) the stress accumulation due to relatively constant plate motion can be modified locally by proximal earthquakes; (2) the strength of the seismogenic zone may change with time, say due to the migration of fluids; (3) predicting the size of an earthquake may be difficult depending on whether a small earthquake triggers a large one; (4) external forces may trigger events as observed in geothermal areas after large earthquakes.

Despite these challenges, the search for the silver bullet – an earthquake precursor – continues. As pointed out by Kanamori (2003), there are two types of precursors. For the purpose of short-term earthquake prediction, identification of a single precursor before *all* large magnitude events is desirable. To date, no such precursor has been identified as far as we know. However, unusual precursory signals have been observed before one, or perhaps a few earthquakes. These precursors may be observed before future earthquakes and are therefore worthy of research effort. The list of observed precursors includes increased seismicity and strain, changes in seismic velocities, electrical resistivity and potential, radio frequency emission, and changes in ground water levels and chemistry (see Rikitake, 1986).

The one successful prediction of a major earthquake was prior to the 1975 M_S 7.3 Haicheng (China) event. More than 1 million people lived near the epicenter, and a recent evaluation of declassified documents concludes that an evacuation ordered by a local county government saved thousands of lives (Wang *et al.*, 2006). There were two official middle-term predictions (1–2 years). On the day of the earthquake, various actions taken by provincial scientists and government officials constituted an imminent prediction, although there was no official short-term (a few months) prediction. A foreshock sequence consisting of several hundred events triggered the imminent prediction; other precursors including geodetic deformation, changes in groundwater level, chemistry, and color, and peculiar animal behavior are also reported to have played a role (Wang *et al.*, 2006). What is not known is how many false predictions were made prior to the evacuation, nor is it known how many earthquake evacuation orders have been made across China. The initial euphoria over the successful evacuation was soon dampened by the Tangshan earthquake the following year for which there was no prediction.

Extensive literature exists detailing the specifics of the various proposed earthquake prediction methodologies and other reported cases of earthquake prediction (Rikitake, 1976; Vogel, 1979; Wyss, 1979; Isikara and Vogel, 1982; Rikitake, 1982; Unesco, 1984; Mogi, 1985; Rikitake, 1986; Gupta and Patwardham, 1988; Olson *et al.*, 1989; Wyss, 1991; Lomnitz, 1994; Gokhberg *et al.*, 1995; Sobolev, 1995; Geller, 1996; Knopoff, 1996; Geller, 1997; Geller *et al.*, 1997; Sykes *et al.*, 1999; Rikitake and Hamada, 2001; Kanamori, 2003; Ikeya, 2004). Expert panels are used in many countries to evaluate earthquake predictions and provide advice to governments and the public. In the US, the National Earthquake Prediction Evaluation Council (NEPEC) provides advice to the director of the US Geological Survey, and the California Earthquake Prediction Council (CEPEC) advises the Governor. No short-term earthquake predictions have been made by these councils to date.

4.21.4 Long-Term Mitigation: Earthquake-Resistant Buildings

The implementation of building codes mandating the use of earthquake-resistant buildings has been highly

successful in mitigating the impact of earthquakes in some regions. The number of fatalities has been reduced, and the majority of direct economic losses in recent US earthquakes (e.g., 1989 Loma Prieta, 1994 Northridge, and 2001 Nisqually) were from damage to buildings and lifelines constructed before 1976 when the Uniform Building Code was updated following the 1971 San Fernando earthquake (National Research Council, 2003). In the past, the improvement of building design was undertaken in response to observations from previous earthquakes. While improvements are still largely in response to past earthquakes today, new seismological and engineering techniques allow the development of design criteria for future likely earthquakes. Building design is also going beyond the prevention of collapse with the goal of reducing the costs of future earthquakes in addition to the number of fatalities. One of the challenges is implementation of earthquake-resistant designs, both for new construction and for the existing building stock.

4.21.4.1 Earthquake-Resistant Design

4.21.4.1.1 Lateral forces

Following the 1891 Nobi, Japan, earthquake that killed 7000 people, John Milne laid the foundation for the building codes that were to follow (Milne and Burton, 1891). He detailed the poor performance of modern masonry construction which had recently been introduced to replace the more traditional wood construction in an effort to mitigate fires, and described the great variability in damage to buildings over short distances due to the effect of soft versus hard ground. He also emphasized the need to design buildings to withstand the horizontal forces associated with earthquakes rather than just vertical forces. Similar observations were made following the 1906 San Francisco earthquake by the Lawson Commission (1908).

After the 1908 Messina-Reggio earthquake in southern Italy, which killed 83 000, Panetti proposed that buildings be designed to withstand a horizontal force in proportion to their vertical load. He suggested that the first story should be able to withstand 1/12th the weight of the overlying stories and the second and third stories should be able to withstand 1/8th (Housner, 1984). In Japan, Toshikata Sano made a similar proposal. In 1915, he recommended that buildings should be able to withstand a lateral force, V, in proportion to their weight, W, such that $V = CW$, where C is the lateral force coefficient expressed as a percentage of gravitational acceleration. But it was not until the 1923 Kanto earthquake which killed 100 000 that Sano's criteria became part of the Japanese Urban Building Law Enforcement Regulations released in 1924 (Whittaker *et al.*, 1998). In the Japanese regulations, C, was set at 10% g. Following the 1925 Santa Barbara earthquake in the US, several communities adopted Sano's criteria with $C = 20\%$g. Sano's recommendation was also adopted in the first release of the US Uniform Building Code in 1927, where the value of C was dependent on the soil conditions (National Research Council, 2002).

4.21.4.1.2 Strong-motion observations

While building codes were mandating earthquake-resistant designs as early as the 1920s, there were still no instrumental observations of the actual ground motions responsible for building damage. Milne and colleagues designed and built the first effective seismographs in Japan in the late 1880s. The first instruments in the US were installed at the Lick Observatory of UC Berkeley in 1887 (Lawson, 1908). By the 1920s, seismological observatories had been established around the world, but they were designed to measure the weak (low-amplitude) motion resulting from distant earthquakes. It was not until the 1930s that broadband strong (high-amplitude) motion instruments were available, capable of recording both the low- and high-frequency shaking responsible for the damage to buildings. The 1933 Long Beach earthquake provided the first instrumental recording in which PGAs of 29%g in the vertical and 20%g in the horizontal were observed. A larger PGA value of 33%g was observed at EI Centro a few years later on an instrument 10 km from the 1940 M 7.1 Imperial Valley earthquake rupture. This remained the largest measured ground motion for 25 years, establishing the EI Centro seismogram as the standard for earthquake engineering in both the US and Japan.

Over the following decades, the strong-motion database grew, but slowly. This changed in 1971 when the M 6.6 San Fernando earthquake struck the Los Angeles region and the number of strong-motion recordings more than doubled. In this earthquake, more than 400 000 people experienced PGA in excess of 20%g, and it became clear that high-frequency PGA varied over short distances while the longer period (10 s) displacements did not (National Research Council, 1971; Hudson, 1972, Hanks, 1975). One instrument located on the abutment of the Pacoima Dam recorded a $1\,\mathrm{m\,s^{-1}}$ velocity pulse

shortly followed by a 120%g acceleration pulse (Boore and Zoback, 1974). The strong-motion database generated by this earthquake played an important role in the updates to the Universal Building Code, which followed in 1976. It is a testament to the importance of strong-motion networks, and the earthquake engineering research they provide for, that the majority of damage in recent US earthquakes (1989 Loma Prieta, 1994 Northridge, and 2001 Nisqually) occurred to structures built prior to the 1976 update to the Uniform Building Code (National Research Council, 2003).

Strong-motion networks continue to provide important waveform data sets for damaging earthquakes. One notable recent example was the 1999 M_w 7.6 Chi-Chi earthquake, which occurred beneath central Taiwan on 20 September 1999. The strong-motion seismic network that had recently been deployed by the Central Weather Bureau across the island provided waveforms at 441 sites, including over 60 recordings within 20 km of the fault ruptures (Lee *et al.*, 2001). In addition to Taiwan, dense strong-motion networks with hundreds of instruments are now operational in Japan and the western US. Many more smaller networks are operational in earthquake prone regions around the world. They all provide crucial data when a large earthquake occurs close by, yet the infrequency of such events makes continuous funding and operation a challenge.

4.21.4.1.3 Strong-motion simulations

Advances in computational capabilities, numerical techniques, and our knowledge of the structure of fault zone regions now make it feasible to simulate earthquakes to provide estimates of likely ground motions in future events. The recent centennial of the 1906 San Francisco earthquake motivated one such study in northern California. In order to simulate ground shaking, a velocity model was first developed for northern California. The geology-based model provides three-dimensional (3-D) velocity and attenuation for the simulation using observed relationships between rock type, depth, and seismic parameters (Brocher, 2005). Seismic and geodetic data available from the 1906 earthquake were used to map the distribution of slip in space and time on the fault plane (Song *et al.*, 2006). Several numerical techniques were then used to simulate the earthquake rupture through the geologic model. The simulations could be calibrated by comparing the calculated peak intensities with observed intensities from the 1906 earthquake which were compiled into a 1906 ShakeMap (Lawson, 1908; Boatwright and Bundock, 2005). Snap-shots from one of the simulations are shown in **Figure 8**, (Aagaard, 2006). The peak intensities generated by the simulations reproduce the prominent features of the 1906 ShakeMap validating the simulations.

Other simulations of the 1989 Loma Prieta earthquake, for which instrumental recording of ground shaking is available, also demonstrate that the simulations replicate the amplitude and duration of the observed shaking at frequencies less than 0.5 Hz (Aagaard, 2006; Dolenc *et al.*, 2006). Given likely slip distributions of future earthquakes, these simulations can now provide estimates of the ground shaking in the form of complete seismic waveforms. The results of one study on the southern San Andreas Fault are shown in **Figure 9** (Olsen *et al.*, 2006). The source rupture is along the San Bernardino Mountains and Coachella Valley segments, which are considered more likely to rupture in the coming decades as they have not ruptured since 1812 and 1960. The slip distribution of the 2002 M_W 7.9 Denali, Alaska, earthquake was used for the rupture after scaling it for an M 7.7 rupture. The velocity structure was provided by the SCEC Community Velocity Model (Kohler *et al.*, 2003), and ground shaking is calculated for frequencies of 0–0.5 Hz just as in the northern California simulations. When the rupture propagates from the southeast to the northwest, the directivity effect produces large amplitude ground motions in the Los Angeles metropolitan region. When the fault rupture is to the east of Los Angeles, the chain of sedimentary basins (the San Bernardino, Chino, San Gabriel, and Los Angeles basins) running westward from the northern termination of the rupture funnels seismic energy toward the downtown. The seismograms superimposed on **Figure 9** show velocities of more than $3\,m\,s^{-1}$. When the rupture propagates to the southeast, the ground shaking in LA is an order of magnitude smaller (Olsen *et al.*, 2006).

These simulations are providing new insights into seismic wave propagation and help identify the geologic structures that control strong ground shaking. The uncertainties in the predicted ground shaking result from limitations in the velocity models, the numerical techniques, and the unknown future slip distributions. However, these simulations allow us to explore the range of possible ground motions that we might expect for earthquake ruptures that are evident in the geologic record but not the historic or instrumental records.

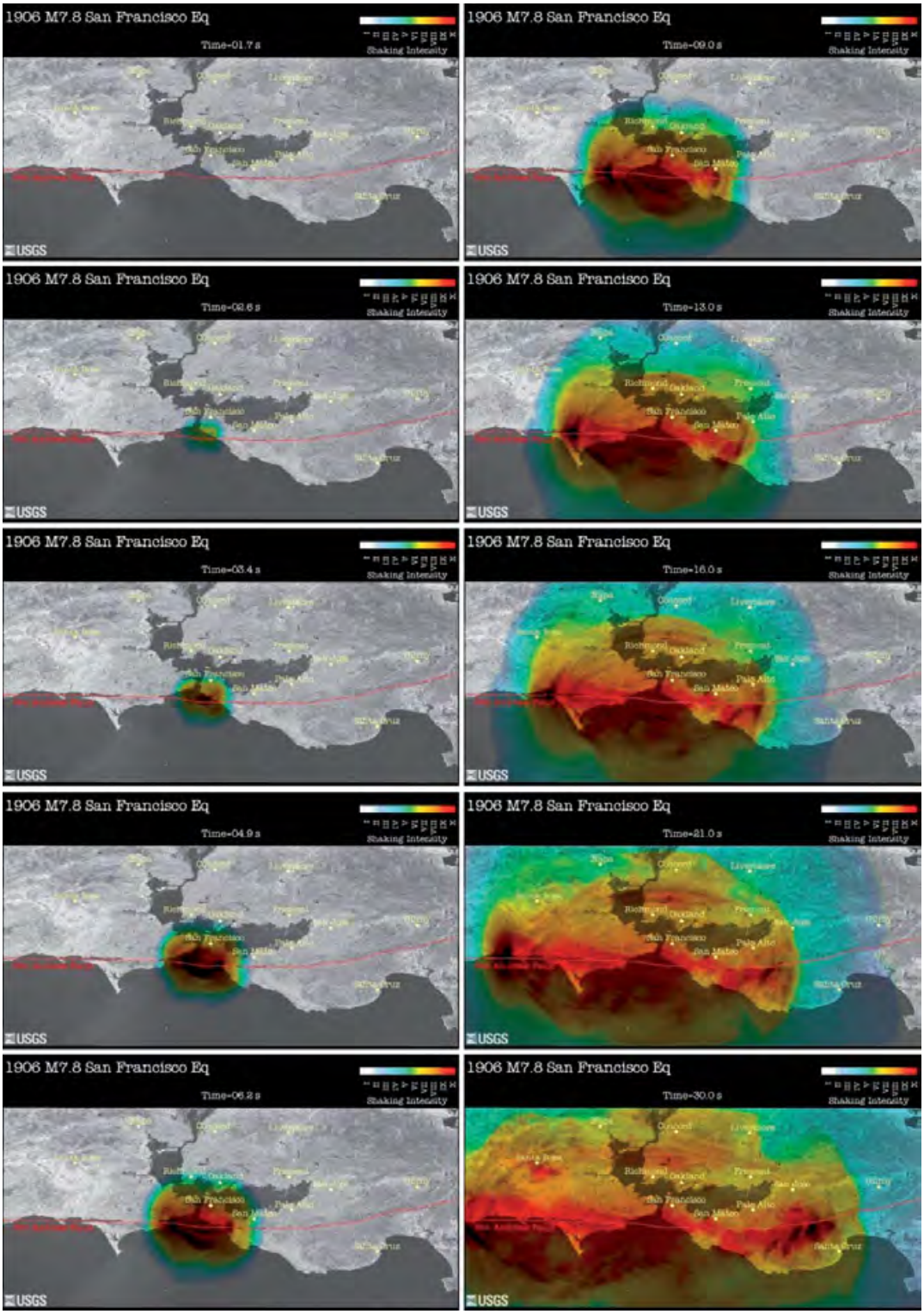


Figure 8 Simulation of the 1906 rupture along the San Andreas Fault. Each map shows the San Francisco Bay Area, north is to the left, and the San Andreas Fault in red. The sequence of snapshots show the peak ground shaking intensity (MMI) 1.7, 2.6, 3.4, 4.9, 6.2, 9.0, 13.0, 16.0, 21.0 and 30.0 s after the rupture initiates. Figures provided by Brad Aagaard (2006). See http://earthquake.usgs.gov/regional/nca/1906/simulations/.

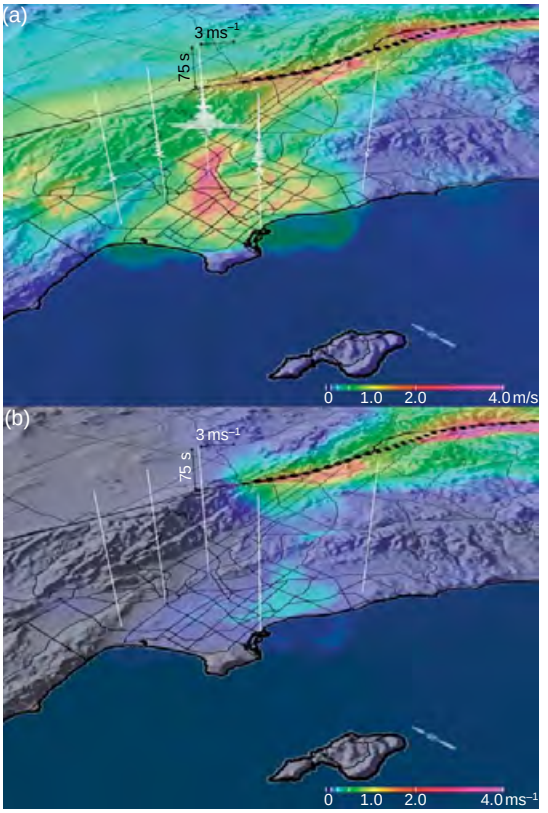


Figure 9 Simulation of an M 7.7 rupture on the southernmost segments of the San Andreas Fault. The section of the fault to rupture is shown by the string of black squares. The color pallet shows the peak ground velocity. (a) Rupture from the south to the north showing the funneling of energy toward the Los Angeles basin west of the rupture. (b) When rupturing from north to south, the amplitudes in the Los Angeles basin are an order of magnitude smaller. From Olsen, *et al*. (2006) Strong shaking in Los Angeles expected from southern San Andreas earthquake. *Geophysical Research Letters* 33: L07305.

4.21.4.1.4 New seismic resistant designs

As strong-motion waveforms became available to the engineering community, the complexity of surface ground motions and their interaction with buildings became apparent. Rather than containing a dominant period, the seismic waveforms were found to be more like white noise over a limited frequency range. Housner *et al.* (1953) proposed to reduce waveforms to a response spectrum, which is the maximum response of single degree-of-freedom oscillators with different natural periods and (typically) 5% internal damping to a recorded waveform. When the response is multiplied by the effective mass of a building, it constrains the lateral force the building would experience and should therefore be able to sustain.

The response spectra are still widely used today, but numerical techniques now allow for much more complex nonlinear modeling of buildings during ground shaking. Such modeling allows testing of new seismic resistant designs using past earthquake recordings as well as future earthquake scenarios. The 1994 M_W 6.7 Northridge earthquake in southern California exposed a vulnerability in steel moment frame buildings. Moment frames resist the lateral forces in an earthquake through bending in the rigidly connected beams and columns. Due to construction practices and the use of nonductile welds, a substantial number of connections fractured in the earthquake. The Universal Building Code was updated accordingly in 1997 (UBC97). But the question remains as to how these building will behave in a future larger magnitude earthquake. Krishnan *et al.* (2006) explored this question using a numerical simulation of two M_W 7.9 earthquakes on the section of the San Andreas that last ruptured in 1857. They first calculated synthetic waveforms at various locations across southern California, and then simulated the effect of the ground shaking at each location on two 18-story steel moment frame buildings, one based on pre-UBC97 code, and one that was post-UBC97. Krishnan *et al.* concluded that if the rupture propagated north-to-south (toward Los Angeles), then the pre-UBC97 building would likely collapse in the San Fernando Valley, Santa Monica, west Los Angeles, Baldwin Park, Compton, and Seal Beach. The post-UBC97 building would likely survive in most locations except the San Fernando Valley. This type of modeling is currently confined to the academic community; however, there is the potential to bring the lessons learned to bear on future construction practices.

Building codes for most buildings are currently focused on 'life safety', the prevention of fatalities in an earthquake. Fatalities mostly occur due to building collapse. The goal of building codes is therefore to prevent collapse in order to get everyone out alive. With a few exceptions, codes are not intended to keep buildings in service after an earthquake and a building that performed 'well' may still need to be demolished. Earthquake engineering is now looking beyond life safety to further reduce the damage to a building at specific levels of ground shaking. Performance-based seismic design (PBSD) is one approach which focuses on what to achieve rather than what to do. The implementation of PBSD concepts will therefore lead to buildings that combine the current prescriptive building codes to prevent

collapse with owner-selected design components to reduce the damage to economically acceptable levels. As a result, we can expect not only reduced fatalities in future earthquakes but also reduced economic losses which would be a reversal of the current trend of increasing economic losses (National Research Council, 2003). This poses challenges for both the seismological and engineering communities. While it is the low-frequency energy that is responsible for damage to buildings, damage to the building content is more sensitive to higher frequencies, greater than the frequency content of current ground motion simulations. For the engineering community, PBDS requires much more detailed knowledge of the performance of building components than the current prescriptive methods.

Building code requirements for critical facilities such as nuclear power plants, dams, hospitals, bridges, and pipelines are usually greater than the life safety standard currently used for homes and offices. The design criteria are continued operation for safety reasons, for example, dams and nuclear power plants, or to provide recovery services in the aftermath of an earthquake, for example, hospitals. The engineering of these facilities is usually site specific. One example of successful engineering of a critical facility is the Trans-Alaska Pipeline, a 48-inch diameter pipeline carrying over 2 million barrels of North Slope oil to the Marine Terminal at Valdez every day. The pipeline crosses three active fault traces and was designed to withstand the maximum credible ground shaking and displacements associated with each. One of the intersected faults is the Denali Fault, where the pipeline was designed to accommodate a right-lateral strike-slip displacement of up to 6 m by constructing the supports on horizontal runners. The 3 November 2002 M_w 7.9 Denali earthquake ruptured over 300 km of the Denali, Totschunda, and Susitna Glacier faults, including the section beneath the pipeline. The displacement at the pipeline was 5.5 m, and there was only minor damage to some of the supports which had been displaced several meters by the rupture (Sorensen and Meyer, 2003).

Structural control is another relatively new approach to reducing the impact of large earthquakes on various structures. The concept is to suppress the response of a building by either changing its vibration characteristics (stiffness and damping) or applying a control force. There are active, semiactive, and passive types of structural control. Active control systems are defined as those that use an external power source. The active mass damper is one such device where an auxiliary mass is driven by actuators to suppress the swaying of a building. Kajima Corporation applied this technique to its first building in 1989, and the device is capable of suppressing the response of the building to strong winds and small to medium earthquakes. The high power demand limits its effectiveness for large earthquakes. Passive systems rely on the viscoelastic, hysteretic, or other natural properties of material to reduce or dampen vibrations. Base isolation is one example of a passive system in which large rubber pads separate a building from the ground. These pads shear during strong shaking, reducing the coupling between the building and the ground. These devices have the advantage that they require no external power, little or no maintenance, and perform well in large earthquakes. There are now over 200 buildings around the world with base isolation systems. Finally, semiactive systems use a combination of the two approaches in that the building response is actively controlled but using a series of passive devices. Active variable stiffness and active variable damping devices are currently being used as part of semiactive systems. These semiactive systems have been installed in a few buildings in Japan as they are still in the development mode, but, as with PBSD, they hold the promise of reducing not only the number of fatalities, but also the economic losses associated with future earthquakes.

4.21.4.2 The Implementation Gap

There are two implementation gaps that seriously negate the effectiveness of earthquake-resilient building design. The first is the large variability in their application or enforcement in different countries; the second is that building codes are generally only applicable to new construction.

4.21.4.2.1 The rich and the poor

Earthquake-resistant design has been proven effective and building codes that include earthquake provisions have been adopted in most countries that have experienced multiple deadly earthquakes (Bilham, 2004). However, while the number of earthquake fatalities in rich countries is estimated to have decreased by a factor of 10, presumably due to better buildings and land use (Tucker, 2004), the number of fatalities in poor countries is projected to increase by a factor of 10. The 1950 M 8.6 Assam earthquake in

India killed 1500 people, but it is estimated that a repeat event in the same location would kill 45 000 people (Wyss, 2004), an increase by a factor of 30 in a region where the population has increased by a factor of 3. Similarly, a repeat of the 1987 M 8.3 Shillong earthquake would kill an estimated 60 times as many people as in 1987 (Wyss, 2004). During that period the population has increased by a factor of 8, again suggesting an order of magnitude increase in the lethality of earthquakes. This increase is largely due to the replacement of single-story bamboo homes with multi-story, poorly constructed, concrete frame structures, often on steep slopes (Tucker, 2004).

Berberian (1990) investigates the earthquake history in Iran. He concludes that the adoption of building codes has had little or no effect, largely due to lack of enforcement. The enforcement gap was also identified after the 1999 Izmit earthquake in Turkey as a major contributor to the 20 000 fatalities. Better implementation and enforcement therefore remain a priority in many earthquake prone regions. However, the socioeconomic situation in many of these countries leaves earthquake risk reduction low on the priority list of development agencies. Most aid organizations continue to operate in a response mode to natural disasters rather than a preventative one. One notable exception is GeoHazards International (http://www.geohaz.org), who are working to introduce earthquake-resistant building practices to local builders in regions of high seismic risk.

4.21.4.2.2 The new and the old

Building codes only apply to new construction. As is clear from the history of earthquake-resistant building design, every major earthquake to date has provided lessons in how not to construct buildings. Unreinforced masonry was banned for public schools in California after the 1933 Long Beach earthquake. In the most recent earthquakes, problems with moment frame buildings and the dangers of soft story buildings were identified. After each of these earthquakes, building codes are updated. The vast majority of buildings are therefore not up to current code. Several hundred billion dollars are spent every year on construction in seismically hazardous areas of the US. It is estimated that the additional earthquake-related requirements of building codes account for ~1% of this investment; the cost of making new buildings seismically safe is therefore small (Office of Technology Assessment, 1995). In contrast, the cost of retrofitting existing buildings is much higher, around 20% of the value of the building for most construction types. In addition to the cost, buildings usually need to be vacated during the retrofit causing additional disruption to the occupants. One example of the retrofitting gap comes from a 2001 study of hospital seismic safety in California (Office of Statewide Health Planning and Development, 2001). The study estimated that over a third of the state's hospitals were vulnerable to collapse in a strong ($6.0 < M < 6.9$) earthquake. In Los Angeles County more than half were vulnerable, and the ratio rises to two in three in San Francisco. The total cost of initial improvements required by state law after the 1994 Northridge earthquake totaled $12 billion; in Los Angeles County, the bill was greater than the total assessed values of all hospital property. Hospitals are considered critical infrastructure, which is why they are required to retrofit by law, but given these economic realities the extent of the retrofits remains to be seen.

The high cost and inconvenience of retrofitting, combined with the uncertainty in the benefit, means that few buildings are retrofitted. However, some institutions and governmental bodies have risen to the challenge. One example of an institution stepping forward to tackle this problem is UC Berkeley (Comerio *et al.*, 2006). The university campus sits astride the Hayward Fault, considered to be one of the most hazardous faults in the SFBA. Since the university was founded, it has had a commitment to the safety of its students, faculty, and staff, and seismic resistant designs have been used across campus. Following the 1971 San Fernando earthquake which caused some damage to another University of California (UC) campus, weaknesses in current building practices were identified and the Universal Building Code was updated in 1976. In 1978, the UC system adopted a seismic safety policy and undertook a review of buildings across the Berkeley campus. Key buildings including University Hall, which housed the system-wide administration at the time, high-rise residence halls, and some key classroom buildings and libraries were retrofitted.

The 1989 Loma Prieta, 1994 Northridge, and 1995 Kobe earthquakes demonstrated how relatively modern buildings were still susceptible to damage during earthquakes and refocused the university on seismic safety. A complete review of campus buildings was ordered in 1996, and it was determined that one-third of all space on campus was rated as poor or very poor, that is, susceptible to collapse in an

earthquake. In 1997, the SAFER program was initiated to retrofit or replace seismically hazardous buildings across campus for life safety. The financial commitment was $20 million per year for 20 years. The most hazardous buildings were retrofitted first and the program continues today. At the same time that the SAFER program was being formulated, Mary Comerio conducted a study of the broader social and economic impacts of future earthquakes. One of the conclusions was that the campus would likely have to close for one or more semesters after an earthquake on the Hayward Fault. This posed a long-term threat to the university's existence as many students, faculty, and staff would likely move elsewhere during this period and not return. The seismic retrofit program was therefore expanded to include business continuity as a goal in addition to life safety and incorporated elements of performance-based design.

The City of Berkeley has also shown leadership in developing innovative programs to motivate the seismic retrofitting of buildings. One such program is the transfer tax incentive. On purchasing a home, one-third of the transfer tax payable to the city is available for approved seismic retrofitting of the home. This typically amounts to several thousand dollars each time a home changes hands. While an individual homeowner may not fully retrofit the home, as properties change hands over time the building stock becomes more seismically safe. This program, in concert with other city retrofit incentives, has resulted in over 80% of single-family homes being at least partially retrofitted in the city, and an estimated 35% are fully retrofitted, making Berkeley one of the most improved cities for seismic safety in the Bay Area (Perkins, 2003).

It is even more of a challenge to motivate retrofitting of buildings that are not owner occupied. In a program initiated in 2006, the City of Berkeley is targeting the large number of soft story apartment buildings. Soft story buildings have large openings in walls on the ground floor, which, as recent earthquakes have demonstrated, makes them vulnerable to collapse. The openings most commonly allow access to parking under the building or store fronts. Under the new city ordinance, soft story buildings are first identified on a city list and owners are notified. The owner is then required to notify existing and future tenants of the earthquake hazard and postprominent seismic hazard signs. The owners are also required to have an engineering assessment of the seismic safety of the buildings and make the information available to the city. The program is designed to provide an incentive for owners to retrofit their buildings. The effectiveness of the program will depend on the extent to which tenants are concerned about seismic safety and whether there are alternative accommodations.

4.21.5 Short-Term Mitigation: Real-Time Earthquake Information

The expansion of regional seismic networks combined with the implementation of digital recording, telemetry, and processing systems provides the basis for rapid earthquake information. This process is often referred to as real-time seismology and involves the collection and analysis of seismic data during and immediately following an earthquake so that the results can be effectively used by the emergency response community and, in some cases, for early warning (Kanamori, 2005). One of the first reported calls for real-time earthquakes information came in 1868 following two damaging earthquakes in SFBA in just 3 years. Following the failure of a 'magnetic indicator' for earthquakes, J. D. Cooper suggested the deployment of mechanical devices around the city to detect approaching ground motion and transmit a warning to the city using telegraph cables (Cooper, 1868). Unfortunately, his system was never implemented.

In California, the first automated notification systems provided earthquake location and magnitude information. They used the Real-Time Picker (RTP) and became operational in the mid-1980s. RTP identified seismic arrivals on single waveforms and estimated the signal duration providing constraints on earthquake location and magnitude (Allen, 1978, 1982). In the early 1990s, the systems were further developed to integrate both short-period and broadband information. The Caltech/USGS Broadcast of Earthquakes (CUBE) (Kanamori *et al.*, 1991) and the Rapid Earthquake Data Integration (REDI) Project (Gee *et al.*, 1996; 2003), in southern and northern California, respectively, provided location and magnitude information to users within minutes via pagers.

In Japan, real-time earthquake information systems have been developed in parallel with those in the US. By the 1960s single seismic stations were already being used to stop trains during earthquakes. After the 1995 Kobe earthquake, the Japanese government initiated a program to significantly increase

the seismic instrumentation across the country with multiple seismic networks. The strong-motion Kyoshin Network (K-Net) has over 1000 stations across the entire country with a constant station spacing of 25 km (Kinoshita, 2003). In addition, most of the ~700 short-period instruments deployed in boreholes (Hi-Net) also have strong-motion instruments at the top and bottom of the borehole (KiK-Net). Finally, a lower-density broadband seismometer network consisting of ~70 instruments with a typical station spacing of 100 km spans the entire country. These networks are operated by the National Research Institute for Earth Science and Disaster Prevention (NIED). All data are telemetered in real-time and is available via the web (http://www.bosai.go.jp). The Japan Meteorological Agency (JMA) also operates a seismic network across the country which is used for real-time earthquake information.

4.21.5.1 Ground Shaking Maps: ShakeMap and Beyond

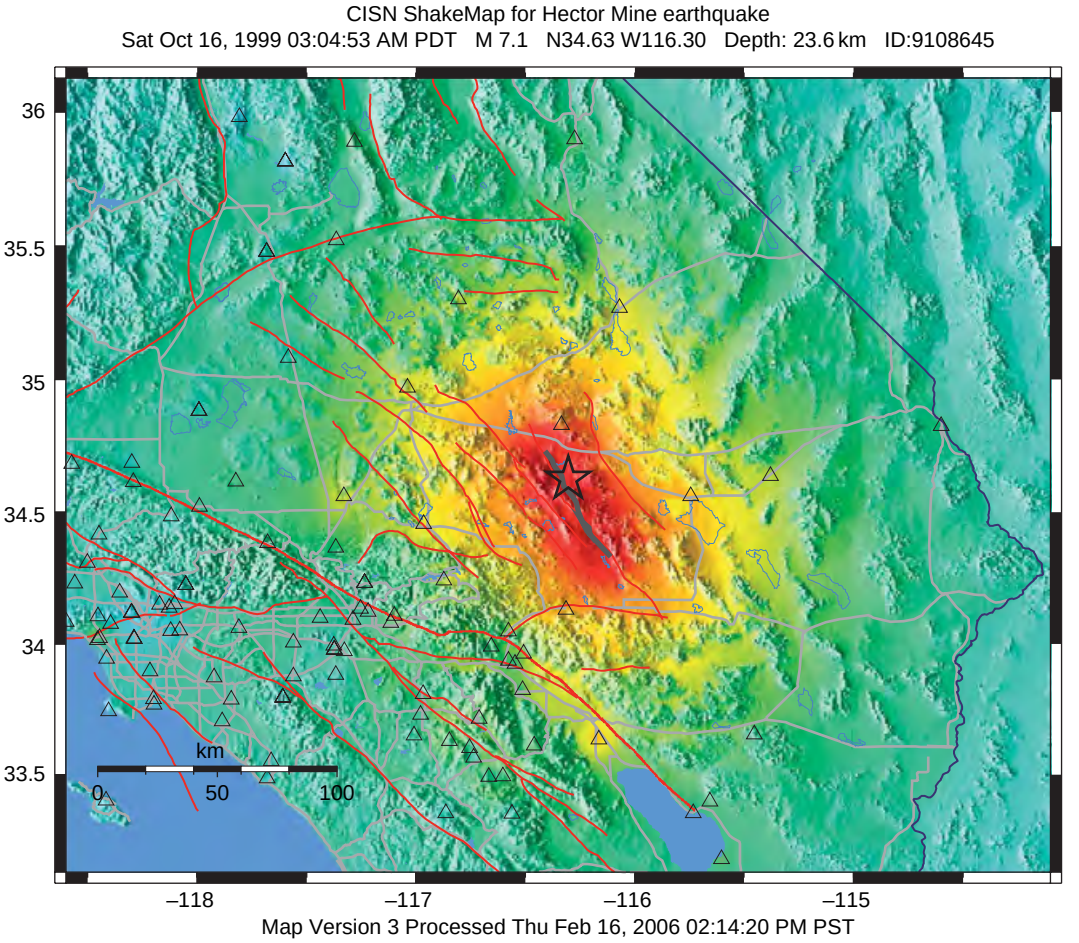
Following the 1994 Northridge earthquake, the TriNet project (Mori *et al.*, 1998; Hauksson *et al.*, 2001) was designed to integrate and expand seismic networks and monitoring in southern California. In both the Northridge and 1989 Loma Prieta earthquakes, strong ground shaking occurred away from the epicenter, and there was a need to go beyond point source information and provide better estimates of the locations of likely damage to the emergency response community. In the 1995 Kobe earthquake, it was many hours until the central government in Tokyo was aware of the full extent of damage to the city of Kobe delaying rescue and recovery efforts (Yamakawa, 1998), again emphasizing the need for rapid automated ground shaking information after major earthquakes.

4.21.5.1.1 ShakeMap

The development and implementation of ShakeMap (Wald *et al.*, 1999) was the response of the seismological community. The ShakeMap concept is to rapidly gather ground shaking information following an earthquake and integrate it into a map of peak ground shaking distribution. While the concept is simple, the implementation is complex, as data from different instrument types with a highly heterogeneous distribution must be integrated. The ShakeMap methodology is triggered by the identification of an earthquake, typically with $M \geq 3$, and first gathers PGA and PGV data from seismic instruments in the proximity of the earthquake. The system must wait several minutes for all stations within a few hundred kilometers to record peak ground shaking and telemeter the data to the central processing site.

Once at the central site, individual station data is first corrected for site amplification effects so they represent observations at uniform 'rock' sites. An empirical attenuation relation for an earthquake of the observed magnitude within the region is then adjusted to provide the best-fit relation for the ground shaking as a function of distance. The attenuation relation is used to generate a map of predicted rock-site ground shaking at all locations. This map is adjusted to match local station observations providing a map of ground shaking controlled by the observations close to seismic stations and the best-fit attenuation relation where there are no data. Finally, adjustments are made for site amplification effects based on mapped geology in the region. In addition to providing maps of PGA and PGV, ShakeMap also combines this data and uses scaling relations to provide estimates of instrumental modified Mercalli intensity (MMI) (Wald *et al.*, 1999). MMI was developed prior to modern seismic instrumentation, but still provides a useful description of the felt ground shaking and damage. More detailed information is available in the ShakeMap manual (Wald *et al.*, 2005).

The methodology was in place for the 1999 M_W 7.1 Hector Mine earthquake providing a test of the real-time earthquake information system (Hauksson *et al.*, 2003). A location and preliminary local magnitude estimate of 6.6 were first available 90 s after the event origin time. An energy magnitude of 7.0 was available 30 s later. These estimates were broadcast via email, the web, and the CUBE pager system within minutes. The first ShakeMap was produced within 4 min of the event. This initial map was generated using observed peak ground shaking and the best-fit attenuation relation, assuming that the ground shaking decayed as a function of distance from the epicenter. As there was only one station within 25 km of the rupture, near-fault ground shaking was estimated based on the attenuation relations. Over the following hours, ShakeMap was updated using information on the finiteness of the fault based on aftershock locations, finite source inversions, and field observations. Broadband waveforms from more distant sites were used to model the rupture improving the estimates of near-fault shaking



Perceived shaking	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
Potential damage	None	None	None	Very light	Light	Moderate	Moderate/heavy	Heavy	Very heavy
Peak acc.(%g)	*<.17*	*.17–1.4*	*1.4–3.9*	*3.9–9.2*	*9.2–18*	18–34	34–65	65–124	>124
Peak vel.(cm s^{-1})	<0.1	0.1–1.1	1.1–3.4	3.4–8.1	8.1–16	*16–31*	*31–60*	*60–116*	*>116*
Instrumental intensity	I	II–III	IV	V	VI	VII	VIII	IX	X+

Figure 10 ShakeMap for the 1999 M_W 7.1 Hector Mine earthquake. This version includes the finiteness of the fault rupture which became available in the hours after the earthquake. There is only one seismic station within 25 km of the rupture, so attenuation relations describing MMI as a function of distance from the fault constrain the near-field MMI estimates. The star shows the epicenter and the line represents the finite extent of the fault. The color pallet shows the instrumental MMI.

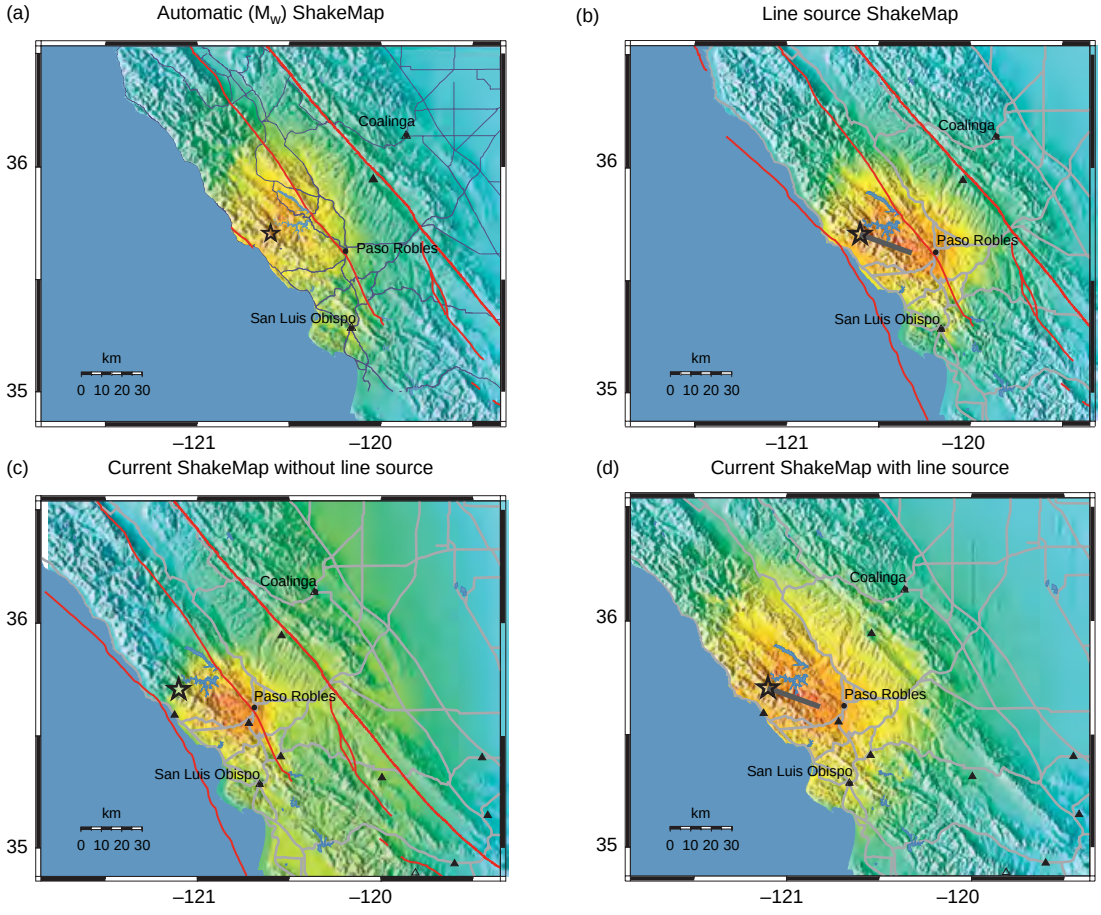
(Dreger and Kaverina, 2000). The final version is shown in **Figure 10**.

4.21.5.1.2 Rapid finite source modeling

The ShakeMap approach works best in regions with dense station coverage. The observed ground motions then control the contouring of the maps. However, the success of ShakeMap has resulted in a desire to generate maps in regions where the station coverage is sparse to nonexistent. Broadband seismic stations can be used to model the finiteness of the source and improve the ShakeMap (e.g., Dreger and Kaverina, 2000; Ji *et al.*, 2004). The integration of rapid and automated finite source modeling into ShakeMap-type products represents one of the new directions in seismic hazard mitigation.

The value of finite source information was demonstrated by the 2003 M_W 6.5 San Simeon earthquake in central California (Hardebeck *et al.*, 2004; Dreger *et al.*, 2005). The seismic station distribution is

sparse in the region, resulting in only three observations of peak ground shaking close to the event in real time. The initial ShakeMap for the event (**Figure 11(a)**) is therefore dominated by the event location and magnitude estimate from which the radial attenuation relation is defined. In fact, the ruptured fault plane extended to the east from the hypocenter, resulting in stronger ground shaking to the east than suggested by this initial ShakeMap. **Figure 11(d)** shows the best estimate of ground



Perceived shaking	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
Potential damage	None	None	None	Very light	Light	Moderate	Moderate/heavy	Heavy	Very heavy
Peak acc.(%g)	*<.17*	*.17–1.4*	*1.4–3.9*	*3.9–9.2*	*9.2–18*	18–34	34–65	65–124	>124
Peak vel.(cm s^{-1})	<0.1	0.1–1.1	1.1–3.4	3.4–8.1	8.1–16	*16–31*	*31–60*	*60–116*	*>116*
Instrumental intensity	I	II–III	IV	V	VI	VII	VIII	IX	X+

Figure 11 ShakeMaps for the 2003 M_W 6.5 San Simeon earthquake in central California. Black triangles are seismic stations, the star is the epicenter, and the black line represents the finite extent of the fault. The color pallet shows the instrumental MMI. (a) The automated ShakeMap generated without any finite source information. There are only three stations constraining the ground shaking estimates on the map. (b) The ShakeMap once the length and geometry of the finite source were included based on the information provided by the real-time finite source model. (c) The ShakeMap derived from all available ground motion observation today (including those for which the waveform data had to be transported by hand on magnetic tape) but without any finite source information. (d) The best estimate of the distribution of ground shaking intensity available today. This incorporates the finite extent of the fault and all stations in the region. Modified from Dreger DS, Gee L, Lombard P, Murray MH and Romanowicz B (2005) Rapid finite-source analysis and near-fault strong ground motions; application to the 2003 M_W 6.5 San Simeon and 2004 M_W 6.0 Parkfield earthquakes. *Seismological Research Letters* 76: 40–48.

shaking available today for comparison; it includes data that were not available in the initial hours after the event.

A real-time finite-source inversion scheme was developed for this scenario by Dreger and Kaverina (2000) using data from the 1992 Landers and 1994 Northridge earthquakes. Although the codes were not automated at the time of 1999 Hector Mine earthquake, they were able to use the offline version to determine finite-source variables and forward calculate ground motions within 5 hours of the event. The now-automated approach (Dreger and Kaverina, 2000) first determines a moment tensor which typically takes 6–9 min. A series of finite-source inversions are then used to explore model space. The moment tensor provides two possible fault planes and the size of the rupture based on moment scaling relations (Somerville *et al.*, 1999). The data are inverted for a series of line sources to test the two moment-tensor nodal planes and a range of rupture velocities. These results are available 11–20 min after the event. At this stage, the orientation and length of the fault plane can be provided to ShakeMap, allowing the ground motion to be estimated as a function of distance from the surface projection of the fault plane rather than distance from the epicenter. A 2-D inversion usually follows, providing a better description of the kinematics of the fault rupture. Finally, the kinematic model can be integrated with near-fault Green's functions to simulate near-fault waveforms, all within ~30 min of an earthquake (Dreger and Kaverina, 2000; Kaverina *et al.*, 2002).

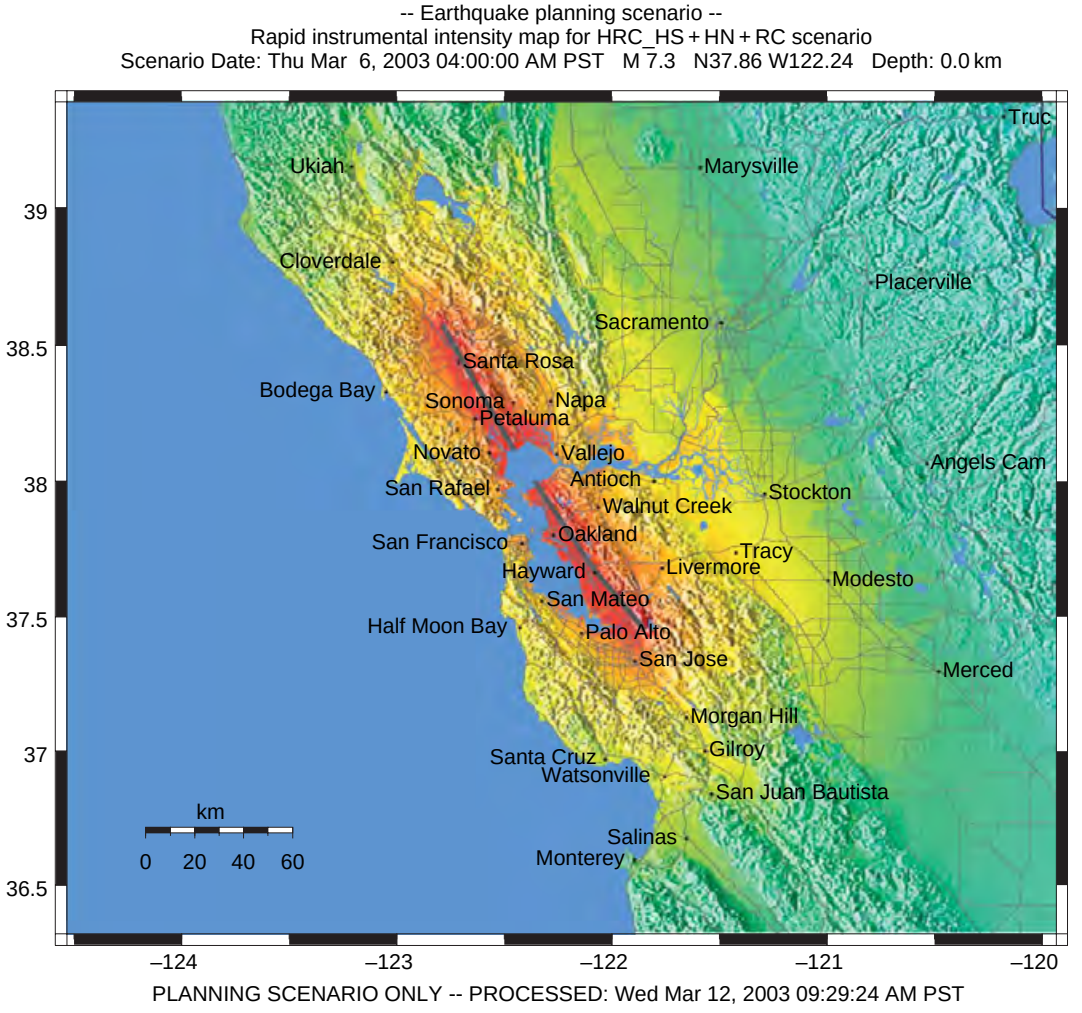
The first event in which the ShakeMap was rapidly updated with finite source information was the 2003 San Simeon earthquake (Dreger *et al.*, 2005). The earthquake occurred in a sparsely populated rural area and most of the damage occurred in the town of Paso Robles 35 km southeast of the rupture where two people were killed. The line-source inversion was complete 8 min after the event and the 2-D inversion and predicted ground motions were available after 30 min. The ShakeMap was updated using the length and geometry of the fault plane derived from the finite source as shown in **Figure 11(b)**. The inclusion of the fault plane resulted in increased estimates of ground shaking at Paso Robles. The initial point-source ShakeMap estimated MMI of V–VI. With the fault plane included in ShakeMap the MMI increased to VII–VIII (compare **Figures 11(a)** and **11(b)**), which is in line with observed damage.

4.21.5.1.3 Applications of ShakeMap

Since its inception, ShakeMap has become a great success, both within the emergency response community for whom it was originally designed, and also with the broader public. While the 1999 Hector Mine earthquake was felt widely across the Los Angeles basin, the ShakeMap showed that the earthquake was fairly distant, centered in the Mohave Desert (**Figure 10**). This information provided for an appropriately scaled response. One Caltrans bridge crew member reported: "I can't tell you how much time and money was saved knowing where to look [for damage]." ShakeMaps are now routinely generated in Nevada, Utah, the Pacific Northwest, and Alaska in addition to California (visit http://earthquake.usgs.gov/eqcenter/shakemap/). Other earthquake-prone regions around the world are also using and developing similar tools. The ShakeMap output also includes GIS shape files of ground shaking levels for use in loss estimation calculations such as HAZUS. These loss estimates are now routinely performed in the hours after moderate and large earthquakes to guide response and recovery.

ShakeMap has also become a tool for public information and education. On the day of the Hector Mine earthquake – ShakeMap's debut – more than 300 000 people visited the website. After smaller, felt earthquakes, website visits reached hundreds per second. In response to this public interest, media maps were designed with the TV audience in mind. These simplified versions of ShakeMap are routinely produced and often used in media coverage following earthquakes. Perhaps the best example of the public interest in the ShakeMap concept is the birth of Community Internet Intensity Maps (CIIMs), better known as "Did you feel it?" These MMI maps are generated automatically using reports of ground shaking intensity provided by the public using an internet portal (http://earthquake.usgs.gov/eqcenter/dyfi.php). These reports are averaged by zip code and provide maps that are very similar to the instrumental MMI ShakeMaps. The CIIMs generate thousands of reports after a felt earthquake, the maximum to date was just under 30 000 after an M 5.2 near Anza California in June 2006 (Wald *et al.*, 2006b). In 2004, the USGS extended the CIIM system to allow for international data collection. These ShakeMap-type products have extended the reach and the complexity of earthquake information provided to the public. This provides an inherent educational benefit as the consumers become more informed about earthquake hazards.

The ShakeMap products for the technical user have also been expanding. Maps of the response spectral acceleration at 0.3, 1.0, and 3.0 s periods are important for estimating the effects of the shaking on particular types of buildings. This information is also available for past significant earthquakes, prior to the inception of ShakeMap, and thus provides a history of the ground shaking experienced by a particular building. These past earthquake maps are also useful for planning and training purposes in preparation for future events. Probabilistic assessments of future likely earthquakes, such as those shown for the SFBA above, have also been used to generate scenario ShakeMaps which can be used in loss estimation and also for training. A scenario ShakeMap for a rupture of the Hayward–Rodgers Creek Fault is shown in **Figure 12**. Finally, ShakeCast is a new mechanism



Perceived shaking	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
Potential damage	None	None	None	Very light	Light	Moderate	Moderate/heavy	Heavy	Very heavy
Peak acc.(%g)	*<.17*	*.17–1.4*	*1.4–3.9*	*3.9–9.2*	*9.2–18*	18–34	34–65	65–124	>124
Peak vel.(cm s^{-1})	<0.1	0.1–1.1	1.1–3.4	3.4–8.1	8.1–16	*16–31*	*31–60*	*60–116*	*>116*
Instrumental intensity	I	II–III	IV	V	VI	VII	VIII	IX	X+

Figure 12 Scenario ShakeMap for an M 7.3 rupture of the Hayward–Rodgers Creek Fault. This is one of the earthquake rupture scenarios identified by WG02 (2003) and was assigned a 1% probability of occurrence by 2032.

for the delivery of ShakeMap which can also be used to trigger user-specific post-earthquake response protocols. For example, utilities, transportation agencies, and other large organizations can automatically determine the shaking at their facilities, set thresholds for notification, and notify responsible staff when appropriate. More information on the range of rapid post-earthquake information products provided by the USGS is available online at http://earthquake.usgs.gov/.

4.21.5.1.4 Global earthquake impact: PAGER

All of the rapid post-earthquake information discussed above is seismic hazard information. However, it is the seismic 'risk', that is, the impact of an earthquake, which is more desirable for most consumers. For emergency services personnel, they respond to locations where the greatest hazard intersects the built environment. ShakeCast is intended to provide sophisticated users with the necessary tools to assess the most likely damage to facilities provided the fragility is known. In an ambitious new project, the USGS National Earthquake Information Center (NEIC) is developing a methodology to convert ground shaking hazard into an assessment of impact on the local population. The Prompt Assessment of Global Urban Earthquakes for Response (PAGER) methodology aims to first estimate the distribution of ground shaking and then estimate the number of fatalities (Earle *et al.*, 2005).

To estimate the distribution of ground shaking, that is, a ShakeMap, for a global event the minimum required data are the earthquake location and magnitude, which are routinely determined for global earthquakes with $M > 5$ by the NEIC. Using available attenuation relations and site corrections derived from the local topography, an initial estimate of the distribution of ground shaking can be made. Additional data that can be input as available include recorded local ground motions, ground shaking intensities reported through the CIIM system, and information about fault finiteness. The finite source information can be derived from a range of sources including aftershock distributions, broadband waveform inversion of teleseismic data (e.g., Ji *et al.*, 2004), and field observation in the hours and days after an event (Wald *et al.*, 2006a). Combining the ShakeMap with population distribution, the number of people experiencing ground shaking at various intensities can be estimated. **Figure 13** shows an example of the PAGER output for the 2005 M_W 7.6 Pakistan earthquake. The methodology estimates that almost 10 million people experienced an MMI of VI, 587 000 experienced MMI IX. Ongoing development of PAGER aims to provide regional fragility information so that these figures can be converted into estimates of the number of casualties.

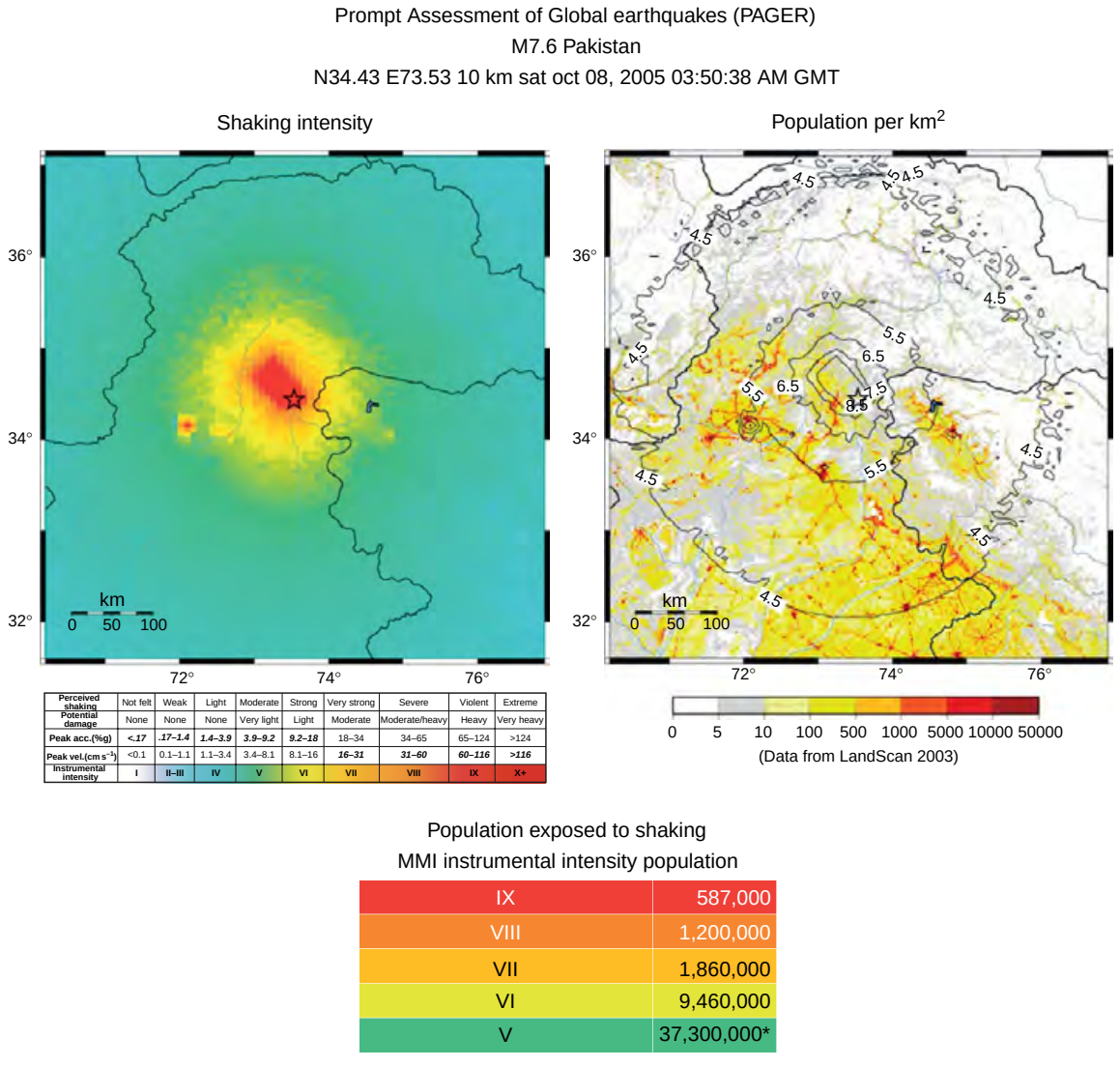
4.21.5.2 Warnings before the Shaking

The tools and methodologies described above provide rapid post-earthquake information in the minutes to hours after an event. This information is critical to the emergency response community and can prevent cascading failures. It is also useful for longer-term planning and training purposes. But the rapid earthquake information system first described by J. D. Cooper (1868) envisioned a warning system designed to provide an alarm prior to ground shaking. Such warning systems could be used for short-term mitigation in the seconds to tens of seconds prior to ground shaking to prevent damage, casualties, and fatalities. The scientific and engineering challenge for any such warning system is to rapidly distinguish between the frequently occurring small and harmless earthquakes and the large damaging ones.

4.21.5.2.1 S-waves versus P-waves

The simplest warning system monitors ground motion and issues an alert or mitigating action when the ground acceleration exceeds some critical threshold. The thresholds are set high, typically $\sim$0.04g (where g is the acceleration due to gravity), which is the level at which buildings and other infrastructure start to experience permanent damage. These systems therefore trigger on S-wave energy and have a zero warning time but also have the benefit that there is no prediction required; the critical ground shaking has been observed when the alert is issued. Such ground shaking detectors are used widely to shut down utility, transportation, and manufacturing systems during earthquakes.

These detectors can be turned into a true warning system, that is, greater than zero seconds warning, by placing them between the earthquake source and the infrastructure or city they are intended to protect. The warning is then transmitted ahead of ground motion electronically. This 'front-detection' approach is being used in Japan and Mexico, where subduction zone earthquakes along the Japan and Middle America Trenches represent a significant hazard for cities further inland. By deploying stations along the coastline adjacent to the earthquake source



Perceived shaking	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
Potential damage	None	None	None	Very light	Light	Moderate	Moderate/heavy	Heavy	Very heavy
Peak acc.(%g)	*<.17*	*.17–1.4*	*1.4–3.9*	*3.9–9.2*	*9.2–18*	18–34	34–65	65–124	>124
Peak vel.(cm s^{-1})	<0.1	0.1–1.1	1.1–3.4	3.4–8.1	8.1–16	*16–31*	*31–60*	*60–116*	*>116*
Instrumental intensity	I	II–III	IV	V	VI	VII	VIII	IX	X+

Population exposed to shaking

MMI instrumental intensity population

IX	587,000
VIII	1,200,000
VII	1,860,000
VI	9,460,000
V	37,300,000*

Figure 13 Output from the USGS National Earthquake Information Center's prototype PAGER system. ShakeMap for the event is shown at upper left, and MMI contours are overlain on the population density (upper right). The number of people experiencing different levels of ground shaking can then be tabulated, below. More information is available at http://earthquake.usgs.gov/eqcenter/pager/.

region, warning can be transmitted electronically ahead of the ground shaking (Nakamura and Tucker, 1988; Espinosa Aranda *et al.*, 1995). A nonzero warning time requires some form of prediction as ground motion parameters must be detected at one location and estimated for another; this introduces uncertainty. In the case of front detection, ground motion parameters close to the epicenter are used to predict ground shaking levels further away. When the geography is conducive, these systems can provide substantial warning times. In the case of the Seismic Alert System in Mexico, the ~300 km between the subduction zone and Mexico City provide for ~70 s of warning as was demonstrated in the 1995 M_W 7.4 Guerrero earthquake (Anderson *et al.*, 1995).

The amount of warning can be increased by using the P-wave rather than the S-wave energy to assess the magnitude or hazard associated with an earthquake. Nakamura (1988) first proposed such an approach which was implemented along the Shinkansen (bullet train) lines in Japan in the 1990s. Nakamura's approach is to use the predominant period, that is, the frequency content, of the first few

seconds of the P-wave to estimate the magnitude of an earthquake. For seismic stations within ~150 km, this measurement is relatively insensitive to epicentral distance and geographical location. Observations from the first few seconds of P-waves recorded within ~150 km of the epicenter of $3 \leq M \leq 8.3$ earthquakes around the world show a scaling relation between magnitude and frequency content, τ_p^{max}, as shown in **Figure 14** (Olson and Allen, 2005). This provides one basis for an early-warning system. The hazard posed by an earthquake is expressed in terms of the magnitude estimate derived from τ_p^{max} of P-waves recorded close to the epicenter. There is uncertainty in magnitude estimates derived from this relation. In the case of the global data set (**Figure 14**) it is ±1 magnitude unit, although these uncertainties can be reduced as discussed below. Similar magnitude–frequency scaling relations have been developed for various regions around the world (Allen and Kanamori, 2003; Nakamura, 2004; Kanamori, 2005; Wu and Kanamori, 2005a, 2005b; Lockman and Allen, 2007; Simons *et al.*, 2006), although the approach also has its detractors (e.g. Rydelek and Horiuchi, 2006) (see also Olson and Allen (2006) response).

In addition to using the frequency content of the P-wave, the amplitude can also be used to assess the forthcoming hazard associated with the S- and surface-wave energy. Wu and Kanamori (2005a, 2005b) explored the use of the peak displacement, velocity, and acceleration within the first 3 s of the P-wave. They found that the lower frequency content of the peak displacement has a high correlation with the peak ground displacement (PGD) and the PGV observed many seconds later. **Figure 15** shows the relation between P_d, the peak ground displacement observed within 3 s of the P-wave arrival, and PGV for 38 $M \geq 5.0$ earthquakes from Taiwan and southern California (Wu *et al.*, in press). P_d observations at a site can therefore be used to assess the forthcoming ground shaking hazard at the same site. P_d, and similar amplitude-derived parameters, can also be used to estimate earthquake magnitude once corrected for attenuation associated with the epicentral distance (Odaka *et al.*, 2003; Kamigaichi, 2004; Wu and Kanamori, 2005; Wu and Zhao, 2006; Wurman *et al.*, in review). In a novel hybrid approach, Cua (2005) uses the amplitude of waveform envelopes to estimate the magnitude of an earthquake. The magnitude determination is derived from the ratio of the peak P-wave displacement and acceleration. Given the different

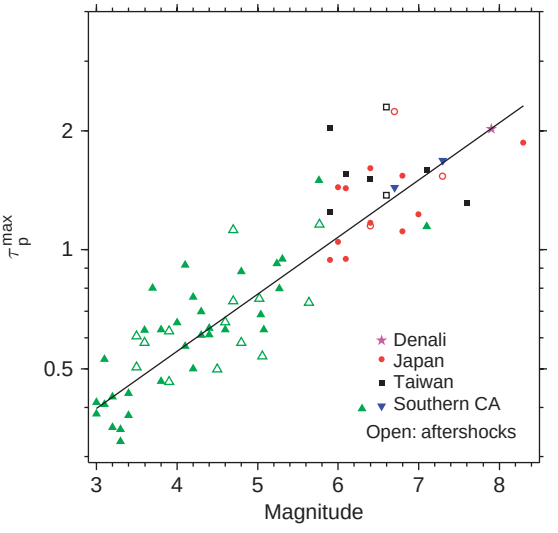


Figure 14 Scaling relation between earthquake magnitude and the frequency content of the first 4 s of the P-wave recorded at stations within 150 km. This global data set consists of 1842 waveforms recorded from 71 earthquakes. The individual values of τ_p^{max} at each station are averaged on this plot. All the event-averaged values fall within a range of ±1 magnitude unit. Modified from Olson E and Allen RM (2005) The deterministic nature of earthquake rupture. *Nature* 438: 212–215.

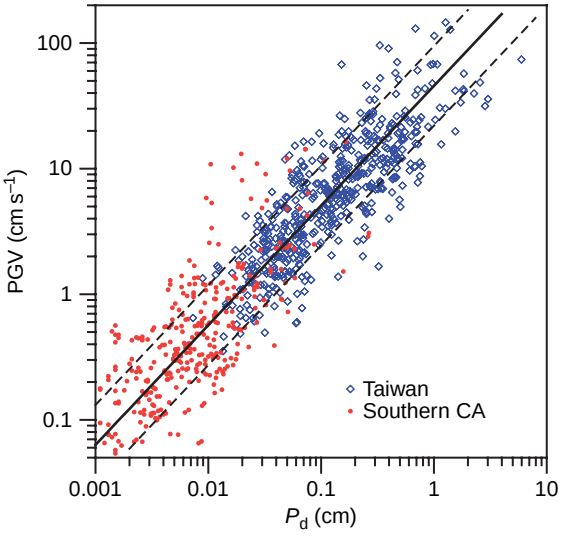


Figure 15 Scaling relation between P_d (the peak displacement observed within the first 3 s of the P-wave) and PGV observed at the same station. Data from 38 M ≥ 5.0 earthquakes in Taiwan (blue) and southern California (red) are shown. Modified from Wu Y-M, Kanamori H, Allen RM, and Hauksson E (in press) Experiment using the tau-c and Pd method for earthquake early warning in southern California. *Geophysical Journal International*.

frequency sensitivities of the acceleration and displacement waveforms, this approach is analogous to the predominant period approach first suggested by Nakamura, but was arrived at independently using a linear discriminate analysis.

4.21.5.2.2 Single-station and network-based warnings

The simplest and most rapid approach to providing a ground shaking warning is to use a single seismic station to record ground motion parameters and issue a warning on site. The UrEDAS system first outlined by Nakamura (1988) provides an estimate of the magnitude and location of an earthquake using just a single three-component seismometer. Criteria for taking mitigating actions are then developed based on the expected peak ground shaking and warning time which are derived from the magnitude and epicentral distance of the event. Alternatively, rather than first estimating the magnitude, the hazard at the station site can be estimated directly. **Figure 15** is an example of this where PGV is estimated directly from P_d. Combining the amplitude and frequency information from P-waves for $M \geq 5.0$ earthquakes in Taiwan, Wu and Kanamori (2005) show that the sites that later experienced damaging ground motion could be distinguished from those that did not. The advantage of this approach is its speed. With this approach, it is possible to provide warning at the epicenter. As soon as information about an earthquake is available at a site, action can be taken. The disadvantage, compared to a multiple-station approach, is greater uncertainties in the hazard estimates and the warning time; in some cases, no estimate of the warning time is available. However, choice of appropriate sites for single-station systems can significantly improve their accuracy. Lockman and Allen (2005) applied a similar methodology to UrEDAS to all broadband velocity stations in southern California. They found one quarter of the stations produced magnitude estimates with errors less than ± 0.3 magnitude units, hypocentral distances within ± 15 km, and back azimuth calculations within ± 20 degrees, but the errors at other stations were larger making some unusable for the purpose of early warning.

A network or regional-based approach is the alternative to single-station systems. By combining information from multiple stations, the uncertainties in hazard estimates and the number of false alarms can be reduced. Network-based approaches typically locate an earthquake and estimate its magnitude as a first step to predicting the expected distribution of ground shaking (Wu and Teng, 2002; Allen and Kanamori, 2003; Kamigaichi, 2004; Cua, 2005; Horiuchi *et al.*, 2005; Allen, in press; Wurman *et al.*, in review). The site-specific peak ground shaking and the time at which it is expected can then be transmitted to users to initiate mitigating actions. When compared with a single-station approach, the cost for users close to the epicenter is a reduced warning time as the system must wait for seismic arrivals at multiple seismic stations and data must be telemetered between sites. However, the introduction of a regional telemetry system increases warning times for users further from the epicenter. For an earthquake detected close to the epicenter, the warning can be transmitted ahead of the ground shaking. This is the front-detection approach described above.

4.21.5.2.3 Warning around the world

It is clear that the most accurate and timely, that is, the most effective, warning systems will combine all of the above approaches making use of information contained in the full waveform and issuing warnings on site as well as taking advantage of a network and telemetry system. **Figure 16** shows the locations of the warning systems now in operation and development around the world. Most make use of hybrid methodologies.

The operational systems are in Japan, Taiwan, Mexico, and Turkey, where warnings are issued to users beyond the seismological community. In Japan, the first alarm seismometers were deployed by Japan Railways in the mid-1960s (Nakamura and Tucker, 1988); these detectors were then developed into the more sophisticated UrEDAS P-wave detection system (Nakamura, 1988) in the early 1990s. Since then, network-based approaches have been developed by both the JMA (Kamigaichi, 2004) and the NIED (Horiuchi *et al.*, 2005). JMA has been testing an early-warning system for general use since February 2004 (Kamigaichi, 2004). In August 2006, they widened the testing to 41 institutions, including railway companies, construction firms, factories, and hospitals. As the public becomes more familiar with the system, they plan to make the information more widely available.

The Central Weather Bureau in Taiwan has been using a virtual subnet approach to rapidly assess magnitude from the S-wave energy of an event. This method requires an average of 22 s for magnitude determination and gives warning to populations

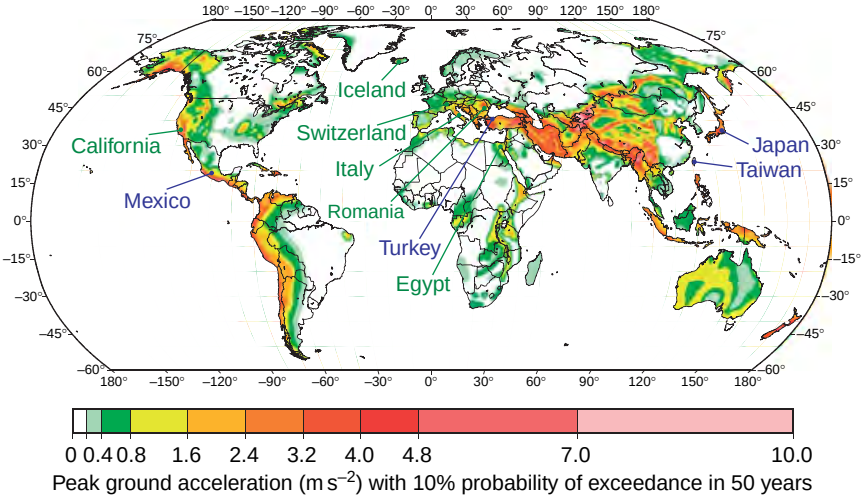


Figure 16 Map showing the locations of earthquake early-warning systems currently in operation (blue) or development (green) around the world. Operational systems include Japan, Taiwan, Mexico, and Turkey. Systems are in development for California, Egypt, Greece, Iceland, Italy, Romania, and Switzerland. An operational system is defined as one that issues warning information to users outside the seismological community. The locations are overlaid on the GSHAP global seismic hazard map (Giardini *et al.*, 1999).

greater than 75 km away (Wu *et al.*, 1998; Wu and Teng, 2002). The development by Wu and colleagues of P-wave methodologies described above is aimed at increasing the warning times and reducing the blind zone where warnings cannot be provided. Using a network approach, it is estimated that the blind zone would be reduced to 20 km (Wu and Kanamori, 2005). Single-station methodologies could provide warnings at smaller epicentral distances (Wu *et al.*, 2006).

Mexico City's Seismic Alert System (SAS) takes advantage of its geographical separation from the seismic source region along Guerrero Gap subduction zone to the southwest. The front-detection system measures the rate of increase of S-wave energy at stations along the coast to estimate magnitude and transmits this information to the population in Mexico City 300 km away (Espinosa-Aranda *et al.*, 1995). It has been operational since 1991 and transmits its warnings to schools, industry, transportation systems, and government agencies. Finally, Turkey is the most recent member of the early-warning club. Their system triggers when the amplitude of ground motion exceeds some threshold at a network of instruments around the Sea of Marma, providing warning to users in Istanbul (Erdik *et al.*, 2003; Boese *et al.*, 2004).

Development of early-warning systems is also underway across Europe and in the United States. The European Community is currently funding the cooperative development and testing of early warning algorithms in Egypt, Greece, Iceland, Italy, Romania, and Switzerland. In the United States, the California Integrated Seismic Network (CISN) has recently embarked on a project to test various early-warning algorithms to evaluate their performance across the state. The test includes two network-based approaches, the Earthquake Alarm System (ElarmS) and the Virtual Seismologist (Cua, 2005), and a single-station approach, the amplitude and period monitor (Wu and Kanamori, 2005). The goal is to evaluate the real-time performance and strengths of these methodologies in order to develop an optimal hybrid system for the state. In order to get a sense of the capabilities of such a future system, we consider the performance of one of these methodologies, the one most familiar to the author, in more detail.

4.21.5.2.4 ElarmS in California

The Earthquake Alarm System, ElarmS, is a network-based approach to earthquake early warning (Allen and Kanamori, 2003; Allen, 2004; Allen, in press; Wurman *et al.*, in review; http://www.ElarmS.org). The methodology uses the first 4 s of the P-wave arrival at stations in the epicentral region to locate earthquakes in progress and estimate their magnitude. An AlertMap is generated, showing the expected distribution of peak ground shaking in terms of PGA, PGV, and MMI. All available data are collected from all stations every second and the AlertMap is updated. Initially, the AlertMap is based on the location and

magnitude estimates only, and an attenuation relation is used to predict ground shaking. As time proceeds, observations of peak ground shaking near the epicenter are incorporated into the estimate of ground shaking at more distant locations. The predictive AlertMap therefore evolves into an observed ShakeMap during the course of an event.

The ElarmS algorithms were developed using calibration datasets for both southern and northern California. Since February 2006, they have been automatically processing all $M \geq 3.0$ earthquakes in northern California. They are not yet part of the real-time system and are running in an off-line mode. On notification of an earthquake from CISN, they sleep for 10 min to allow waveform data to populate the archive. They then gather all available data and process it without human interaction to generate a timeseries of AlertMaps. Between February and September 2006, there were 83 events processed in this fashion. **Figure 17** shows the AlertMap output for one of the largest events during this period, the M_L 4.7 earthquake near Santa Rosa on 2 August 2006 (local time). The time histories of the magnitude, PGA, PGV, and MMI prediction errors are shown in **Figure 18**. This event was near the Rodgers Creek Fault in a similar location to one of the future hazardous scenario events in the region (WG02, 2003).

The initial detection occurs 3 s after the event origin time (**Figure 17(a)**). The event is located (red star) at the station to trigger (grey triangle) and the warning time across the region is estimated (concentric circles). One second later (**Figure 17(b)**), an additional two stations trigger and the event is relocated using the grid search method. The initial magnitude estimate is also available, derived from the first second of data from the first station to trigger.

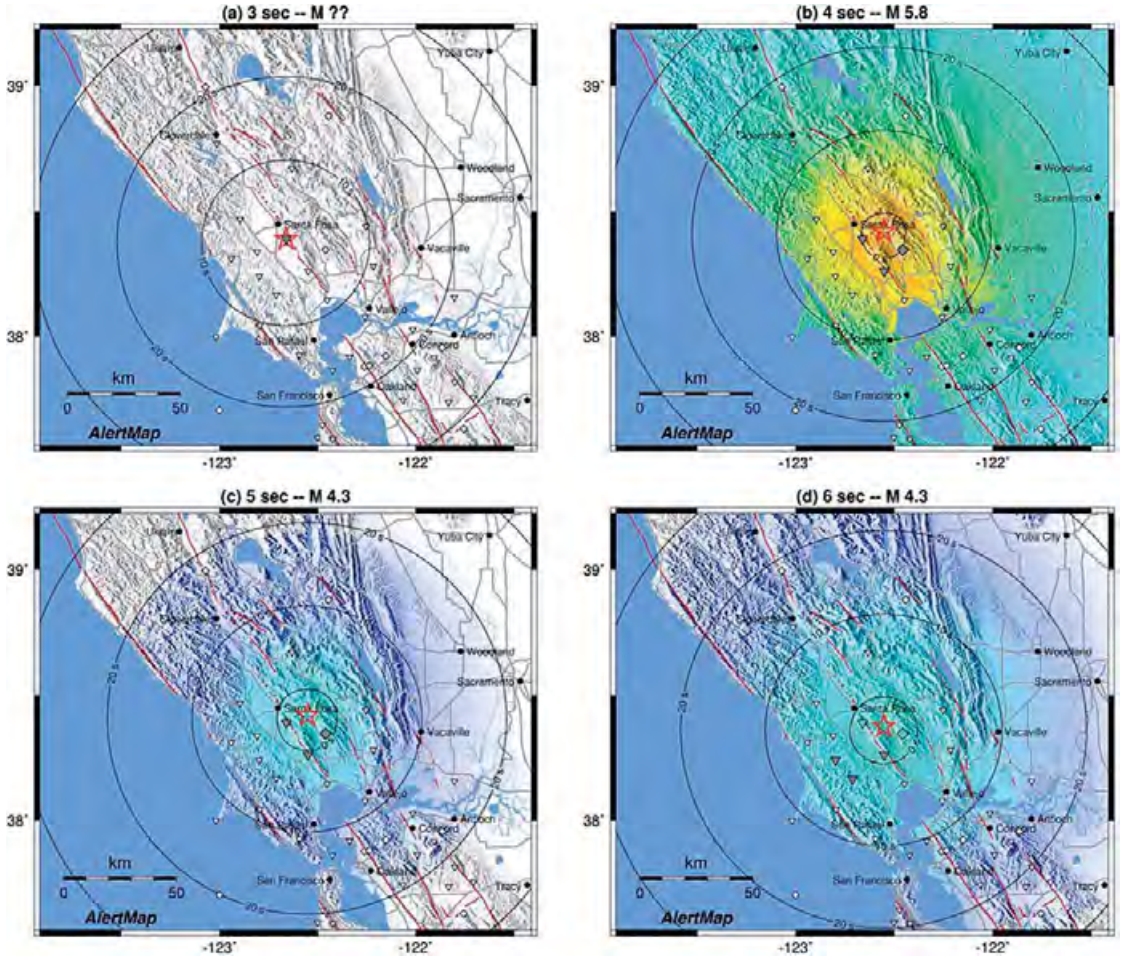
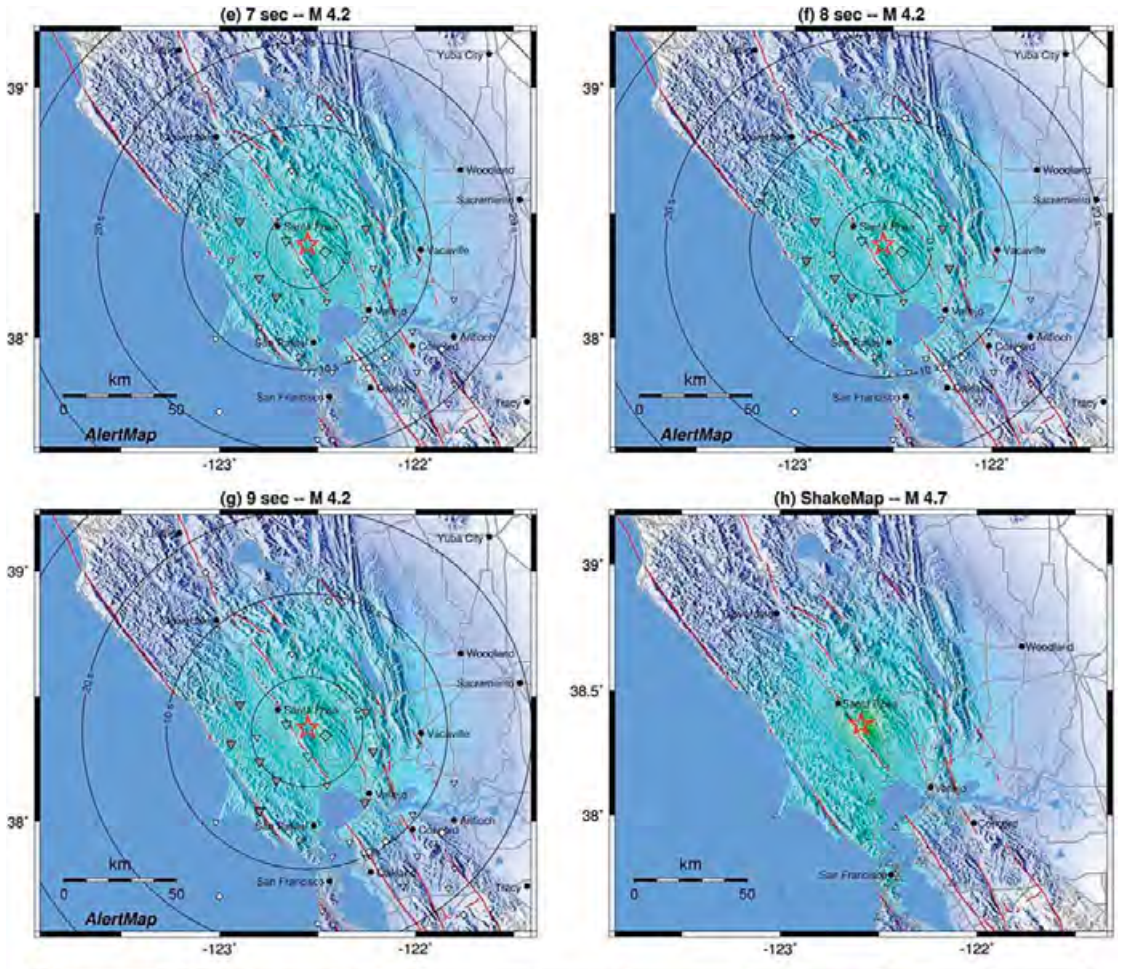


Figure 17 (*Continued*)



PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	*<.17*	*.17-1.4*	*1.4-3.9*	*3.9-9.2*	*9.2-18*	18-34	34-65	65-124	>124
PEAK VEL.(cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	*8.1-16*	*16-31*	*31-60*	*60-116*	*>116*
INSTRUMENTAL INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+

Figure 17 Performance of ElarmS for the M_L 4.7 earthquake near Santa Rosa on 2 August 2006 (local time). (a–g) AlertMap output from the time of initial detection, 3 s after event origin time, for 8 consecutive seconds. (h) The event Shake Map for comparison. The red star is the event epicenter, concentric circles indicate the warning time. Triangles (broadband velocity), inverted triangles (strong motion), and diamonds (collocated velocity and strong motion) show the locations of seismic station. The symbols turn gray when the station triggers and are colored according to the peak ground shaking at the site once it has occurred. The color pallet shows the predicted instrumental MMI for the AlertMaps (a–g) and the 'observed' for the ShakeMap (h).

The initial estimate is high, M 5.8, and the predicted distribution of peak ground shaking is correspondingly high (color pallet). The MMI estimates exceed the actual observations by up to 2 MMI units. One second later (**Figure 17(c)**), magnitude estimates are available from the additional two triggered stations providing an updated event magnitude estimate of M 4.3. This reduces the predicted MMI intensities and reduces the errors in all output parameters (**Figure 18**). This illustrates the benefit of using multiple stations. In this case, waiting one additional second so that magnitude information is available from three rather than one stations significantly reduces the error.

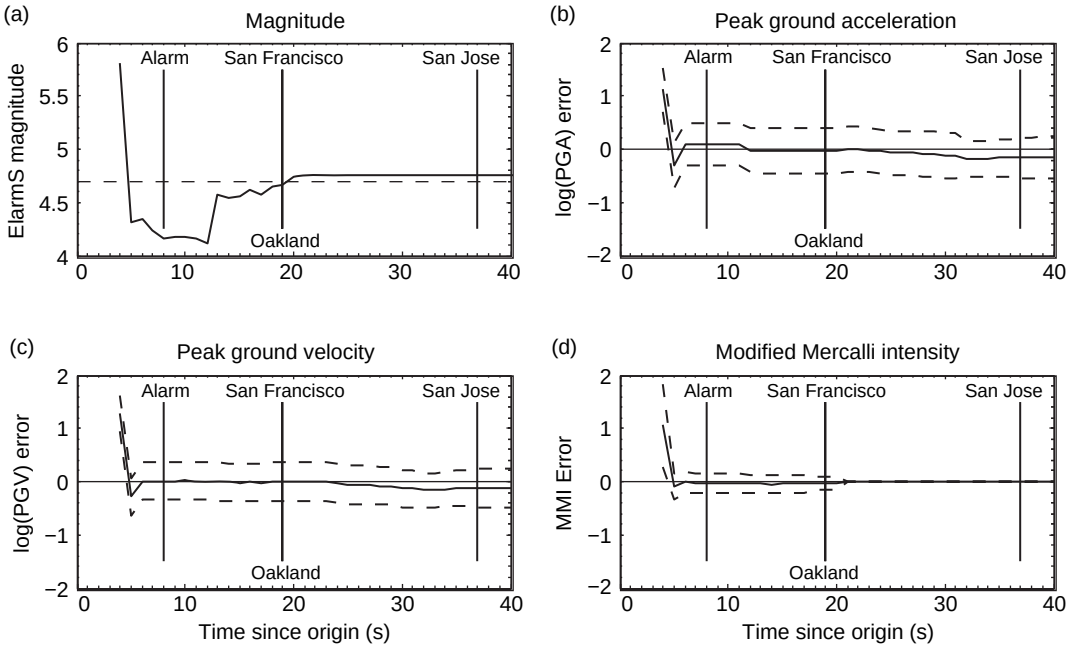


Figure 18 Performance of ElarmS for the M_L 4.7 earthquake near Santa Rosa on 2 August 2006 (local time) as a function of time. (a) ElarmS magnitude estimate; the dashed line is the CISN magnitude of M_L 4.7. (b) Errors in the predicted PGA determined by subtracting the logarithm of the observed from the logarithm of the predicted. Only stations where the peak ground shaking has not yet been observed are included. The dashed lines represent the one-sigma error envelope. (c) Errors in PGV. (d) Errors in MMI. The MMI error goes to zero, as all stations that have not yet observed peak ground shaking after 20 s had a predicted MMI value of I and an observed value of I. The vertical bars indicate the alarm time (4 sec of P-wave data available from 4 sensors) and the time of peak ground shaking in the cities of San Francisco, Oakland, and San Jose.

One second later, just 3 seconds after the initial detection, peak ground shaking is observed at two stations (**Figure 17(d)** – colored triangle and diamond), and these observations are used to adjust the attenuation relations for the region. While the magnitude estimate remains 0.4–0.5 units low for the following 6 s (**Figure 18(a)**), the effect on ground shaking estimates is reduced by the inclusion of these peak ground shaking estimates at the closest stations (**Figures 18(b), 18(c), and 18(d)**). AlertMaps for the following 3s are shown in **Figures 17(e), 17(f),** and **17(g)**. Additional stations trigger providing information for the magnitude estimate, and peak ground shaking is observed at additional sites, but the predicted distribution of ground shaking does not change noticeably. The CISN ShakeMap for this event is shown in **Figure 17(h)** for comparison. The AlertMap from 6 s onward is very similar, the main difference being the slightly stronger ground shaking at the epicenter on the ShakeMap. This is due to the underestimates of the ElarmS magnitude which remains low until 13 s, when it reaches M 4.6. Details of the ElarmS methodology and performance in northern California can be found in Wurman *et al.* (in review).

The continuum of information available about an ongoing earthquake is illustrated in **Figure 18** which shows the changing error in the predictions. Any individual user can decide whether they would rather react to earlier information which has greater uncertainty but also greater warning time, or wait a few seconds for the uncertainty to reduce. This decision can be made in a probabilistic framework (Grasso, 2005; Iervolino *et al.*, in press; Grasso and Allen, in review). When the cost of inaction in a damaging earthquake and the cost of taking mitigating action are known, the appropriate predicted ground shaking threshold for talking action can be defined provided the uncertainty in the prediction is also known. By only taking action when this threshold is reached, the total cost of an earthquake is minimized.

4.21.5.2.5 Warning times

The maximum warning time for the Santa Rosa event is 15 s for San Francisco and Oakland, and 33 s for San Jose in the south bay. This is the time from the initial magnitude estimation until maximum ground shaking in the cities. However, the initial prediction is high, so it would be preferable to wait

at least a few seconds before taking any actions. The 'alarm time' is defined in this chapter as the time at which 4 s of P-wave data are available from four seismic instruments. Application of ElarmS to data sets from southern California, northern California, and Japan shows that the average absolute magnitude error at this time is 0.5 units (Allen, in press; Wurman *et al.*, in review). The alarm time for the Santa Rosa event is shown in **Figure 18**; from alarm time, there is still 11 s warning for Oakland and San Francisco, and 24 s for San Jose. A second M_L 4.7 earthquake occurred in northern California since the automated ElarmS processing began. It occurred on 15 June 2006 near Gilroy south of the bay, and was almost the same distance from San Francisco and Oakland as the 1989 Loma Prieta earthquake. At alarm time for the Gilroy event, when the magnitude estimate was 4.3, there was 3 s of warning for San Jose, 20 s warning for Oakland, and 22 s for San Francisco. In the Loma Prieta earthquake, 84% of the fatalities occurred in Oakland and San Francisco. Therefore, in a repeat of the Loma Prieta earthquake with a warning system in place, there could be ~20 s warning in the locations where most casualties occur.

Warning times for earthquakes in California range from zero seconds up to over a minute depending on the location of the earthquake with respect to a population center. Heaton (1985) used a theoretical distribution of earthquakes in southern California to estimate the range of warning times as a function of ground shaking intensities at the warning location. He showed that for the larger, most damaging earthquakes there could be more than 1 min of warning.

Using the ElarmS methodology, we can estimate the warning time for any earthquake location. **Figure 6** contours the warning time the city of San Francisco would have for an earthquake with an epicenter at any location across the region. The warning time is the difference between the alarm time for the earthquake given the current distribution of real-time seismic stations and the time at which peak ground shaking would occur in San Francisco. An additional 5.5 s has been deducted from the warning time to account for telemetry delays of the existing network (which could be reduced).

While **Figure 6** shows the warning time for all earthquake locations, future damaging events will likely occur on specific faults. These likely future damaging earthquake scenarios were identified by WG02. As probabilities are associated with each earthquake scenario, probabilities that an earthquake with a particular warning time will occur by 2032 can be estimated. **Figure 19** shows that distribution of the warning times for these scenario earthquakes ranges from −7 to 77 s where a negative warning time means the alert time was after the peak ground shaking in San Francisco. The most likely warning times range from −7 to 25 s, which are due to earthquakes on the numerous faults throughout the SFBA (**Figure 6**). The long tail extending to 77 s is due to events on the San Andreas extending to the north. The scenario ShakeMaps for each event (e.g., **Figure 13**) provide an estimate of the ground shaking intensity in San Francisco. The probability distribution shown in **Figure 19** is colored accordingly. The inset to **Figure 19** shows the probability there will be more or less than 0, 5, 10, 20, and 30 s warning and shows that it is more likely that there will be more than 10 sec of warning for the most damaging events. If the telemetry delay was reduced, or more stations were deployed to the north of the SFBA, then more than 20 s warning is likely for these most damaging earthquakes. One of the most deadly scenarios for the city of San Francisco is an M 8, 1906-type earthquake, with a rupture initiating near Cape Mendocino and propagating south. In this scenario, there could be over 1 min of warning time. Probabilistic warning time distributions for various other locations are also available (Allen, 2006).

4.21.5.2.6 Future development

The large-magnitude, most damaging earthquakes are when a warning is of most value and also when the warning times can be the greatest. The accuracy of the ground shaking predictions for these large-magnitude events is significantly improved by knowledge of the finiteness of the rupture. Neither ElarmS, nor any of the other operational early warning systems, currently account for fault finiteness. This is therefore an active area of research. One approach is to monitor the displacement across fault traces allowing instantaneous identification of rupture. This requires instrumentation along all faults and also that the rupture occurs on a previously identified fault at the surface. Some of the earliest proposals for warning systems used wires across fault traces to detect slip. Today, real-time GPS stations could be used to monitor displacement and would be sensitive to slip on fault planes at greater distances. An alternative approach is identifying which seismometers are near-field and which are far-field during the rupture in order to map the rupture extent. Yamada and Heaton (2006) are using the radiated high-frequency energy at near-field stations to approximate the rupture area and the evolving

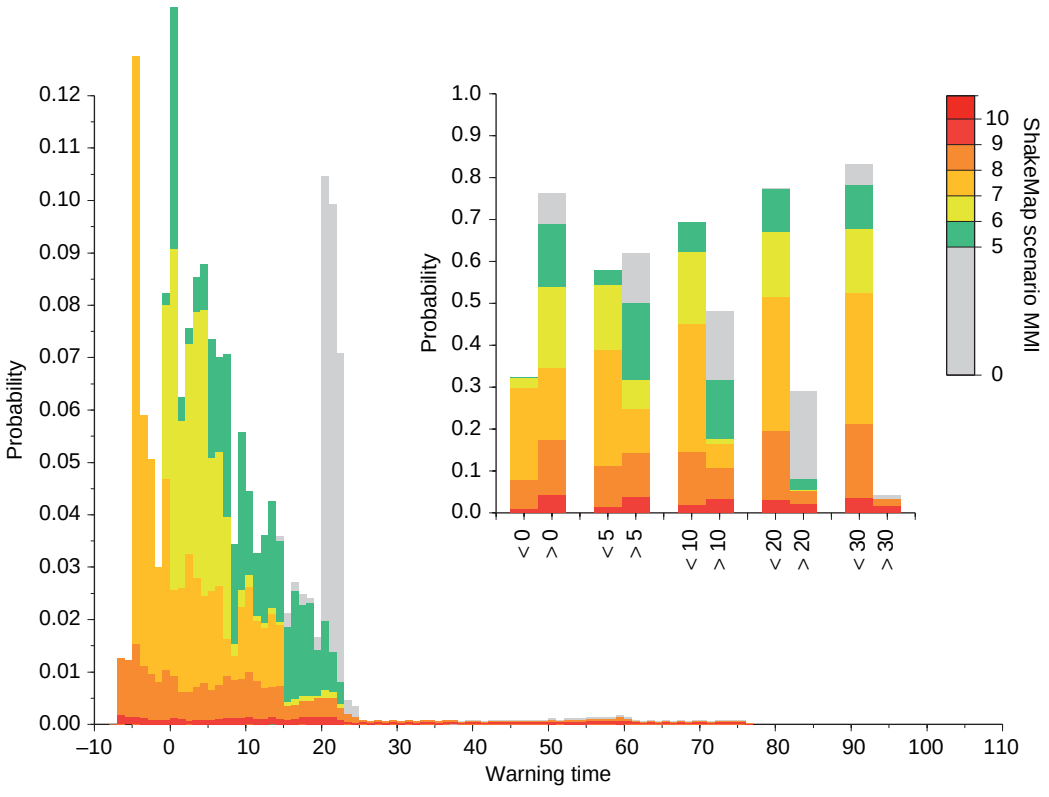


Figure 19 Warning time probability density function for the city of San Francisco. The warning times for all earthquake scenarios identified by WG02 were estimated given the current seismic network and telemetry delays using ElarmS. The range of warning times is −7 to 77 s where a negative warning time means peak ground shaking occurs before the warning is available. The most probable warning times range from −7 to 25 s; the long tail extending to 77 s is due to the San Andreas Fault. The color shows the predicted intensity of ground shaking in the city. The inset shows the probability of more or less than 0, 5, 10, 20, and 30 s warning. It is much more likely there will be greater than zero seconds warning, and the warning times are greater for the most damaging earthquakes. Modified from Allen RM (2006) Probabilistic warning times for earthquake ground shaking in the San Francisco Bay Area. *Seismological Research Letters* 77: 371–376.

moment magnitude in order to estimate the probable rupture length. As these real-time finite-fault techniques are developed, it will be important to incorporate them into early-warning systems.

4.21.5.2.7 Benefits and costs

Warning information from the operational warning systems in Japan, Taiwan, Mexico, and Turkey are currently used by transportation systems such as rail and metro systems, as well as private industries, including construction, manufacturing, and chemical plants. They are also used by utility companies to shut down generation plants and dams, and emergency response personnel to initiate action before ground shaking. In addition, schools receive the warnings allowing children to take cover beneath desks, housing units automatically switch off gas and open doors and windows, and entire complexes evacuate. These same applications would be appropriate for early-warning implementations in many regions around the world and include both automated response by a computerized control system as well as human response (both for personal protection and reduction of damage to infrastructure).

Looking to the future, earthquake engineering is already evolving to incorporate real-time earthquake information from early-warning systems. In Japan, most new high-rise buildings are 'dynamic intelligent buildings' which contain structural control devices to select or change the vibration characteristics of a building, that is, the stiffness or damping (e.g., Housner *et al.*, 1997). Some of these buildings have active control systems which use external power to change or control the building's response to vibrations. Others have passive devices that use hysteretic or viscoelastic properties of material to reduce

vibrations with no external power. More recently, semiactive systems have been developed which use passive devices that are actively put into operation when necessary. Early-warning information is of value to both the active and semiactive types. As more information about the characteristics of forthcoming ground shaking becomes available (such as amplitude and frequency content), the more effective the building's response systems can be.

For personal protection, early warning systems could perhaps be of most value in regions with high seismic hazard and poor implementation of earthquake-resistant building practices. In many of these underdeveloped environments, buildings are typically small single-story dwellings. Homes may be built by the owner using local materials such as mud bricks. Earthquakes in these regions have high fatality rates as buildings collapse on their occupants. For example, the recent 2003 Bam (Iran) and 2005 Pakistan earthquakes together killed over 100 000 people. In these environments, it only takes a few seconds to get out of these buildings, and early-warning systems could provide that time.

The costs of early-warning systems are substantial, but so are the costs of other mitigation strategies and the earthquakes themselves. California currently has ~300 seismic stations that are telemetered in real-time and appropriate for use in an early warning system. Broad implementation of earthquake early warning in the region would require a more robust and redundant seismic network. To install an additional 600 instruments would cost between $6 and $30 million, depending on the instrumentation used. To operate that network would cost between $2 and $6 million per year. In addition to these costs, a system to transmit the warning information would be needed as well as an educational program to teach people how to use the information. For comparison, UC Berkeley is currently retrofitting campus buildings to prevent collapse in future earthquakes. The cost per building is typically $10 to $30 million; retrofit of the historical Hearst Mining building cost $80 million and was made possible by a generous donation. UC Berkeley is spending $20 million per year for 20 years to protect its students and staff in an earthquake, and indeed its very own existence, against a significant earthquake in the region. Implementation of an early-warning system in California is not a replacement for earthquake-resistant buildings and retrofit programs, but there are hundreds of buildings in the SFBA alone like those currently being retrofit on the Berkeley campus which will not be retrofitted. An early-warning system would allow some short-term mitigation strategies for everyone.

Similarly, in regions where there is little or no implementation of earthquake-resistant building practices, a warning system would provide some mitigation of earthquake effects. The costs could perhaps be reduced by using clusters of stations to improve on single-station performance without requiring a full seismic network. The operation of such systems would have to be done locally, requiring a local seismological skill base. Developing this skill base will also perhaps assist in the improvement of building practices, so both long-term building and short-term warning can be used to reduce the costs of future earthquakes.

4.21.6 Conclusion

Progress in seismic hazard mitigation has been substantial – near-zero fatalities from all earthquakes are within our technical capabilities – and yet the cost of earthquakes is still rising, and the number of fatalities continues to increase.

Reducing the cost and fatalities in future earthquakes requires first identifying the hazard and then implementing appropriate mitigation strategies. Our understanding of the earthquake process allows effective long-term forecasts of hazard expressed as the probability of ground shaking above some threshold. Plate tectonics provides the framework for understanding where most future earthquakes will occur. When considered as a stationary time series, the likelihood of future events can be estimated with a degree of confidence. This provides earthquake probability forecasts on timescales of fifty to hundreds of years. Yet, most in the seismology community would agree that there is a time dependence to earthquake hazard, and the probability of a large earthquake increases with time since the last event as stress increases on a fault. The challenge is to estimate the likely time until the next rupture, which is dependent not only on the rate of increasing stress, but also the initial stress, activity on surrounding faults, and changes in the physical properties of the crust. Given these limitations, the uncertainty in hazard forecasts increases as the forecast timescale decreases.

While the public continues to identify short-term earthquake prediction – the high probability of a clearly defined earthquake in a short period of

time – as the solution to earthquake disasters, few seismologists see such predictions as feasible within the foreseeable future. Existing mitigation strategies, when fully implemented, could reduce the impact of earthquakes more than even the most accurate short-term predictions. This is because predictions would only allow people to get out of the danger area, but the infrastructure on which their lives depend would remain.

Mitigation strategies fall into two categories: long term and short term. Long-term mitigation focuses on building infrastructure capable of withstanding earthquake shaking. This approach has been very effective in reducing the number of fatalities in earthquakes, but still new lessons are learned each time there is a large damaging earthquake. New techniques now allow engineers to test designs against the shaking anticipated from future earthquakes. This provides the opportunity to move beyond the current mode dominated by response to what did not work in the last earthquake. Performance-based seismic design is now also providing a framework for reducing the economic impacts of earthquakes in addition to preventing fatalities.

Short-term mitigation is provided by rapid earthquake information systems. Modern seismic networks have been providing location, magnitude, and ground shaking information in the minutes after an event for over a decade. This information has now been widely integrated into emergency response, allowing for more efficient and effective rescue and recovery efforts. But today, many earthquake-prone regions are pushing the limits of rapid earthquake information systems in an effort to provide similar information in the seconds to tens of seconds before the ground shaking. These warning systems provide another opportunity to further reduce the costs and casualties in future earthquakes.

The reduction of seismic risk will be most effective when multiple approaches are used. There is still a surprise component to all large-magnitude earthquakes, which acts as a reminder that we need to be wary of becoming too tuned in our mitigation efforts. By combining earthquake-resistant design to prevent building collapse, warning systems to isolate toxic systems, and rapid response to critical facilities identified as potentially damaged, we can reduce the impact of an earthquake and also accommodate the failure of one component in the system. In another situation, one mitigation strategy might not be economically feasible while another is. It is therefore important to continue development of a full range of methodologies.

Perhaps the greatest challenge in seismic hazard mitigation is implementation of these mitigation strategies in all earthquake-prone regions. While the hazard is now clearly identified on a global scale, implementation is extremely variable. All mitigation is local, and the challenge is to provide the necessary resources to the communities that need them. Implementation requires two components: education and incentives. Education about the risk and available mitigation approaches is the first component. But, even when this information is provided, it can be difficult to motivate action for an event that may or may not occur within any individual's lifetime. Incentives are therefore also necessary and can be offered through legal mandate or economic benefit. As the population continues to grow in underdeveloped nations, where cities are increasingly concentrated in earthquake-prone locations and where current mitigation is least effective, the challenge to bridge the implementation gap could not be greater and of more importance.

Acknowledgments

Gilead Wurman provided several of the figures which were generated with GMT (Wessel and Smith, 1995). Support for this research was provided by USGS/NEHRP Grant # 06HQAG0147.

References

Aagaard B (2006) Finite-element simulations of ground motions in the San Francisco Bay Area from large earthquakes on th San Andreas Fault (abstract). *Seismological Research Letters* 77: 275.

Aki K (1980) Possibilities of seismology in the 1980s. *Bulletin of the Seismological Society of America* 70: 1969–1976.

Allen CR (1976) Responsibilities in earthquake prediction. *Bulletin of the Seismological Society of America* 66: 2069–2074.

Allen R (1978) Automatic earthquake recognition and timing from single traces. *BSSA* 68: 1521–1532.

Allen R (1982) Automatic phase pickers: Their present use and future prospects. *BSSA* 72: S225–S242.

Allen RM (2004) Rapid magnitude determination for earthquake early warning. In: Pecce M, Manfredi G, and Zollo A (eds.) *The Many Facets of Seismic Risk*, pp. 15–24. Napoli: Universita degli Studi di Napoli "Federico II".

Allen RM (2006) Probabilistic warning times for earthquake ground shaking in the San Francisco Bay Area. *Seismological Research Letters* 77: 371–376.

Allen RM (in press) The Elarms earthquake warning methodology and application across California. In: Gasparini P and Zschau J (eds.) *Seismic Early Warning*. Springer.

Allen RM and Kanamori H (2003) The potential for earthquake early warning in southern California. *Science* 300: 786–789.

Anderson J, Quaas R, Singh SK, *et al.* (1995) The Copala, Guerrero, Mexico earthquake of September 14, 1995 ($M_W = 7.4$): A preliminary report. *Seismological Research Letters* 66: 11–39.

Bakun WH (1999) Seismic activity of the San Francisco Bay Region. *Bulletin of the Seismological Society of America* 89: 764–784.

Bakun WH and Lindh AG (1985) The Parkfield, California, earthquake prediction experiment. *Science* 229: 619–624.

Bakun WH and McEvilly TV (1979) Earthquakes near Parkfield, California: Comparing the 1934 and 1966 sequences. *Science* 205: 1375–1377.

Berberian M (1990) Natural hazards and the first earthquake catalogue of Iran. Historical hazards in Iran prior to 1900, vol. 1, 649 pp. Tehran: IIEES.

Bilham R (1988) Earthquakes and urban-growth. *Nature* 336: 625–626.

Bilham R (1996) Global fatalities from earthquakes in the past 2000 years: Prognosis for the next 30. In: Rundel JB, Turcotte DL, and Klein W (eds.) *Reduction and Predicatability of Natural Disasters*, pp. 19–32. Reading, MA: Addison-Wesley.

Bilham R (1998) Death toll from earthquakes. *Geotimes* 43: 4.

Bilham R (2004) Urban earthquake fatalities: A Safer world, or worse to come? *Seismological Research Letters* 75: 706–712.

Boatwright J and Bundock H (2005) Modified Mercalli Intensity Maps for the 1906 San Francisco earthquake plotted in ShakeMap Format, Series. *USGS Open File Report*. 2005–1135.

Boese M, Erdik M, and Wenzel F (2004) Real-Time prediction of ground motion from P-wave records. *EOS, Transactions of the American Geophysical Union Fall Meeting Supplement* 85: Abstract S21A.0251.

Boore DM and Zoback MD (1974) Two dimentional kinematic fault modeling of the Pacoima Dam strong motion recordings of February 9, 1971, San Fernando Earthquake. *Bulletin of the Seismological Society of America* 64: 555–570.

Brocher TM (2005) Compressional and shear wave velocity versus depth in the San Francisco Bay area, California: Rules for USGS Bay Area Velocity Model 05.0.0, Series. *USGS Open-File Report.* 2005–1317.

Burridge R and Knopoff L (1967) Model and theoretical seismicity. *Bulletin of the Seismological Society of America* 57: 341–371.

Comerio MC, Tobriner S, and Fehrenkamp A (2006) Bracing Berkeley: A guide to seismic safety on the UC Berkeley campus, Pacific Earthquake Engineering Research Center, Series, PEER 2006/01.

Cooper JD (1868) Earthquake indicator. *San Francisco Evening Bulletin*, 3 November 1868, p. 10.

Cornell CA (1968) Engineering seismic risk analysis. *Bulletin of the Seismological Society of America* 58: 1583–1606.

Cua GB (2005) *Creating the Virtual Seismologist: Developments in Ground Motion Characterization and Seismic Early Warning*. PhD Thesis, Caltech.

d'Alessio MA, Johanson IA, Burgmann R, Schmidt DA, and Murray MH (2005) Slicing up the San Francisco Bay Area: Block kinematics and fault slip rates from GPS-derived surface velocities. *Journal of Geophysical Research* 110: B06403.

Dolenc D, Dreger D, and Larsen S (2006) 3D Simulations of ground motions in northern California using the USGS SF06 velocity model (abstract). *Seismological Research Letters* 77: 300.

Dreger D and Kaverina A (2000) Seismic remote sensing for the earthquake source process and near-source strong shaking: A case study of the October 16, 1999 hector mine earthquake. *Geophysical Research Letters* 27: 1941–1944.

Dreger DS, Gee L, Lombard P, Murray MH, and Romanowicz B (2005) Rapid finite-source analysis and near-fault strong ground motions; application to the 2003 Mw 6.5 San Simeon and 2004 Mw 6.0 Parkfield earthquakes. *Seismological Research Letters* 76: 40–48.

Dunbar PK, Lockridge PA, and Whitewide LS (2006) Catalog of significant earthquakes 2150 B.C. to the present, Series. *NOAA National Geophysical Data Center Report SE-49*. http://www.ngdc.noaa.gov/nndc/struts/form?t=101650&s=1&d=1 (accessed Jan 2007).

Earle PS, Wald DJ, and Lastowka LA (2005) PAGER – Rapid assessment and notification of an earthquake's impact, Series, USGS Fact Sheet 2005–2035.

Erdik MO, Fahjan Y, Ozel O, Alcik H, Aydin M, and Gul M (2003) Istanbul Earthquake Early Warning and Rapid Response System. *EOS, Transactions of the American Geophysical Union Fall Meeting Supplement* 84: Abstract S42B.0153.

Espinosa-Aranda JM, Jimenez A, Ibarrola G, *et al.* (1995) Mexico City seismic alert system. *Seismological Research Letters* 66: 42–53.

Fedotov SA (1965) Regularities of distribution of strong earthquakes in Kamchatka, the Kuril Islands and northern Japan (in Russian). *Akad. Nauk. SSSR Inst. Fiziki Zemli Trudi* 36: 66–93.

Gee LS, Neuhauser DS, Dreger DS, Pasyanos ME, Uhrhammer RA, and Romanowicz B (1996) Real-time seismology at UC Berkeley: The Rapid Earthquake Data Integration project. *Bulletin of the Seismological Society of America* 86: 936–945.

Gee L, Neuhauser D, Dreger D, Uhrhammer R, Romanowicz B, and Pasyanos M (2003) The rapid earthquake integration project. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake & Engineering Seismology*, pp. 1261–1273. San Diego: Academic Press.

Geller RJ (1996) Debate on evaluation of the VAN method: Editor's introduction. *Geophysical Research Letters* 23: 1291.

Geller RJ (1997) Earthquake prediction: a critical review. *Geophysical Journal International* 131: 425–450.

Geller RJ, Jackson DD, Kagan YY, and Mulargia F (1997) Geoscience - Earthquakes cannot be predicted. *Science* 275: 1616.

Giardini D (1999) The Global Seismic Hazard Assessment Program (GSHAP) - 1992/1999. *Annali Di Geofisica* 42: 957–974.

Giardini D, Grunthal G, Shedlock KM, and Zhang PZ (1999) The GSHAP Global Seismic Hazard Map. *Annali Di Geofisica* 42: 1225–1230.

Gokhberg MB, Morgounov VA, and Pokhotelov OA (1995) *Earthquake Prediction: Seismo-Electromagnetic Phenomina*, 193 pp. Singapore: Gordon and Breach Publishers.

Grasso VF (2005) *Seismic Early Warning Systems: Procedure for Automated Decision Making.* PhD Thesis, Universita delgi Studi di Napoli Federico II.

Grasso VF and Allen RM (in review) Uncertainty in real-time earthquake hazard predictions. *Bulletin of the Seismological Society of America*.

Grunthal G, Bosse C, Sellami S, Mayer-Rosa D, and Giardini D (1999) Compilation of the GSHAP regional seismic hazard for Europe, Africa and the Middle East. *Annali Di Geofisica* 42: 1215–1224.

Gupta SK and Patwardham AM (1988) Earthquake prediction: Present status, 280 pp. Pune: University of Poona.

Hanks TC (1975) Strong ground motion of the San Fernando, California earthquake: Ground displacements. *Bulletin of the Seismological Society of America* 65: 193–225.

Hardebeck JL, Boatwright J, Dreger D, *et al.* (2004) Preliminary Report on the 22 December 2003, M 6.5 San Simeon, California Earthquake. *Seismological Research Letters* 75: 155–172.

Harris RA and Simpson RW (1998) Suppression of large earthquakes by stress shadows: A comparison of Coulomb and rate-and-state failure. *Journal of Geophysical Research* 103: 24439–24451.

Hauksson E, Small P, Hafner K, *et al.* (2001) Southern California Seismic Network: Caltech/USGS element of TriNet 1997–2001. *Seismological Research Letters* 72: 690–704.

Hauksson E, Jones LM, and Shakal AF (2003) TriNet: A modern ground-motion seismic network. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake & Engineering Seismology*, pp. 1276–1284. San Diego: Academic Press.

Heaton TH (1985) A model for a seismic computerized alert network. *Science* 228: 987–990.

Horiuchi S, Negishi H, Abe K, Kamimura A, and Fujinawa Y (2005) An automatic processing system for broadcasting earthquake alarms. *Bulletin of the Seismological Society of America* 95: 708–718.

Housner GW (1984) Historical view of earthquake engineering, *Proceedings of the Eigth World Conference on Earthquake Engineering*, San Francisco, 25–39.

Housner GW, Bergman LA, Caughey TK, *et al.* (1997) Structural control: Past, present, and future. *Journal of Engineering Mechanics-Asce* 123: 897–971.

Housner GW, Martel R, and Alford JL (1953) Spectrum analysis of strong motion earthquakes. *Bulletin of the Seismological Society of America* 42: 97–120.

Hudson DE (1972) Local distribution of strong earthquake ground motions. *Bulletin of the Seismological Society of America* 62: 1765–1786.

Iervolino I, Convertito V, Giorgio M, Manfredi G, and Zollo A (in press) Real-time risk analysis for hybrid earthquake early warning systems. *Journal of Earthquake Engineering*.

Ikeya M (2004) *Earthquakes and Animals*, 295 pp. Singapore: World Scientific Publishing Company.

Isikara AM and Vogel A (1982) *Multidisciplinary Approach to Earthquake Prediction*, 578 pp. Braunschweig: Friodr. Vieweg & Sohn.

Jackson DD and Kagan YY (1993) Comment on ''Seismic gap hypothesis: Ten years after,'' Reply to S.P. Nishenko and L.R. Sykes. *Journal of Geophysical Research* 98: 9917–9920.

Ji C, Helmberger DV, and Wald DJ (2004) A teleseismic study of the 2002 Denali fault, Alaska, earthquake and implications for rapid strong-motion estimation. *Earthquake Spectra* 20: 617–637.

Kagan YY and Jackson DD (1991) Seismic gap hypothesis. *Journal of Geophysical Research* 96: 21419–21431.

Kagan YY and Jackson DD (1995) New seismic gap hypothesis: Five years after. *Journal of Geophysical Research* 100: 3943–3959.

Kamigaichi O (2004) JMA Earthquake Early Warning. *J. Japan Assoc. Earthquake Eng.* 4.

Kanamori H (2003) Earthquake prediction: An overview. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake & Engineering Seismology*, pp. 1205–1216. San Diego: Academic Press.

Kanamori H (2005) Real-time seismology and earthquake damage mitigation. *Annual Review of Earth and Planetary Sciences* 33: 195–214.

Kanamori H, Hauksson E, and Heaton T (1991) TERRAscope and CUBE Project at Caltech. *EOS, Transactions, American Geophysical Union* 72: 564.

Kaverina A, Dreger D, and Price E (2002) The combined inversion of seismic and geodetic data for the source process of the 16 October 1999 M-w 7.1 Hector Mine, California, earthquake. *Bulletin of the Seismological Society of America* 92: 1266–1280.

Kelleher JA, Sykes LR, and Oliver J (1973) Criteria for prediction of earthquake locations, Pacific and Caribbean. *Journal of Geophysical Research* 78: 2547–2585.

Kinoshita S (2003) Kyoshin Net (K-Net), Japan. In: Lee WHK, Kanamori H, Jennings PC, and Kisslinger C (eds.) *International Handbook of Earthquake & Engineering Seismology*, pp. 1049–1056. San Diego: Academic Press.

Kircher CA, Seligson HA, Bouabid J, and Morrow GC (2006) When the big one strikes again – estimated losses due to a repeat of the 1906 San Francisco earthquake. *Earthquake Spectra* 22: S297–S339.

Knopoff L (1996) Earthquake prediction: The scientific challenge. *Proceedings of the National Academy of Sciences of the United States of America* 93: 3719–3720.

Kohler M, Magistrale H, and Clayton R (2003) Mantle heterogeneities and the SCEC three-dimensional seismic velocity model version 3. *Bulletin of the Seismological Society of America* 93: 757–774.

Krishnan S, Ji C, Komatitsch D, and Tromp J (2006) Case studies of damage to tall steel momentframe buildings in southern california during large San Andreas earthquakes. *Bulletin of the Seismological Society of America* 96: 1523–1537.

Langbein J, Borcherdt R, and Dreger D (2005) Preliminary report on the 28 September 2004, M 6.0 Parkfield, Claifornia Earthquake. *Seismological Research Letters* 76: 10–26.

Lawson AC (1908) *The California Earthquake of April 18, 1906*. Washington, DC: Carnegie Institution of Washington.

Lee WHK, Shin TC, Kuo KW, Chen KC, and Wu CF (2001) CWB Free-Field Strong-Motion Data from the 21 September Chi-Chi, Taiwan, Earthquake. *Bulletin of the Seismological Society of America* 91: 1370–1376.

Lockman A and Allen RM (2005) Single station earthquake characterization for early warning. *Bulletin of the Seismological Society of America* 95: 2029–2039.

Lockman A and Allen RM (2007) Magnitude-period scaling relations for Japan and the Pacific Northwest: Implications for earthquake early warning. *Bulletin of the Seismological Society of America* 97: 140–150.

Lomnitz C (1994) *Fundamentals of Earthquake Prediction*, 326 pp. New York: John Wiley & Sons.

McCann WR, Nishenko SP, Sykes LR, and Krause J (1979) Seismic gaps and plate tectonics: Seismic potential for major boundaries. *Journal of Pure and Applied Geophysics* 117: 1082–1147.

McCue K (1999) Seismic hazard mapping in Australia, the southwest Pacific and southeast Asia. *Annali Di Geofisica* 42: 1191–1198.

McGuire RK (1993) Computation of seismic hazard. *Annali Di Geofisica* 34: 181–200.

Milne J and Burton WK (1981) *The Great Earthquake in Japan, 1891, 2nd edn., Lane*, 69 pp. Japan: Crawford & Co, Yokohama, and 30 plates.

Mogi K (1985) *Earthquake Prediction*, 335 pp. London: Academic Press.

Mori J, Kanamori H, Davis J, and Hauksson E (1998) Major improvements in progress for southern California earthquake monitoring. *EOS* 79: 217–221.

Nakamura Y (1988) On the urgent earthquake detection and alarm system (UrEDAS). *Proc. 9th World Conf. Earthquake Eng* VII: 673–678.

Nakamura Y (2004) UrEDAS, Urgent Earthquake Detection and Alarm System, Now and Future. *Proc. 13th World Conf. Earthquake Eng.* August 2004, Paper No. 908.

Nakamura Y and Tucker BE (1988) Earthquake warning system for Japan Railways' Bullet Trains: Implications for disaster prevention in California. *Earthquakes and Volcanoes* 20: 140–155.

National Research Council (1971) *The San Fernando Earthquake of February 9, 1971: Lessons from a Moderate Earthquake on the Fringe of a Densely Populated Region*, 24 pp. Washington, DC: National Academy Press.

National Research Council (2002) *Living on an Active Earth: Perspectives on Earthquake Science*, 418 pp. Washington, DC: National Academies Press.

National Research Council (2003) *Preventing Earthquake Disasters: The Grand Challenge in Earthquake Engineering*, 172 pp. Washington, DC: National Academy Press.

Nishenko SP (1989) Earthquakes, hazards and predictions. In: James DE (ed.) *The Encyclopedia of Solid-Earth Geophysics*. New York: Van Nostrand Reinhold 260–268.

Nishenko SP (1991) Circum-Pacific seismic potential: 1989–1999. *Journal of Pure and Applied Geophysics* 135: 169–259.

Nishenko SP and Sykes LR (1993) Comment on ''Seismic gap hypothesis: Ten years after'' by Y.Y. Kagan and D.D. Jackson. *Journal of Geophysical Research* 98: 9909–9916.

Odaka T, Ashiya K, Tsukada S, Sato S, Ohtake K, and Nozaka D (2003) A new method of quickly estimating epicentral distance and magnitude from a single seismic record. *Bulletin of the Seismological Society of America* 93: 526–532.

Office of Statewide Health Planning and Development (2001) Summary of hopital performance ratings, Series.

Office of Technology Assessment (1995) *Reducting Earthquake Losses*. Washington, DC: U.S. Government Printing Office.

Olsen KB, Day SM, and Minster JB (2006) Strong shaking in Los Angeles expected from southern San Andreas earthquake. *Geophysical Research Letters* 33: L07305.

Olson E and Allen RM (2005) The deterministic nature of earthquake rupture. *Nature* 438: 212–215.

Olson EL and Allen RM (2006) Is earthquake rupture deterministic? (Reply). *Nature* 442: E6.

Olson RS, Podesta B, and Nigg JM (1989) *The Politics of Earthquake Prediction*, 187 pp. Princeton: Princeton University Press.

Otsuka M (1972) A chain-reaction-type source model as a tool to interpret the magnitude-frequency relation of earthquakes. *Journal of Physics of the Earth* 20: 35–45.

Perkins J (2003) San Francisco and the Bay Area earthquake nightmare, Series, Association of Bay Area Governments.

Reid HF (1910) The mechanics of the earthquake, 192 pp. Washington, DC: Carnegie Institution of Washington.

Rikitake T (1976) *Earthquake Prediction*, 357 pp. Amsterdam: Elsevier.

Rikitake T (1982) *Earthquake Forecasting and Warning*, 402 pp. Dordrecht: D. Reidel Publishing.

Rikitake T (1986) *Earthquake Premonitory Phenomena: Database for Earthquake Prediction*, 232 pp. Tokyo: Tokyo University Press.

Rikitake T and Hamada K (2001) Earthquake prediction. In: Meyers RA (ed.) *Encyclopedia of Physical Science and Technology* 3rd edn., pp. 743–760. San Diego: Academic Press.

RMS (1995) (Risk Management Solutions, Inc.) What if the 1906 earthquake strikes again? A San Francisco Bay Area scenario, Series, *RMS Report*.

Romanowicz B (1993) Spatiotemporal patterns in the energy release of great earthquakes. *Science* 260: 1923–1926.

Rowshandel B (2006) Estimation of future earthquake losses in California. In: Loyd, Mattison, and Wilson (eds.) *Earthquakes of the San Francisco Bay Area and Northern California*. Sacremento: California Geological Survey.

Rydelek P and Horiuchi S (2006) Is earthquake rupture deterministic? *Nature* 442: E5–E6.

Shedlock KM and Tanner JG (1999) Seismic hazard map of the western hemisphere. *Annali Di Geofisica* 42: 1199–1214.

Simons FJ, Dando B, and Allen RM (2007) Automatic detection and rapid determination of earthquake magnitude by wavelet multiscale analysis of the primary arrival. *Earth and Planetary Science Letters* 250: 214–223.

Sobolev GA (1995) *Fundamentals of Earthquake Prediction*, 161 pp. Moscow: Electromagnetic Research Center.

Somerville P, Irikura K, Graves RP, Sawada S, and Wald D (1999) Characterizing crustal earthquake slip models for the prediction of strong motion. *Seismological Research Letters* 70: 59–80.

Song S, Beroza G, and Segall P (2006) A unified source model for the 1906 San Francisco Earthquake (abstract). *Seismological Research Letters* 77: 271.

Sorensen SP and Meyer KJ (2003) Effect of the Denali Fault rupture on the Trans-Alaska Pipeline, *Proceedings of the 6th U.S. Conference on Lifeline Earthquake Engineering*, Long Beach, CA, August 2003, 1–9.

Sykes LR, Shaw BE, and Scholz CH (1999) Rethinking earthquake prediction. *Journal of Pure and Applied Geophysics* 155: 207–232.

Tucker BE (2004) Trends in global urban earthquake risk: A call to the international earth science and earthquake engineering communities. *Seismological Research Letters* 75: 695–700.

Turcotte DL (1992) *Fractals and Chaos in Geology and Geophysics*, 221 pp. Cambridge: Cambridge University Press.

Unesco (1984) *Earthquake Prediction*, 995 pp. Tokyo: Terra Scientific Publishing Company.

United Nations (2004) World urbanization prospects. The 2003 Revision, 185 pp. New York: United Nations.

Veneziano D, Cornell CA, and O-Hara T (1984) Historical method of seismic hazard analysis, Series, NP-3438.

Vogel A (1979) *Terrestrial and Space Techniques in Earthquake Prediction Research*, 712 pp. Braunschweig: Friodt. Vieweg & Sohn.

Wald DJ, Earle PS, Lin K-W, Quitoriano V, and Worden B (2006a) Challenges in Rapid Ground Motion Estimation for the Prompt Assessment of Global Urban Earthquakes, *Proceedings of the 2nd International Workshop on Strong Ground Motion Prediction and Earthquake Tectonics in Urban Areas*, *ERI*, Tokyo, BERI 8 pp.

Wald DJ, Quitoriano V, and Dewey JW (2006b) USGS ''Did you feel it?'' community internet intensity maps: Macroseismic data collection via the internet. *ISEE Proceedings*, Geneva.

Wald DJ, Quitoriano V, Heaton TH, and Kanamori H (1999a) Relationships between peak ground acceleration, peak ground velocity, and Modified Mercalli Intensity in California. *Earthquake Spectra* 15: 557–564.

Wald DJ, Quitoriano V, Heaton TH, Kanamori H, Scrivner CW, and Worden CB (1999b) TriNet ''ShakeMaps'': Rapid generation of peak ground motion and intensity maps for earthquakes in southern California. *Earthquake Spectra* 15: 537–555.

Wald DJ, Worden CB, Quitoriano V, and Pankow KL (2005) ShakeMap Manual: Technical Manual, Users Guide, and Software Guide, U.S. Geological Survey, Series, Techniques and Methods 12–A1.

Wang K, Chen Q-F, Sun S, and Wang A (2006) Predicting the 1975 Haicheng earthquake. *Bulletin of the Seismological Society of America* 96: 757–795.

Wessel P and Smith WHF (1995) New version of the Generic Mapping Tools released. *Supplement EOS Trans. Amer. Geophy. Union* 76: 329.

WG02 (2003) Earthquake probabilities in the San Francisco Bay Region: 2002–2031, US Geological Survey, Series, Open File Report 03–214.

Whittaker A, Moehle J, and Higashino M (1998) Evolution of seismic design practice in Japan. *Structural Design of Tall Buildings* 7: 93–111.

Wu Y-M and Kanamori H (2005a) Experiment on an Onsite Early Warning Method for the Taiwan Early Warning System. *Bulletin of the Seismological Society of America* 95: 347–353.

Wu Y-M, Kanamori H, Allen RM, and Hauksson E (in press) Experiment using the tau-c and Pd method for earthquake early warning in southern California. *Geophysical Journal International*.

Wu Y-M and Teng T-I (2002) A virtual subnetwork approach to earthquake early warning. *Bulletin of the Seismological Society of America* 92: 2008–2018.

Wu Y-M, Yen H-Y, Zhao L, Huang B-S, and Liang W-T (2006) Magnitude determination using initial P waves: A single-station approach. *Geophysical Research Letters* 33: L05306.

Wu YM and Kanamori H (2005b) Rapid assessment of damage potential of earthquakes in Taiwan from the beginning of P waves. *Bulletin of the Seismological Society of America* 95: 1181–1185.

Wu YM, Shin TC, and Tsai YB (1998) Quick and reliable determination of magnitude for seismic early warning. *Bulletin of the Seismological Society of America* 88: 1254–1259.

Wu YM and Teng TL (2002) A virtual subnetwork approach to earthquake early warning. *Bulletin of the Seismological Society of America* 92: 2008–2018.

Wu YM and Zhao L (2006) Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters* 33: L16312.

Wurman G, Allen RM, and Lombard P (in review). Toward earthquake early warning in northern California. *Journal of Geophysical Research*.

Wyss M (1979) Earthquake prediction and seismicity patterns (Reprinted from Pure Appl. Geophys., Vol 117, No 6, 1979), 237 pp. Basel: Birkhauser Verlag.

Wyss M (1991) *Evaluation of Proposed Earthquake Precursors*, 94 pp. Washington DC: Am. Geophys. Un.

Wyss M (2004) Human losses expected in Himalayan earthquakes. *Natural Hazards* 34: 305–314.

Yamada M and Heaton T (2006) Early warning systems for large earthquakes: Extending the virtual seismologist to finite ruptures (abstract). *Seismological Research Letters* 77: 313.

Yamakawa K (1998) The Prime Minister and the earthquake: Emergency management leadership of Prime Minister Marayama on the occasion of the Great Hanshin-Awaji earthquake disaster. *Kansai Univ. Rev. Law and Politics* 19: 13–55.

Zhang P, Yang Z-X, Gupta HK, Bhatia SC, and Shedlock KM (1999) Global seismic hazard assessment program (GSHAP) in continental Asia. *Annali Di Geofisica* 42: 1167–1190.

4.22 The Role of Fault Zone Drilling

M. D. Zoback, Stanford University, Stanford, CA, USA

S. Hickman and W. Ellsworth, US Geological Survey, Menlo Park, CA, USA

4.22.1 Introduction – Why Drill to Study Earthquakes?

The objective of fault-zone drilling projects is to directly study the physical and chemical processes that control deformation and earthquake generation within active fault zones. An enormous amount of field, laboratory, and theoretical work has been directed toward the mechanical and hydrological behavior of faults over the past several decades. Nonetheless, it is currently impossible to differentiate between – or even adequately constrain – the numerous conceptual models of active faults proposed over the years. For this reason, the Earth science community is left in the untenable position of having no generally accepted paradigm for the mechanical behavior of faults at depth. One of the primary causes for this dilemma is the difficulty of either directly observing or inferring physical properties and deformation mechanisms along faults at depth, as well as the need to observe directly key parameters such as the state of stress acting on faults at depth, pore fluid pressure (and its possible variation in space and time), and processes associated with earthquake nucleation and rupture. Today, we know very little about the composition of active faults at depth, their constitutive properties, the state of *in situ* stress or pore pressure within fault zones, the origin of fault-zone pore fluids, or the nature and significance of time-dependent fault-zone processes.

Most of what we now know about the structure, composition, and deformation mechanisms of crustal faults has been learned from geological investigations of exhumed faults, particularly where erosion has exposed previously deeply buried fault rocks. These field observations have proved to be particularly useful for several reasons. First, field observations of exhumed faults allow broad coverage with respect to

variations in faulting style (e.g., comparing strike slip, normal and reverse faults), fault movement history, and local geology. Second, where sufficient surface outcrops exist, field observations can readily address issues related to geometrical complexity and spatial heterogeneity in physical properties and fluid composition (e.g., Kerrich *et al*., 1984; Parry, 1994; Evans and Chester, 1995).

Constraints on the mechanical state and physical properties of active fault zones (e.g., fluid pressure, stress and permeability) from surface observations are indirect and subject to alternate interpretations. Stress heterogeneities induced by fault slip can lead to considerable uncertainties in inferring past fluid pressures from observations of vein geometry in outcrop. In all of these investigations, a complex history of uplift and denudation may have severely altered evidence of deformation mechanisms, fault-zone mineralogy, and fluid composition operative during fault slip. This problem is especially acute for solution–transport–deformation mechanisms (e.g., pressure solution and crack healing/sealing) and other low-activation-energy processes, as the deformation microstructures formed at depth are easily overprinted by ongoing deformation as the fault rocks are brought to the surface. Thus, the importance of fluids in earthquake generation and rupture is impossible to assess with any degree of certainty based solely on studies of exhumed fault rocks.

Fault-zone drilling projects allow us to address a number of first-order questions related to fault mechanics:

What are the mineralogy, deformation mechanisms and constitutive properties of fault gouge? Why do some faults creep? What are the strength and frictional properties of recovered fault rocks at realistic *in situ* conditions of stress, fluid pressure, temperature, strain rate, and pore fluid chemistry? What determines the depth of the shallow seismic-to-aseismic transition? What do mineralogical, geochemical, and microstructural analyses reveal about the nature and extent of water–rock interaction?

What is the fluid pressure and permeability within and adjacent to fault zones? Are there superhydrostatic fluid pressures within some fault zones and through what mechanisms are these pressures generated and/or maintained? How does fluid pressure vary during deformation and episodic fault slip (creep and earthquakes)? Do fluid pressure seals exist within or adjacent to fault zones and at what scales?

What are the composition and origin of fault-zone fluids and gasses? Are these fluids of meteoric, metamorphic, or mantle origin (or combinations of the three)? Is fluid chemistry relatively homogeneous, indicating pervasive fluid flow and mixing, or heterogeneous, indicating channelized flow and/or fluid compartmentalization?

How do stress orientations and magnitudes vary across fault zones? Are principal stress directions and magnitudes different within the deforming core of weak fault zones than in the adjacent (stronger) country rock, as predicted by some theoretical models? How does fault strength measured in the near field compare with depth-averaged strengths inferred from heat flow and regional stress directions? What is the nature and origin of stress heterogeneity near active faults?

How do earthquakes nucleate? Does seismic slip begin suddenly or do earthquakes begin slowly with accelerating fault slip? Do the size and duration of this precursory slip episode, if it occurs, scale with the magnitude of the eventual earthquake? Are there other precursors to an impending earthquake, such as changes in pore pressure, fluid flow, crustal strain, or electromagnetic field?

How do earthquake ruptures propagate? Do earthquake ruptures propagate as a uniformly expanding crack or as a 'slip pulse'? What is the effective (dynamic) stress during seismic faulting? How important are processes such as shear heating, transient increases in fluid pressure, and fault-normal opening modes in lowering the dynamic frictional resistance to rupture propagation?

How do earthquake source parameters scale with magnitude and depth? What is the minimum size earthquake that occurs on faults? How is long-term energy release rate partitioned between creep dissipation, seismic radiation, dynamic frictional resistance, and grain size reduction (i.e., by integrating fault-zone monitoring with laboratory observations on core)?

What are the physical properties of fault-zone materials and country rock (seismic velocities, electrical resistivity, density, porosity)? How do physical properties from core samples and downhole measurements compare with properties inferred from surface geophysical observations? What are the dilational, thermoelastic, and fluid-transport properties of fault and country rocks and how might they interact to promote either slip stabilization or transient overpressurization during faulting?

What processes control the localization of slip and strain? Are fault surfaces defined by background microearthquakes and creep the same? Would active slip surfaces be recognizable through core analysis and downhole measurements in the absence of seismicity and/or creep?

4.22.2 Fluids and Faulting

Among the many compelling reasons for drilling into active faults is the possibility to study the role of fluids, fluid pressure, and fluid flow in earthquake processes. A long-standing (and still-growing) body of evidence suggests that fluids are intimately linked to a variety of faulting processes (see review by Hickman *et al.* (1995)). These include the long-term structural and compositional evolution of fault zones; fault creep; and the nucleation, propagation, arrest, and recurrence of earthquake ruptures.

The concept that high fluid pressures and the localization of deformation are often linked is widely accepted in the structural geology literature (e.g., Hubbert and Rubey, 1959; Fyfe *et al.*, 1978), and has been reinforced by studies of active accretionary prisms in subduction complexes and their fossil equivalents (e.g., Dahlen, 1990; Fisher, 1996; Saffer and Bekins, 2006). Better understanding of the role of fluids in faulting in accretionary prisms is one goal behind the plan to drill in the Nankai subduction zone (the NanTroSeize project), as discussed below. A point to note is that there is good evidence that seismic rupturing, in at least some instances (e.g., the Western Taiwan fold and thrust belt and the western margin of the Great Valley adjacent to the San Andreas Fault), is occurring in fluid-overpressured crust (Davis *et al.*, 1983; Sibson, 1990). The Taiwan Chelungpu Fault Drilling Project (TCDP), discussed below, is related to drilling into the fault responsible for the M 7.7 Chi–Chi earthquake of 1999, which occurred along the western margins of the Western Taiwan fold and thrust belt.

4.22.2.1 Sources of Fault-Zone Fluids

Potential sources of fluids in brittle faults and shear zones include metamorphic fluid generated by dehydration of minerals during prograde metamorphism (including shear heating), fluid trapped in pore space as sedimentary formation brines, meteoric water carried downward by circulation, and release of volatiles from molten magma or the upper mantle (e.g., Kerrich *et al.*, 1984; Hacker *et al.*, 1995; Wakita and Sano, 1987; Ko *et al.*, 1997). The high fluid pressures that have been postulated within the San Andreas Fault Zone (being tested as part of the San Andreas Fault Observatory at Depth (SAFOD) drilling project discussed below) might be generated and maintained by continued upwelling of overpressured fluids within the fault zone and leakage of these fluids into the country rock (Rice, 1992). Alternatively, high fluid pressures might result from the sealing of locally derived high-pressure fluids within the fault zone once pressure gradients drop below a critical 'threshold' required to overcome forces between molecular water and mineral surfaces in very small cracks and pores (Byerlee, 1990).

Kennedy *et al.* (1997) argued that elevated $^{3}He/^{4}He$ ratios they observed in springs and wells located along a broad zone encompassing the San Andreas Fault system indicate that significant quantities of mantle-derived fluids are entering the fault zone through the ductile lower crust at near lithostatic pressure. However, without direct sampling of fluids from within the San Andreas Fault Zone at depth it is unclear whether these fluids are ascending through a broad, fractured, and faulted zone associated with the overall plate boundary or are narrowly focused within the (permeable) core of the San Andreas Fault itself, and hence intimately involved in the physics of faulting as envisioned by Rice (1992) and others.

4.22.2.2 Fault-Zone Permeability

The permeability structure of shear zones and brittle faults has recently been the focus of field studies that both confirm and extend observations made years ago by mining geologists. Large faults are not discrete surfaces but rather are a braided array of slip surfaces encased in a highly fractured and often hydrothermally altered transition or 'damage' zone (Smith *et al.*, 1990; Bruhn *et al.*, 1990, 1994; Chester *et al.*, 1993; Schulz and Evans, 2000). Structural and mineralogical textures indicate that episodic fracturing and brecciation are followed by cementation and crack healing, leading to cycles of permeability enhancement and reduction accompanied by episodic fluid flow along faults (e.g., Eichhubl and Boles, 2000).

Theoretical modeling (Sleep and Blanpied, 1992, 1994; Sleep, 1995) showed that the generation of dilatant pores and microcracks during earthquakes in a hydraulically isolated fault zone, followed by creep compaction between earthquakes, might lead to cyclically high fluid pressures along faults. Miller (1996) and Fitzenz and Miller (2003) have used the Sleep and Blanpied (1992) model to numerically simulate temporal variations in fluid pressure within faults and their role in controlling earthquake interactions and periodicity, making assumptions about the rheological and hydrological properties of fault

and country rocks. However, using these models to predict the behavior of real faults in the Earth will require direct *in situ* measurements and sampling in active fault zones at seismogenic depths.

A possibly relevant development from studies of fluid pressure in sedimentary basins has been the revelation from borehole measurements of abrupt transitions, both vertically and laterally, between distinct fluid pressure regimes in some sedimentary basins. These 'fluid pressure compartments' are bounded by seals which in some cases are stratigraphic (e.g., shale horizons) but in others are gouge-rich faults or thin zones of hydrothermal cementation which cut across stratigraphy (Hunt, 1990; Powley, 1990; Dewers and Ortoleva, 1994; Martinsen, 1997). By analogy with these observations, Byerlee (1993) proposed a model in which contiguous vertical and horizontal seals within a fault zone would lead to discrete fluid pressure compartments (i.e., tabular lenses), the rupture of which might be important in earthquake nucleation and propagation (see Lockner and Byerlee (1995)). Although direct evidence for these fault-zone fluid compartments in active fault zones is lacking, negative polarity reflections (bright spots) on seismic reflection images acquired over some accretionary prisms have been interpreted to indicate the existence of high-pressure fluid compartments along the basal decollements (Moore and Vrolijk, 1992; Shipley *et al.*, 1994; Moore *et al.*, 1995; Bangs *et al.*, 2004).

4.22.2.3 Transient Fluid Pressure Effects

A range of physical effects arising from the mechanical response of fluid-saturated crust has been invoked to account for time-dependent phenomena associated with faulting such as slow earthquakes, creep events, earthquake swarms, and aftershock activity and its decay (e.g., Nur and Booker, 1972; Rice and Cleary, 1976; Dreger *et al.*, 2000; Hainzl and Ogata, 2005). Transient changes in fluid pressure and effective stress have also been suggested to play a direct role in rupture propagation and arrest. Shear resistance on the rupture surface may be dramatically lowered by localized increases in fluid pressure from frictional heating or locally elevated as a consequence of pore fluid diffusion and dilatant hardening at fault jogs and other irregularities (Sibson, 1973, 1985; Lachenbruch, 1980; Mase and Smith, 1987; Rudnicki, 1988; Sleep, 1995; Segall and Rice, 1995; Andrews, 2002). Continuous monitoring of fluid pressure within active, seismogenic faults are critical objectives of both the SAFOD and Gulf of Corinth scientific drilling project discussed below.

4.22.2.4 Chemical Effects of Fluids on Fault-Zone Rheology

Over the past several years a number of fault mechanics models have either been developed or refined that incorporate solution transport deformation mechanisms that may weaken and/or destabilize the fault zone. However, complicating this issue enormously is the fact that under only slightly varied environmental and mineralogical conditions similar processes can act to cement the fault zone together, thereby increasing fault strength (see Hickman and Evans (1992)). The experimental and theoretical studies on which these models are based are now focusing on processes that have long been inferred as being important from field observations of natural fault and shear zones, such as pressure solution, fluid-assisted retrograde mineral reactions, crack healing and cementation (e.g., Kerrich *et al.*, 1984; Power and Tullis, 1989; Bruhn *et al.*, 1990; Boullier and Robert, 1992; Chester *et al.*, 1993; Eichhubl and Boles, 2000). These deformation mechanisms are all interrelated, in that they depend upon thermally activated chemical reactions between the rock and pore fluid as well as the rates at which dissolved species are transported through the pore fluid.

Laboratory and theoretical investigations have shown that pressure solution may be important in reducing long-term fault strength and in promoting aseismic slip (i.e., creep) along faults (e.g., Rutter and Mainprice, 1979; Chester and Higgs, 1992; Bos and Spiers, 2001). In contrast, solution transport processes such as crack healing and sealing and cementation may cause the welding together of asperities or fault gouge, leading to time-dependent fault strengthening between earthquakes (e.g., Angevine *et al.*, 1982; Hickman and Evans, 1992; Karner *et al.*, 1997). Laboratory friction experiments conducted under hydrothermal conditions suggest that a change in dominant deformation mechanism with increasing depth from brittle deformation to solution transport creep might control the depth at which the seismic-to-aseismic transition occurs in the crust (Blanpied *et al.*, 1991). Observations of aligned fibrous serpentine in an exhumed branch of the San Andreas Fault system suggest that dissolution–diffusion–crystalization processes may be important in reducing fault strength and promoting aseismic slip (creep) along faults (Andreani *et al.*, 2005).

Hydrothermal mineral reactions can also weaken crustal rocks when the reaction products are weaker than the reactants (see Wintsch *et al.*, 1995). Based upon observations of exhumed shear zones in granite, Janecke and Evans (1988) argued that muscovite formed from the breakdown of feldspar might dramatically lower the ductile shear strength of the granite, even at temperatures well below those necessary for the plastic flow of quartz. At least at shallow depths, fault zones such as the San Andreas are mostly composed of clay- and mica-rich gouge resulting from the hydrolysis of feldspar (e.g., Wu, 1978), suggesting an enhancement of the feldspar breakdown reaction within the fault zone. Reactions in the olivine–talc–serpentine–water system have been demonstrated to dramatically lower the shear strength of ultramafic rocks in laboratory friction experiments (Pinkston *et al.*, 1987)

4.22.3 Frictional Strength of Faults

In addition to studying the role of fluids in faulting, fault-zone drilling provides the opportunity to directly measure the state of stress acting within and adjacent to active faults. As reviewed by Zoback and Healy (1984), Hickman (1991), and Townend and Zoback (2000), *in situ* stress measurements in a variety of faulting regimes, in conjunction with information on the attitude of nearby active faults, indicate fault strengths in intraplate areas that are comparable to those predicted by combining Coulomb faulting theory and laboratory derived coefficients of friction between 0.6 and 1.0 (Byerlee, 1978). This is sometimes known as Byerlee's law (Brace and Kohlstedt, 1980). When pore pressure is near hydrostatic (as has found to be the case at all sites where deep drilling has occurred in crystalline rock (see review in Townend and Zoback (2000)), high stress magnitudes, sometimes referred to as 'hydrostatic' Byerlee's law, are predicted. The high stress at depth consistent with Byerlee's law have been measured at a number of sites, including the Rocky Mountain Arsenal, Denver (Healy *et al.*, 1968); Rangely, Colorado (Raleigh *et al.*, 1972; Zoback and Healy, 1984); the Nevada Test Site (Stock *et al.*, 1985); the Fenton Hill geothermal site, New Mexico (Barton *et al.*, 1988; Fehler, 1989); Moodus, Connecticut (Baumgärtner and Zoback, 1989; Mrotek *et al.*, 1988); Dixie Valley, Nevada (Hickman *et al.*, 1997); the Siljian deep borehole in Sweden (Lund and Zoback, 1999) and to ~8 km depth in the KTB drilling project, Oberfalz, West Germany (Brudy *et al.*, 1997). Thus, Byerlee's law (Brace and Kohlstedt, 1980), which was established on the basis of simple faulting theory and laboratory friction experiments (Byerlee, 1978), appears valid for faults within plate interiors. Studies of lithospheric flexure in response to sediment, volcanic, and internal loads (e.g., McNutt, 1980; McNutt and Menard, 1982; Kirby, 1983) also indicate that differential stresses in much of the Earth's crust are high and can approach the magnitude of stress predicted by Byerlee's law. Hence, such sources of stress are capable of generating crustal stresses large enough to induce faulting.

Figure 1 (from Townend and Zoback (2000)) shows a compilation of stress measurements in relatively deep wells and boreholes in various parts of the world. As shown, the ratio of the maximum and minimum effective stresses corresponds to a crust in frictional failure equilibrium with a coefficient of friction ranging between 0.6 and 1.0. The rate at which stress increases with depth is based on the assumption that the ratio of maximum to minimum effective stress is limited by ratio of shear to normal stress on pre-existing faults well-oriented to slip in the current stress field (see Zoback and Healy (1984)). Pore pressure was observed to be essentially equal to hydrostatic pressure at all the sites where deep drilling into crystalline rocks in the crust has taken place. Because stress magnitudes increase with depth due to increasing overburden stress, mean effective stress (mean stress minus the pore pressure, P_f) at depth scales approximately with depth as shown in the figure and discussed by Townend and Zoback (2000).

There are two implications of the data shown in **Figure 1**. First, Byerlee's law, defined on the basis of hundreds of laboratory experiments, appears to be applicable to faults *in situ*. The consistency of *in situ* friction coefficients and those measured in the lab is a rather amazing result when one considers the huge difference between the size of samples used for friction experiments in the lab and the size of real faults *in situ*, the variability of roughness of the sliding surface, and variability in the rock types encountered in the field and studied in the lab. Second, everywhere that stress magnitudes have been measured at appreciable depth, they indicate that they are controlled by the frictional strength of pre-existing faults in the crust. In other words, the Earth's crust appears to be in a state of frictional failure equilibrium and the law that describes that state is simple Coulomb friction.

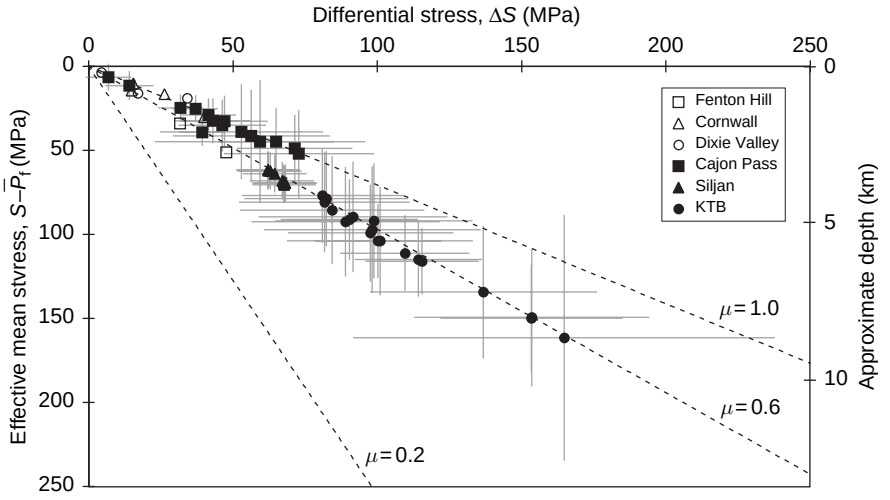


Figure 1 Stress measurements in deep boreholes indicate high crustal strength in accordance with Coulomb faulting theory with laboratory-determined coefficients of friction between 0.6 and 1.0 and hydrostatic pore pressure, which is observed in each borehole (after Townend and Zoback, 2000). It should be noted that the uncertainty estimates in the figure are likely significantly smaller than those shown.

4.22.3.1 Weak Plate-Bounding Faults

A substantial (and growing) body of evidence indicates that slip in crustal earthquakes along major plate-bounding faults (like the San Andreas) occurs at much low levels of shear stress than those shown in **Figure 1**. This hypothesis (sometimes referred to as the San Andreas Stress/Heat Flow Paradox, e.g., Lachenbruch and Sass, 1980, 1992; Zoback *et al.*, 1987; see review by Hickman (1991)). In the context of the San Andreas, there are two principal lines of evidence that indicate that the fault has low frictional strength – the absence of frictionally generated heat and the orientation of the maximum principal stress in the crust adjacent to the fault. A large number of heat flow measurements show no evidence of frictionally generated heat adjacent to the San Andreas Fault (Brune *et al.*, 1969; Brune, 1970; Henyey and Wasserburg, 1971; Lachenbruch and Sass, 1980, 1992; Williams *et al.*, 2004), which implies that shear motion along the fault is resisted by shear stresses approximately a factor of 5 less than those shown in **Figure 1**. Saffer *et al.* (2003) show that it is highly unlikely that topographically driven fluid flow has an appreciable effect on these heat flow measurements, indicating that the lack of frictionally generated heat in the vicinity of the San Andreas Fault is indeed indicative of average shear stress levels acting on the fault at depth.

In addition to the heat flow data, the orientation of principal stresses in the vicinity of the fault also indicates that right-lateral strike slip motion on the fault occurs in response to low levels of shear stress (Zoback *et al.*, 1987; Mount and Suppe, 1987; Oppenheimer *et al.*, 1988; Zoback and Beroza, 1993; Townend and Zoback, 2001, 2004). **Figure 2** (from Townend and Zoback (2004)) shows the direction of relative motion of the Pacific plate with respect to the North American plate indicated by GPS measurements with blue arrows (the map projection is about the pole of relative motion) and the direction of maximum horizontal stress (black, inward pointed arrows). The direction of maximum horizontal compression is clearly at a very high angle to the fault, resulting in low levels of resolved shear stress on the fault.

The significance of regional stress orientation for the state of stress on the fault at depth has been challenged by Scholz (2000) who argues that if stress orientations near the fault rotate to being ~45° from the strike of the fault, both the fault and the crust could have comparably high frictional strength. As summarized by Zoback (2000), there is appreciable evidence that even close to the fault the direction of maximum horizontal compression is at a high angle to the fault (Oppenheimer *et al.*, 1988; Zoback and Beroza, 1993; Townend and Zoback, 2001, 2004). In fact, stress measurements in the SAFOD pilot hole (Hickman and Zoback, 2004) located only 1.8 km from the San Andreas Fault in central California (discussed below) provide still more evidence in

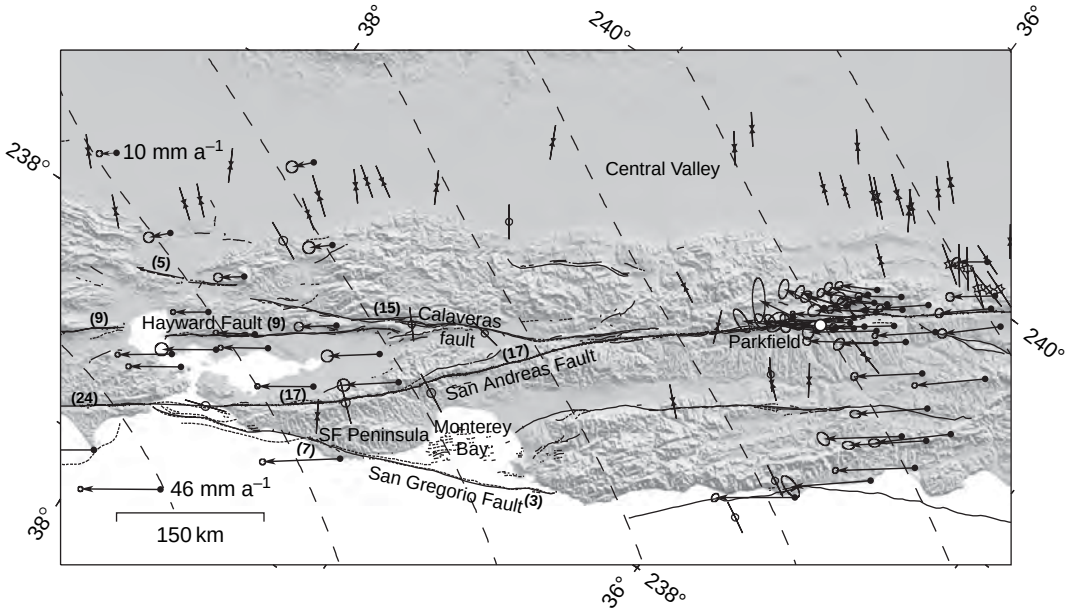


Figure 2 Maximum horizontal compressive stress (S_{Hmax}) and crustal velocity data from central California shown in an oblique Mercator projection about the North America–Pacific Euler pole. S_{Hmax} directions determined from borehole breakouts are shown by inward-pointing arrows, from hydraulic fracturing experiments by short line segments with stars, and from earthquake focal mechanism inversions by short line segments with split circles. The dashed trajectories show regional S_{Hmax} directions calculated using a model of lithospheric buoyancy and plate interaction by Flesch *et al.* (2000). The vectors illustrate crustal velocity data relative to North America (from Murray and Segall, 2001). The numbers in parentheses are Quaternary fault slip rates. The location of the SAFOD drillhole is shown by the open circle near Parkfield. After Townend J and Zoback MD (2004) Regional tectonic stress near the San Andreas fault in Central and Northern California. *Geophysical Research Letters* 31(13): L15S11.

support of the hypothesis that the San Andreas is a weak fault imbedded in a strong crust. In other words, stress magnitudes in the crust adjacent to the fault are high but shear stresses resolved on the fault are low.

Horizontal differential stress magnitudes in the pilot hole were found to be high adjacent to the fault (i.e., consistent with measurements shown in **Figure 1**) while the orientation of maximum horizontal stress at depth in the hole indicates low levels of shear stress resolved onto the San Andreas Fault itself (Hickman and Zoback, 2004). Shear-wave anisotropy measurements in the SAFOD main hole further indicate that the direction of maximum horizontal compression remains nearly perpendicular to the fault to within a few hundred meters of the active fault trace at depth (Boness and Zoback, 2006). In addition, heat flow measurements (to 2 km depth) in the SAFOD pilot hole are consistent with shallower data in the region which show no evidence of frictionally generated heat (Williams *et al.*, 2004).

There is mounting evidence that a number of other plate-boundary faults are similarly weak (see Lachenbruch and Thompson (1972), Kanamori (1980), Wilcock *et al.* (1990), Mount and Suppe (1992), Ben-Avraham and Zoback (1992), Magee and Zoback (1993), and Wang *et al.* (1995), and review by Hickman (1991). The concept that plate-bounding faults like the San Andreas are weak faults imbedded in a strong crust has become widely accepted in recent years and earthquake researchers are now faced with the problem of explaining why major plate boundary faults are substantially weaker than the surrounding crust. In fact, the question of how crustal faults lose their strength is critically important in crustal mechanics and earthquake hazard reduction for a number of reasons:

- As the weakness of plate boundaries (relative to plate interiors) is a fundamental aspect of plate tectonics, how and why plate boundary faults lose their strength is of fundamental global importance for understanding where plate boundaries form, how they evolve with time, and how deformation is partitioned along them.
- The roles of fluid pressure, intrinsic rock friction, chemical reactions, and the physical state of

active fault zones in controlling fault strength must be known in order to simulate earthquakes in the laboratory and on the computer using representative fault-zone properties and physical conditions. This information will also allow for improved models of static stress transfer and earthquake triggering at a regional scale and between specific faults, as needed for intermediate-term seismic hazard forecasting following large earthquakes.

- Through long-term fault-zone monitoring and *in situ* observations of the earthquake source, it will be possible to improve models for earthquake rupture dynamics, including such effects as transient changes in fluid pressure, fault normal opening modes and variations in slip pulse duration. These observations can be used directly in attempts to generate improved predictions of near-field strong ground motion (amplitude, frequency content, and temporal characteristics) and more reliable models for dynamic stress transfer and rupture propagation. The latter processes are believed to control earthquake size (i.e., whether or not a small earthquake will grow into a large one) and, hence, are crucial to long-term probabilistic assessments of earthquake hazard.
- The results of fault-zone drilling experiments are critical to the development of more realistic models for the seismic cycle and assessment of the practicality of short-term earthquake prediction in two ways. During fault-zone monitoring it will be possible to determine if earthquakes are preceded by accelerating fault slip (e.g., a nucleation phase) and/or transient changes in fluid pressure. Second, we will be able to determine whether or not factors which might dramatically lower fault strength (high pore pressure and/or chemical fluid–rock interactions, for example) are closely related to the processes controlling earthquake nucleation. Our current knowledge of fault-zone processes is so poor that not only are we unable to make reliable short-term earthquake predictions, we cannot scientifically assess whether or not such predictions are even possible.

In summary, while essentially all available data indicates that the frictional strength of intraplate crust is high, the frictional strength of plate-bounding faults is anomalously low. Taken together, the heat-flow data and the directional constraint (i.e., S_{Hmax} at 65–85° to the San Andreas Fault) suggest that the San Andreas Fault is weak in both an absolute and relative sense. In other words, the average shear stress required to cause faulting is low (comparable to earthquake stress drops), whereas the frictional strength of faults in the adjacent crust is much higher, consistent with Byerlee's law. Despite the fundamental nature of this finding, we have no direct *in situ* evidence indicating why this might be so, whether the mechanisms responsible for low strength along the San Andreas are likely to be found in other major fault systems or what role that these mechanisms might play in the processes of earthquake nucleation and propagation.

Numerous theories have been proposed over the past decade that are related to the weakness of the San Andreas and other plate-bounding faults. Knowledge of the *in situ* frictional properties of the San Andreas and other major active faults is not only of considerable scientific interest but is also critical for assessing the nature and potential magnitude of static stress transfer and earthquake triggering following large-to-intermediate size earthquakes.

Although the causes for the weakness of these faults are unknown, four general classes of explanations have been suggested:

- elevated fluid pressure (as discussed above);
- intrinsically low coefficients of friction of fault-zone materials;
- solution–transport reactions and related low stress deformation mechanisms (also discussed above); and
- dynamic weakening mechanisms such as shear heating/thermal pressurization or acoustic fluidization.

If the coefficient of friction, μ, is equal to 0.6–0.9 on the San Andreas Fault, as predicted by Byerlee's law, then the heat-flow constraint could be satisfied if the *in situ* pore pressure, P_p, is greater than twice hydrostatic (Lachenbruch and Sass, 1980, 1992). However, if one assumes that principal stress magnitudes are constant across the fault zone and that $\mu \geq 0.6$, then high fluid pressures alone cannot explain the directional constraint as P_p would exceed the least principal stress once the angle between S_{Hmax} and the fault exceeds about 60° (e.g., Zoback *et al.*, 1987; Scholz, 1989; Lachenbruch and McGarr, 1990). It has been suggested that large-scale yielding could lead to an increase in the magnitudes of the principal stresses within the fault zone relative to their values immediately outside of the fault (Rice, 1992). If so, this would allow P_p within the fault zone to exceed significantly the external magnitude of the least principal stress (Byerlee, 1990; Rice, 1992). In this manner, permanently high pore pressures within an

intrinsically strong (i.e., high coefficient of friction) San Andreas Fault Zone, in conjunction with much lower fluid pressures in the surrounding rock, could lower the fault strength sufficiently to satisfy both the heat-flow and directional constraints.

A model that is, in many respects, similar to that of Rice (1992) was proposed earlier by Byrne and Fisher (1990) to explain the apparent weakness of the basal décollement beneath the Kodiak accretionary prism in southwest Alaska. Magee and Zoback (1993) applied the Rice model to explain the low frictional strength of the subduction zone associated with the $M \sim 8.2$ Tokachi-oki earthquake off northern Honshu, Japan. Although not requiring localized increases in stress magnitude, Fournier (1996) has presented a model in which near-lithostatic fluid pressures may be maintained within the San Andreas Fault Zone at depths greater than about 6–10 km if the maximum differential stress at these depths is quite low (~10–30 MPa) and the tensile strength of the rock outside the fault zone remains high due to pervasive crack healing at elevated temperatures.

Alternatively, if one assumes that the fault is optimally oriented with respect to the principal stresses and that fluid pressures are hydrostatic, then the heat-flow constraint can be satisfied if the friction coefficient along the fault is less than about 0.2 (Lachenbruch and Sass, 1992). Similarly, at least in central California where S_{Hmax} is at about 75–85° to the San Andreas Fault, the heat-flow and directional constraints can be simultaneously satisfied under conditions of uniformly hydrostatic fluid pressures if the friction coefficient is extremely low – about 0.1 or less – along the fault and Byerlee's law is applicable outside the fault zone (Lachenbruch and McGarr, 1990; Lachenbruch and Sass, 1992). It is often proposed that the presence of clays or other weak minerals along the San Andreas and other faults might lead to anomalously low frictional resistance (e.g., Wu 1978; Janecke and Evans, 1988; Wintsch *et al.*, 1995). This inference has been supported by laboratory sliding experiments on synthetic clay-rich fault gouges (e.g., Wang *et al.*, 1980; Shimamoto and Logan, 1981; Bird, 1984; Logan and Rauenzahn, 1987), on synthetic serpentinite gouges (Reinen *et al.*, 1994; Reinen and Tullis, 1995) and on synthetic laumontite gouge (Hacker *et al.*, 1995). However, these experiments are all at low-to-moderate confining pressures and temperatures. In contrast, experiments on natural clay-rich fault gouges collected from the San Andreas at depths of less than 0.4 km (Morrow *et al.*, 1982), on synthetic clay-rich fault gouges (Morrow *et al.*, 1992) and on synthetic serpentinite gouges (Moore *et al.*, 1997) at high temperatures and/or confining pressures and hydrostatic fluid pressures indicate coefficients of friction at *in situ* conditions that are too high to be reconciled with either the heat-flow or directional constraints. In addition, both natural and synthetic fault gouges deformed in the laboratory generally fail to exhibit the slip-weakening or velocity-weakening behavior required for the generation of earthquakes (e.g., Byerlee and Summers, 1976; Logan and Rauenzahn, 1987; Marone *et al.*, 1990; Morrow *et al.*, 1992; Reinen *et al.*, 1994). Thus, the importance of these materials in the rheology of the San Andreas Fault at seismogenic depths is unclear.

Another important class of models that might explain the low long-term strength of seismically active segments of the San Andreas and other major faults have called upon processes directly associated with earthquake rupture propagation. These dynamic weakening mechanisms include shear heating during slip, leading to transient high fluid pressures (Lachenbruch, 1980; Andrews, 2002) or melting (Sibson 1973; Spray, 1987; Rice, 2006); reductions in normal stress accompanying the propagation of dilational waves along the fault (Brune *et al.*, 1993; Andrews and Ben-Zion, 1997); the fluidization of fault-zone materials due to the channeling of coseismic acoustic energy (Melosh, 1979, 1996); and hydrodynamic lubrication (Brodsky and Kanamori, 2001). While such processes may be operative during an earthquake, they do not relate to the problem of how earthquakes initiate and each requires that very specific fault-zone conditions exist to be viable. Thus, to assess the likelihood that dynamic weakening mechanisms might operate along active faults we need to compare the structure and physical properties of actual fault zones with parameters required by these various models and then combine these observations with the results of long-term fault-zone monitoring in the near field of small-to-moderate size earthquakes.

4.22.4 Near-Field Observations of Earthquake Nucleation and Propagation

Understanding the physical processes operating during both nucleation and rupture propagation can prove to be critical to understanding why plate

boundary faults tend to be anomolously weak, especially if dynamic weakening mechanisms are important. In addition, by drilling into active faults at seismogenic depths and observing repeating earthquakes at very short distances, one can observe near-field phenomena for earthquakes of $M \sim 1$ or larger, thereby providing a new window into the physics of the earthquake source. Ideally, in addition to instrumenting fault-zone drill holes with seismometers, it would be ideal to place instruments within or immediately adjacent to active sliding surfaces to directly monitor fault displacement, deformation, pore pressure, and heat generated during sliding. The work of Brune and co-workers on foam rubber models of earthquakes (Brune *et al.*, 1993; Anooshehpoor and Brune, 1994) illustrates the advantages of making measurements at or very near the sliding surface. Many of the objectives for near-field observation would be met by sensors placed within a few hundred meters of the earthquake source. At these distances, near-field waves will be of significant amplitude compared to the far-field waves for $M \geq 0$ events, and static strains will be well within the resolution of borehole strainmeters.

The process by which the fault becomes unstable and initiates a dynamically propagating rupture is central to understanding how earthquakes work. It has recently been proposed that the very beginnings of rupture for earthquakes in the magnitude range from at least $M_w = 1$–8 characteristically involve a period of slow growth of the seismic moment (Iio, 1992, 1995; Ellsworth and Beroza, 1995, 1998; Beroza and Ellsworth, 1996). The characteristics of this process, called the seismic nucleation phase, rule out self-similar models for the nucleation and growth of rupture including the standard model of a dynamically growing crack (Kostrov, 1964). Although a range of hypotheses have been proposed to explain this slow beginning to earthquakes, far-field observations have thus far proved to be inadequate to determine if the seismic nucleation phase represents a cascade of smaller events, in which case the dynamically expanding crack model might apply, or if it represents a transition from an aseismic (stable) sliding to dynamic rupture, as required by laboratory-based and theoretical models of rupture initiation (Dieterich, 1992; Ohnaka, 1992). Observations of the nucleation process made within the near field in drill holes passing close to or through active faults at seismogenic depths have the potential to resolve this process, as they will not be distorted by attenuation or scattering, which limits the interpretation of available data (Iio, 1995).

The physics of earthquake rupture propagation has also been the subject of intensive investigation in recent years (e.g., Heaton, 1990; Brune *et al.*, 1993; Melosh, 1996; Peyrat *et al.*, 2004; Harris, 2004). New data have again drawn into question the standard model of a dynamically expanding crack that heals inward from its outer boundary (Madariaga, 1976). There is now evidence from large earthquakes that the rupture may propagate as a 'slip pulse' (e.g., Wald and Heaton, 1994), yet we know little about how such a concentrated slip zone is generated or maintained, or why the fault comes to rest so abruptly. Brune *et al.* (1993) have further proposed that tensile opening of the fault accompanies the shear displacement in the slip pulse. If correct, it would be a mechanism by which a fault can have high static strength, but slide without generating heat. However, sliding at near-zero normal stress implies that the dynamic stress drop should equal the tectonic stress (see Lachenbruch and Sass (1980)) resulting in near-zero shear stress on the fault after rupture (e.g., Zoback and Beroza (1993) for the Loma Prieta earthquake). Thus, measuring the dynamic stress drop in the near-field region will give us a direct test of the high static strength/low dynamic friction hypothesis.

Recent observations of microearthquakes in moderately deep boreholes (2–2.5 km) at Cajon Pass, California (Abercrombie and Leary, 1993; Abercrombie, 1995), the SAFOD pilot hole (Imanishi *et al.*, 2004), and in Long Valley, California (Prejean and Ellsworth, 2001) demonstrate that ultra-high-fidelity recordings can be made in downhole observatories. Imanishi and Ellsworth (in press) recently used high-frequency recordings from the SAFOD Pilot Hole array to study the scaling of apparent stress (seismic wave energy/seismic moment). They found that this ratio shares the same upper bound over the magnitude range from $M9$ to $M0$, providing strong support for the self-similar model of the earthquake source.

4.22.5 Fault-Zone Drilling Projects

In this section we review several recent, on-going, and planned scientific drilling projects related to earthquake studies. Many of the key questions to be addressed by deep drilling into active faults were enumerated in December 1992 at a workshop (attended by 113 scientists and engineers from seven countries) on San Andreas Fault-Zone drilling at the Asilomar Conference Center in Pacific Grove,

California. While the purpose of this workshop was to initiate a broad-based scientific discussion of the issues that could be addressed by drilling and direct experimentation in the San Andreas Fault, the arguments put forth for scientific drilling into fault zones are generally applicable to all of the fault-zone drilling projects discussed in this section.

The questions to be addressed by fault-zone drilling projects are far reaching, and include the following:

4.22.5.1 Fault Behavior

- What is the static strength of active faults? Why are some faults anomalously weak?
- Why are some faults (or fault segments) creeping and some locked?
- What factors control the localization of slip and strain?
- How are faults stressed at different crustal levels?
- How does strain communication occur within fault zones over different time scales?
- How is energy partitioned within fault zones between seismic radiation, frictional dissipation, grain-size reduction, and chemical reactions?
- Can the frequency–magnitude relationship for earthquakes be extrapolated to smaller magnitudes?

4.22.5.2 Fluid Pressure

- What is the vertical and lateral distribution of fluid pressure regimes in faults?
- Do fault fluid pressure compartments exist?
- If so, what is the nature of the seals between these compartments?
- What is the time dependence of fluid pressure within fault zones?
- What is extent of vertical and lateral fluid migration during a seismic stress cycle?

4.22.5.3 Fault Fluids

- What is the origin and composition of fault zone fluids?
- What are the permeabilities of fault-zone materials and country rock?
- What are the fluid transport mechanisms in and adjacent to fault zones and what physical processes lead to fluid redistribution?
- What is the interplay between water–rock interaction and rheology at different structural levels?

4.22.5.4 Fault-Zone Properties and Physical Parameters

- How does the stress tensor vary in the vicinity of fault zones?
- How do pre- and postfailure stress states compare?
- What, if any, form of cyclical dilatancy operates in the vicinity of fault zones?
- How do physical properties relate to the fault zone fabric?
- What is the origin of low-velocity zones associated with fault zones?
- How well and in what manner do physical properties and heterogeneity measured in boreholes correlate with geophysical observables?

4.22.5.5 Fault Structure and Materials

- How does the width and character of active slip zones vary with depth?
- What is the thermal structure of active fault zones?
- How do mineralogy and deformation mechanisms within fault zones change with depth, temperature, and country-rock geology?
- What determines the maximum depth of seismic activity?
- At what temperature do mineral reaction kinetics operate at the timescale of an earthquake cycle?
- How accurate are inferences drawn from deformation microstructures, piezometers, and fluid inclusions and how might one assess their survivability?

Fundamental questions about faulting and earthquakes such as these have gone unanswered due to the complete lack of data on the physical and chemical processes operating on faults at depth. Hence, the principal reasons for drilling into active faults are to conduct extensive investigations *in situ* and on exhumed materials that are representative of the faults at the pressures, temperatures, and conditions at which major earthquakes nucleate.

4.22.5.6 San Andreas Fault Observatory at Depth

The site identified for this 3.0 km deep drilling project is located in central California (**Figure 2**), about 10 km northwest of the town of Parkfield. This site is located at the northwestern end of the rupture zone of the 1966 and 2004 $M = 6$ Parkfield earthquakes, in the transition between the creeping and locked sections of the San

Andreas Fault. The San Andreas displays a range of behaviors at this site. At the surface, the fault is creeping at a rate of 1.8 cm yr^{-1}, with most of the fault displacement localized to a zone no more than 10 m wide (Burford and Harsh, 1980; Schulz, 1989). Numerous earthquakes occur directly on the San Andreas Fault in the depth interval from about 3 to 12 km. The shallow seismicity at Parkfield occurs in tight clusters of activity (Nadeau *et al.*, 1994; 1995, 2004) that have remained spatially stationary for at least the past 20 years.

An important feature of the microearthquakes beneath Middle Mountain is that they occur in families of repeating events which provide a reliable target for guiding drilling (**Figure 3**). Individual earthquakes have been observed to recur numerous times using the U.C. Berkeley High Resolution Seismic Network (HRSN), at precisely the same location and with the same magnitude (Nadeau *et al.*, 1994, 1995, 2004; Nadeau and McEvilly, 1997). Repeating sources of up to $M = 2$ are located at drillable depths beneath the proposed drill site. A major goal of this experiment is to drill as close as possible to one or more of these sources and to follow the buildup of strain and its release through multiple earthquake cycles during the monitoring phase of the experiment.

Illustrated in **Figure 3**, a 2 km deep pilot hole drilled at the SAFOD site in 2002, 1.8 km to the southwest of the surface trace of the San Andreas Fault (see review by Hickman *et al.* (2004) and associated papers published in special issues of Geophysical Research Letters). During the summers of 2004 and 2005, the SAFOD mainhole was rotary drilled through the zone of the repeating microearthquakes. Note that in the subsurface, the position of the fault is to the southwest of the surface trace.

Four major geologic units were encountered along the trajectory of the SAFOD main hole: In the vertical section of the wellbore, the near-surface Quaternary and Tertiary sediments were found to be underlain by Salinian granite at a depth of ~700 m. After deviating the borehole toward the fault, arkosic sediments (most likely locally derived from Salinian granite) were encountered about 300 m NE of the drill-site, perhaps after crossing the Buzzard Canyon fault, a NW trending strike-slip fault exposed at the surface that trends subparallel to the San Andreas (M. Rymer, personal communications). Approximately, 1200 m NE of the drillsite, a possibly ancestral trace of the San Andreas was crossed as the lithology changed abruptly to claystones and siltstones of the Great Valley formation, found throughout central California on the east side of the San Andreas.

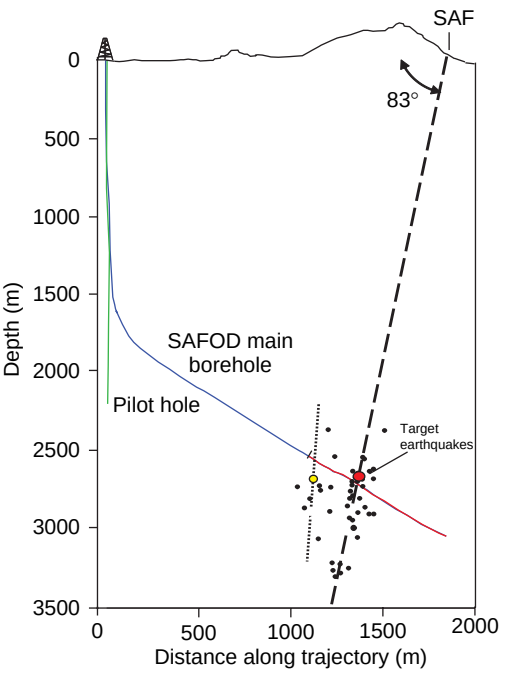


Figure 3 Schematic diagram of the SAFOD pilot hole and main hole. The precise location of the target earthquakes is approximate although the relative locations are exact. As indicated by the distribution of microearthquakes, slip is accommodated by multiple faults at depth. Simply connecting the position of the San Andreas Fault at depth with the surface position implies that the fault dips steeply to the west.

Geophysical logs and cuttings analysis indicate that the San Andreas is a broad zone of anomalously low P- and S-wave velocity and resistivity that define a relatively broad damage zone as indicated in **Figure 4**. The locations of active fault traces are revealed by casing deformation (indicated by the red line in **Figure 4**) and the location of a $M \sim 0$ microearthquake that occurred in May 2005 (indicated in **Figure 4**). Note that these active faults are associated with narrower, more highly localized zones of low P- and S-wave velocity and resistivity embedded within the broader damage zone. In a third phase of the project, coring will be done in multilateral holes to directly sample the damage zone and both creeping/seismically active faults at depth. Preliminary results from the project were presented in several special sessions of the December 2005 American Geophysical Union meeting (published in EOS, v. 86, 2005), and numerous publications are in print or under review detailing early results from SAFOD.

An array of downhole seismometers, accelerometers, and other sensors will be deployed near

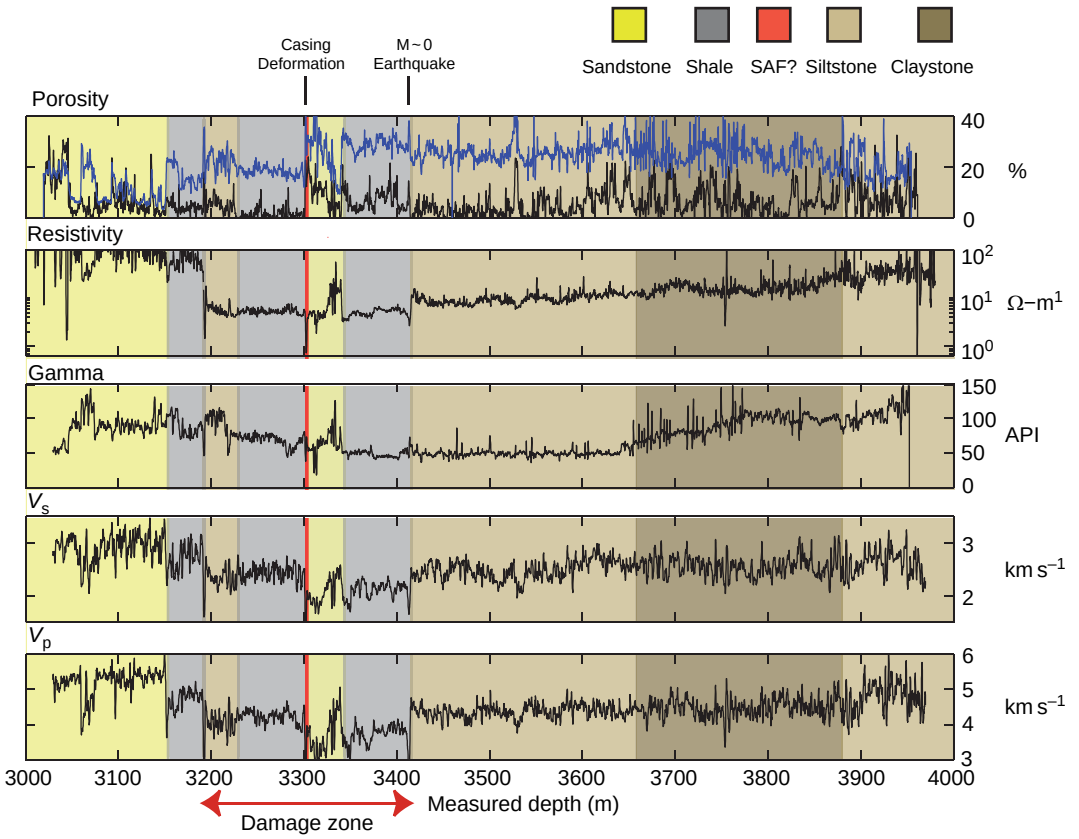


Figure 4 Geophysical logs and generalized lithology from phase 2 of the SAFOD project (shown in red in **Figure 3**). Anomalously low V_p, V_s, and resistivity indicate a broad low damage zone ~250 m wide. Multiple active fault traces appear to occur within the damage zone. The red line at 3300 m depth indicates an area of localized casing deformation. A $M \sim 0$ earthquake was recorded by a seismic array in the borehole appears to be associated with the localized low velocity zone at 3410 m.

the fault zone after drilling. This will allow for accurate determinations of the radiated energy, seismic moment, and earthquake locations can be made, the detailed velocity structure of the fault zone can be investigated, and earthquake nucleation and propagation can be studied with unprecedented detail. High-frequency signals are expected from the very local events, with good signal-to-noise above 1 kHz. Thus, good sensitivity to the higher frequencies becomes the most important seismic monitoring design criteria. Peak acceleration estimates for a magnitude 1.0 earthquake located at a distance of 1000 and 100 m are 1*g* and 10*g*, respectively.

4.22.5.7 Nojiima Fault

In 1995, a M 6.9 strike-slip faulting earthquake struck Kobe, Japan, resulting in 6432 fatalities. The earthquake occurred on the Nojiima Fault (**Figure 5**), a right-lateral strike-slip fault that trends NE–SW through Awaji Island and beneath Osaka bay (Ando, 2001). Several research groups drilled boreholes (ranging in depth from 747–1800 m) through the fault along segments of the Nojima Fault that exhibited surface rupture during the earthquake. Note the numerous shallow (2–5 km depth) occurred on the Nojiima Fault beneath Awaji Island. Extensive geophysical logging, detailed core analysis, stress measurements, and repeated hydrological tests and repeated injection tests were carried out. The preliminary results were published in the special volume of the Island Arc (Oshiman *et al.*, 2001).

Two boreholes were drilled at Hirabayashi (where the maximum slip was 2 m) that penetrated the core of the Nojima Fault. The structure of the Nojima Fault Zone is characterized by a narrow fault core with various types of fault gouge. The hanging wall of the fault displays many minor shear zones that increase in frequency toward the fault core (**Figure 6**). The footwall of the fault is significantly

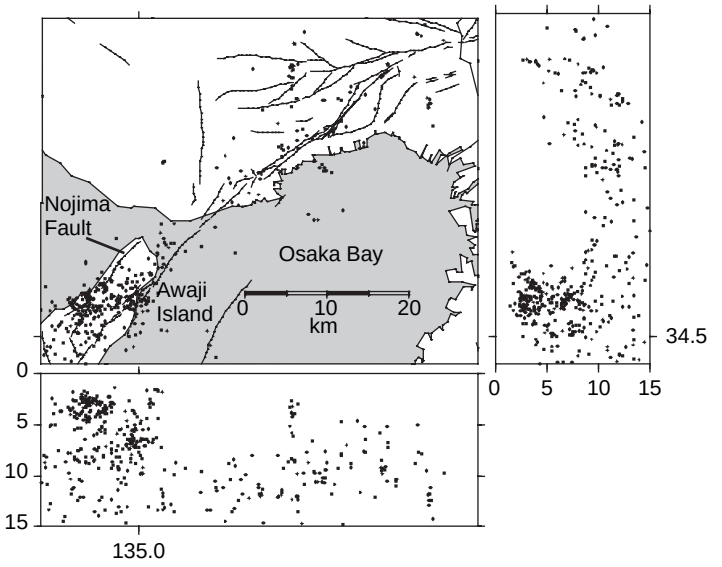


Figure 5 The Nojima Fault – source of the *M* 6.9 1995 Kobe earthquake – is an NE trending strike-slip fault exposed on Awaji Island and extends beneath Osaka Bay. Numerous shallow aftershocks (black dots) occurred on the section of the fault where drilling took place. East–west and north–south cross-sections show the distribution of aftershocks with depth. As shown, many of the aftershocks beneath Awaji Island are extremely shallow (2–5 km deep), with others extending to typical depths of 10–15 km. Courtesy of H. Ito.

less deformed and less altered with respect to the hanging wall. The physical properties of fault zone and country rocks have been extensively studied (Tanaka *et al.*, 2001; Ohtani *et al.*, 2001; Fujimoto *et al.*, 2001; Boullier *et al.*, 2001).

Integrated borehole monitoring systems (three-component strain meter, three-component seismometer, and water-level sensors) have been installed along two different segments of the Nojima Fault.

4.22.5.8 Chelungpu Fault

The 1999 *M* 7.7 Chi–Chi earthquake in Taiwan produced large zones of surface rupture, with a maximum displacement of 8 m on the Chelungpu Fault. Along nearly all of its length, the earthquake fault strikes ~N–S (**Figure 7(a)**) with the rocks to the east thrust over those to the west along an east-dipping thrust faults (**Figure 7(b)**). The northern portion of the fault is characterized by very large slip but relatively low ground acceleration during the *M* 7.7 earthquake. The Taiwan Chelungpu Fault Drilling Project (TCDP) was carried out near the town of Dakeng in a zone of large surface offset (and subsurface slip) near the northern end of the N–S trending rupture surface.

The main scientific objectives for drilling were related to testing various mechanisms of nucleation and rupture of large earthquakes and provide answers to fundamental questions pertaining to the relationship between the Chelungpu thrust and regional tectonics (Tanaka *et al.*, 2002; Ma *et al.*, 2003). TCDP drilling was started in 2004 and two boreholes were completed in 2005, including successful continuous coring, logging, and borehole monitoring. The direction of maximum horizontal compression found in one of the TCDP boreholes is ~N 120° E, consistent with the regional stress field (Kao and Angelier, 2001) and relative motion of the Phillipine Sea Plate with respect to Taiwan (Wu *et al.*, 2007).

As shown in **Figure 8**, two holes (A and B) were drilled through the main fault zone that ruptured during the Chi–Chi earthquake (**Figure 8**). In hole A, multiple fault zones were identified in the Pliocene Chinshui Shale and Miocene Kueichulin Formation, but the primary slip zone in hole A was encountered at a depth of 1111 m (see photo in **Figure 8**) and 1136 m in hole B (see photo in **Figure 8**). This fault is associated with bedding-parallel thrusting with a gentle dip of about 20° and is characterized by over 1 m thickness of gouge (fault core) and gradational breccia in a damage zone found in both the upper and lower blocks. Wu *et al.* (2007) noted a major perturbation of the stress field in hole A in the vicinity of the Chelungpu Fault.

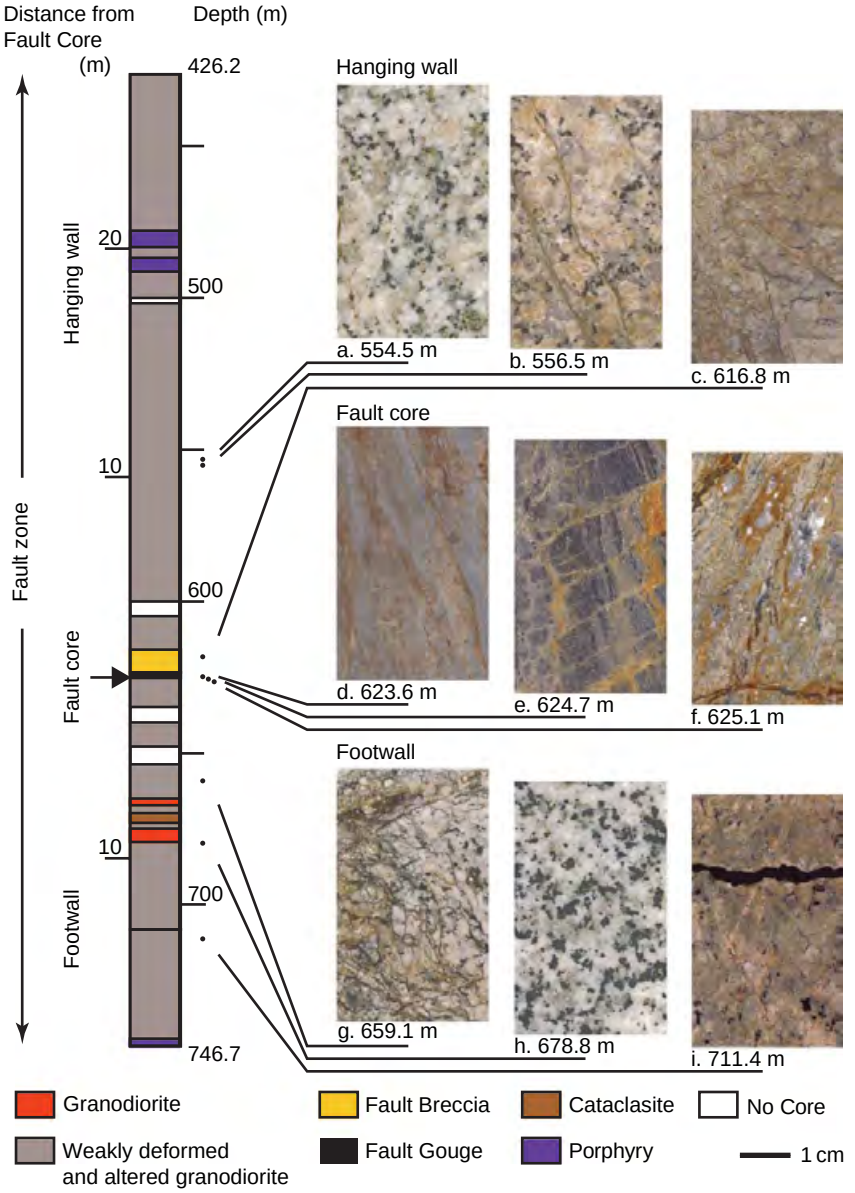


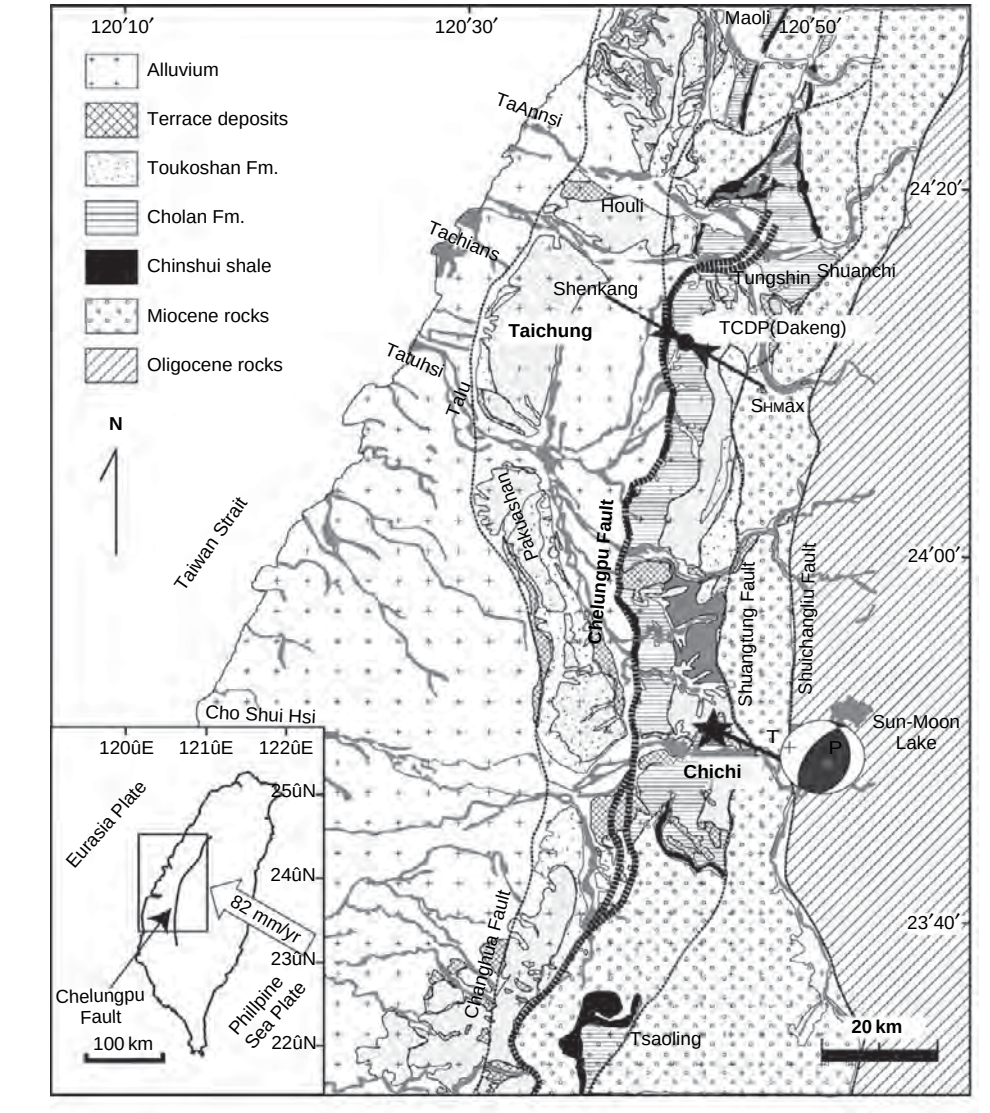
Figure 6 Lithologies encountered in the Nojima Fault and photographs of rocks from the hanging wall, fault core, and footwall. After Ando M (2001) Geological and geophysical studies of the Nojima Fault from drilling: An outline of the Nojima Fault Zone Probe. *The Island Arc* 10: 206–214.

To address key questions of earthquake rupture dynamics such as energy dissipation and fracture energy, Ma *et al.* (2006) analyzed the grain size distribution of the fault gouge. They estimated that 6% of the total earthquake energy is associated with gouge formation. Tanaka *et al.* (2007) and Kano *et al.* (2006) discuss the possibility that a thermal anomaly measured at the trace of the fault is associated with frictional heat generated during faulting.

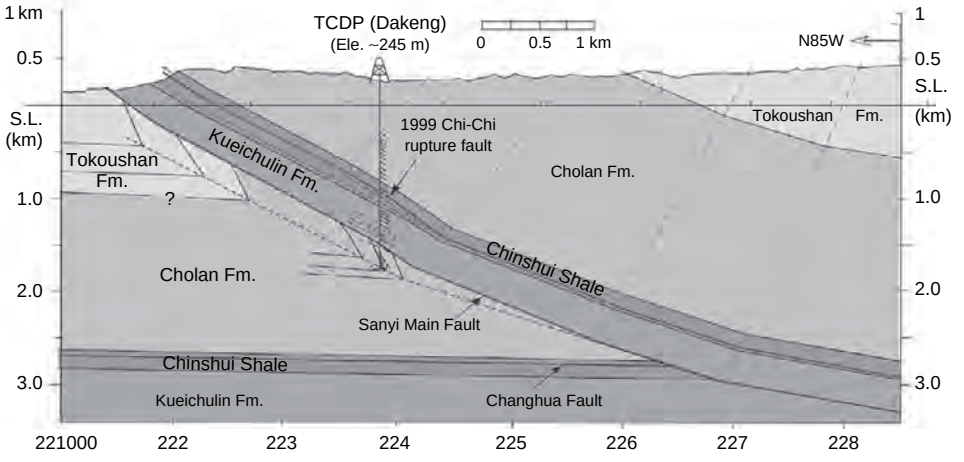
4.22.5.9 Gulf of Corinth

The Gulf of Corinth, Greece, is the most actively deforming region in Europe and one of the most seismically active areas. Deformation in the area is dominated by north–south directed back-arc extension associated with the Hellenic trench. The Gulf of Corinth is an east–west trending graben bordered on the north and south by east–west trending normal faults (**Figure 9**). The extension rate in the region

(a)


120′10′
120′30′
Maoli
120′50′
Alluvium
Terrace deposits
Toukoshan Fm.
Cholan Fm.
Chinshui shale
Miocene rocks
Oligocene rocks
N
TaAnnsi
Houli
24′20′
Tachians
Tungshin
Shuanchi
Shenkang
Taichung
TCDP(Dakeng)
Tatuhsi
Tadu
SHMAX
Taiwan Strait
Paknashan
Chelungpu Fault
24′00′
Shuangtung Fault
Shuichangliu Fault
Cho Shui Hsi
120ûE
121ûE
122ûE
25ûN
24ûN
23ûN
22ûN
Eurasia Plate
82 mm/yr
Chelungpu Fault
100 km
Philippine Sea Plate
T
P
Sun-Moon Lake
Chichi
Changhua Fault
23′40′
Tsaoling
20 km

(b)


1 km
0.5
S.L.
(km)
1.0
2.0
3.0
TCDP (Dakeng)
(Ele. ~245 m)
0
0.5
1 km
N85W
1
0.5
S.L.
(km)
Tokoushan Fm.
1999 Chi-Chi rupture fault
Kueichulin Fm.
Tokoushan Fm.
?
Cholan Fm.
Chinshui Shale
Cholan Fm.
Sanyi Main Fault
Chinshui Shale
Kueichulin Fm.
Changhua Fault
1.0
2.0
3.0
221000
222
223
224
225
226
227
228

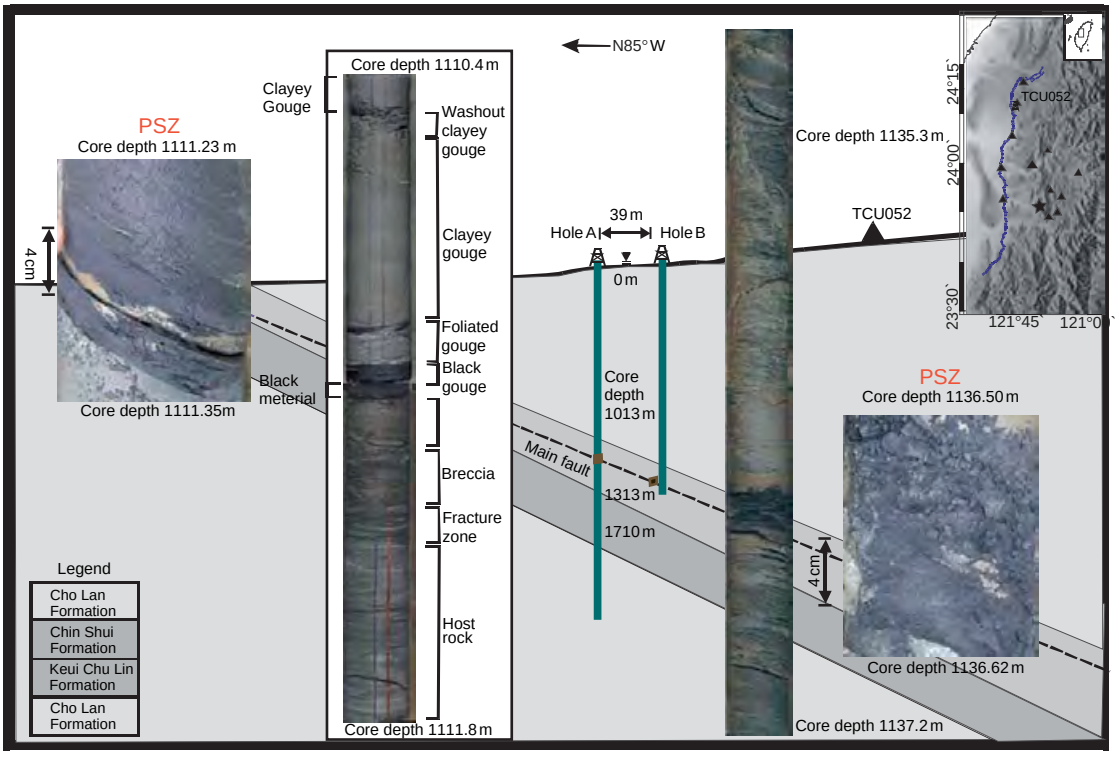


Figure 8 Detailed ithologies and fault rocks encountered in the two TCDP core holes shown on a background of the generalized geology as shown in **Figure 7(b)**. The primary slip zone in hole A was encountered at a depth of 1112.23–1111.35 m (see photo) and 1136.50–1136.62 m in hole B (see photo). Courtesy of K.-F. Ma.

ranges between 1 and 1.5 cm yr^{-1}. A number of large earthquakes have occurred in the region in the past century.

The Corinth Rift Laboratory (CRL) project involved drilling into the Aigion Fault on the south side of the Gulf. This site offers excellent conditions for *in situ* investigation of active normal faulting, the rifting process and for monitoring fluid–fault interactions (Cornet *et al.*, 2004).

The AIG10 borehole was drilled in 2002 to 1000 m depth and intersected the Aigion Fault at 760 m (**Figure 10**). A key goal of the laboratory is to understand the relationships between outcrops of steeply dipping fault zones and the seismogenic faults at great depth. Special attention is directed to the interactions between circulating fluids and fault mechanics, including hydro-thermo-mechanical coupling and the role of geochemical healing and alteration. After passing through an overburden of sands, conglomerates, and clays, the borehole penetrated alternating Mesozoic radiolarite and limestone in the hanging wall before intersecting the main fault zone at a depth of 760 m, where cataclastic fault rocks were encountered. Below the fault, a homogeneous, heavily karstified limestone was encountered down to 1000 m.

The Aigion Fault Zone forms a hydraulic barrier that sustains about a 0.5 MPa differential pressure across it. There is a slight overpressure (~0.5 MPa) in the hanging wall immediately above the fault with a larger overpressure (0.9 ± 0.1 MPa) in the foot-wall resulting from the topographic relief from the footwall to the hanging wall sides of the fault. Nearly

Figure 7 (a) Generalized map of surface geology in western Taiwan (H. Tanaka, personal communication, 2006), the trace of the Chelungpu Fault that slipped in the *M* 7.7, 1999 Chi–Chi earthquake epicenter and the location of the TCDP boreholes near Dakeng are shown. The average ~N120° E direction of maximum horizontal compression observed in the borehole (Wu *et al.*, 2007) is the same as the regional stress field (Kao and Angelier, 2001) and the motion direction of the Phillipine Sea Plate relative to Taiwan (see inset). (b) Geologic cross-section derived from high-resolution seismic reflection data showing the location of the TCDP borehole with respect to the east dipping Chelungpu Fault. After Yue LF, Suppe J and Hung JH (2005) Structural geology of a classic thrust belt earthquake: The 1999 Chi-Chi earthquake Taiwan (M 7.6). *Journal of Structural Geology* 27: 2058–2083.

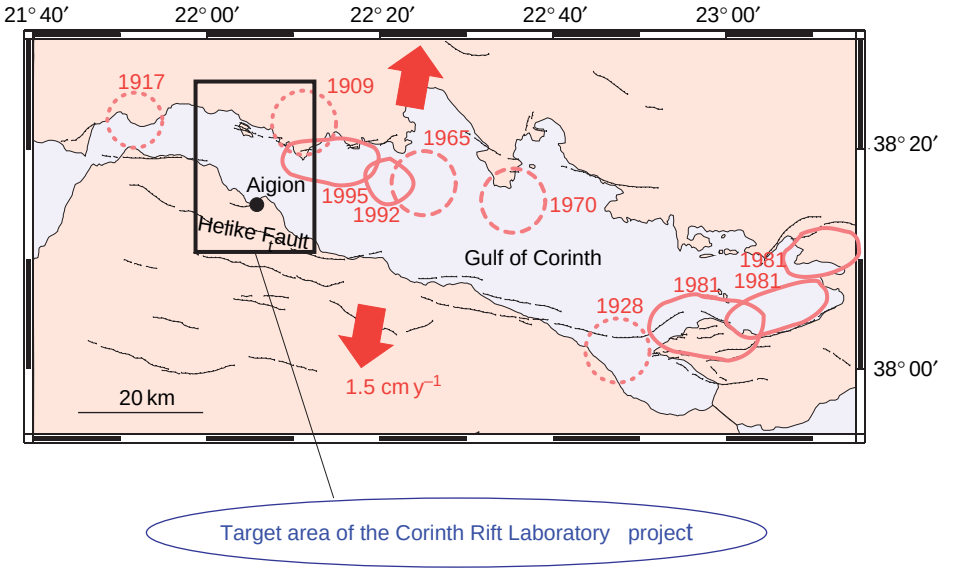


Figure 9 Active tectonics of the Gulf of Corinth area showing the locations of active normal faults, large earthquakes of the past 100 years, and geodetically determined extension directions and rates. Courtesy of F. Cornet.

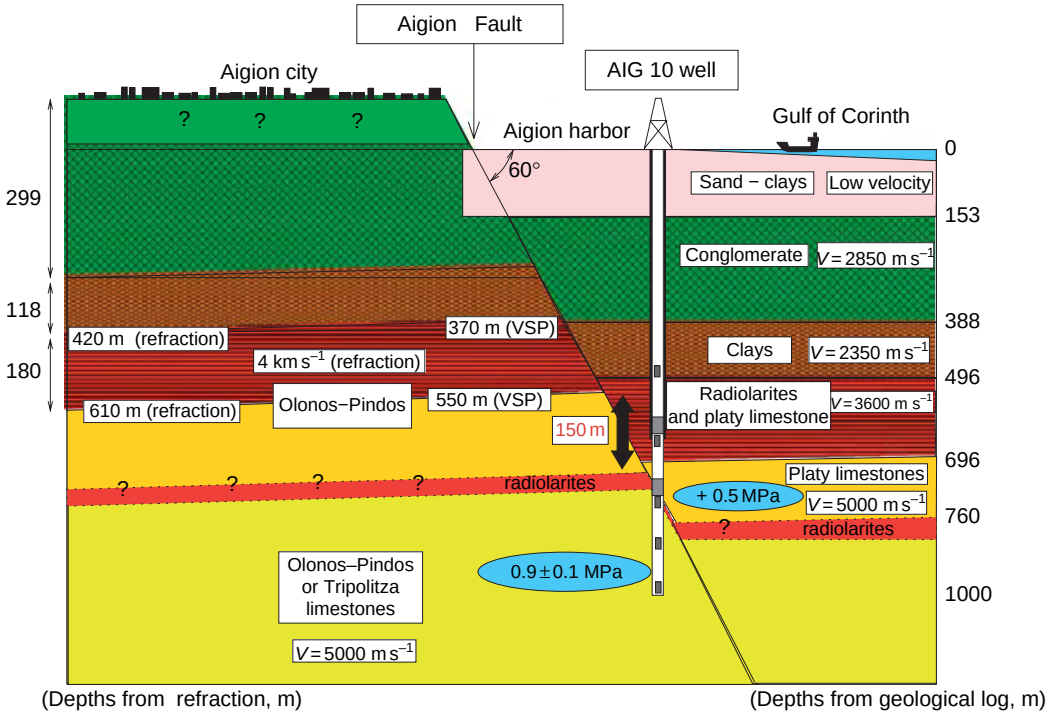


Figure 10 Geology in the vicinity of the AIG 10 well that was cored through the Aigion Fault. The fault was encountered at a depth of ~760 m. Pore pressure on the footwall side of the fault at depth was approximately 1 MPa in excess of hydrostatic, consistent with the fact that the footwall elevation is approximately 100 m higher than the elevation of the well. Courtesy of F. Cornet.

isothermal temperatures (31°C) below the fault is consistent with local convection in the karstic limestone while an absence there is an absence of flow above the fault. Geochemical data indicate a shallow continental origin of water in and adjacent to the fault zone, with a notable absence of deep fluid input from the mantle. The borehole is being used to make continuous measurements of pressure in the fault zone.

The data acquired to data show tidal variations, as well as pressure variations induced by remote earthquakes (Cornet *et al.*, 2004). The Corinth science team proposes to drill a 4.5 km deep borehole to monitor transients in pore pressure within the seismogenic zone and provide clues to the origin of deep fluids.

4.22.5.10 NELSAM

The deep gold mines of South Africa offer unique environments to study earthquakes by providing access to the focal area. The mining operations generate thousands of earthquakes per day, and some of these events approach M 5. The Natural Earthquake Laboratory in South African Mines (NELSAM) is a project that utilizes the availability of these deep mines for earthquake research (Reches *et al.*, 2006). The central part of this project is dense instrumentation and detailed characterization of a large fault zone in TauTona mine, which is the deepest mine on Earth (approaching 4 km, depth) located within the Western Deep Levels of the Witwatersrand basin, South Africa (**Figure 11**).

The laboratory is built around the Pretorius Fault Zone (shown in pink in the lower part of **Figure 11**), which is at least 10 km long with 30–200 m of displacement and which appears to have been inactive for the past 2.5 Ga. The mining plan for the next few years is likely to induce earthquakes of significant magnitude ($M > 2$) along this fault. NELSAM science activities at the site started in 2005 with site characterization, including mapping of 3D structure and composition of the Pretorius fault zones at 3.54 km depth. The emphasis of the investigations has been on segments of the fault zone that were reactivated during recent earthquakes. Analysis of the stress state both near and away from the excavation will be carried out using techniques developed for the oil industry in which compressive and tensile wellbore failures are studied in wells of different orientation (e.g., Zoback *et al.*, 2003). In addition, five boreholes, 20–60 m in length, have been drilled across the Pretorius Fault Zone. Once completed, the earthquake laboratory will include a dense array (250 m footprint) of three-component broadband accelerometers, seismometers, strain meters, temperature sensors, creep meters, electromagnetic radiation sensors, and acoustic emissions sensors. Fault-zone fluid chemistry will be monitored with an onsite mass spectrometer.

A M 2.2 earthquake that occurred on 12 December 2004 in the center of the planned earthquake laboratory that reactivated several segments of the Pretorius Fault (Heesakkers *et al.*, 2005). The mapping of the rupture zone of this earthquake in three dimensions revealed quasi-planar, cross-cutting reactivated segments with inclinations ranging from 21° to 90°. The rupturing

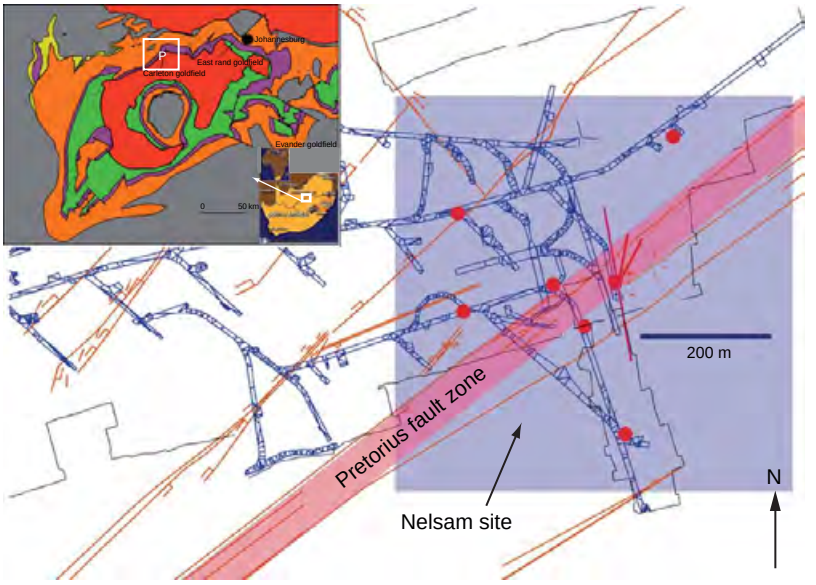


Figure 11 The TauTona mine is located in the Carleton goldfield, approximately 80 km west of Johannesburg, South Africa (see inset). Scientific drilling is occurring in the vicinity of the Pretorius Fault Zone (shown by pink shading) where a number of recent earthquakes have occurred recently. The figure shows the excavations at a depth of 3597–3657 m. Most of the faults exposed in the mine tunnels trend NE–SW, similar to the overall trend of the Pretorius Fault Zone. Courtesy of Z. Reches.

formed fresh fine-grained white rock powder almost exclusively along the contacts of the ancient, sintered cataclasite, and the quartzitic host rock.

4.22.5.11 NanTroSeize

An extensive fault-zone drilling campaign is planed as part of NanTroSeize experiment (**Figure 12(a)**) to be carried out by the Integrated Ocean Drilling Program in the Nankai trough just off southern Honshu, Japan. The extent of the 1944 Tonankai earthquake slip zone, the presumed extent of the locked zone (indicated by the heavy white line), and the presumed seismogenic zone are also shown.

The first-order objectives are to:

- document the material properties and state of the plate boundary fault system at several P–T and

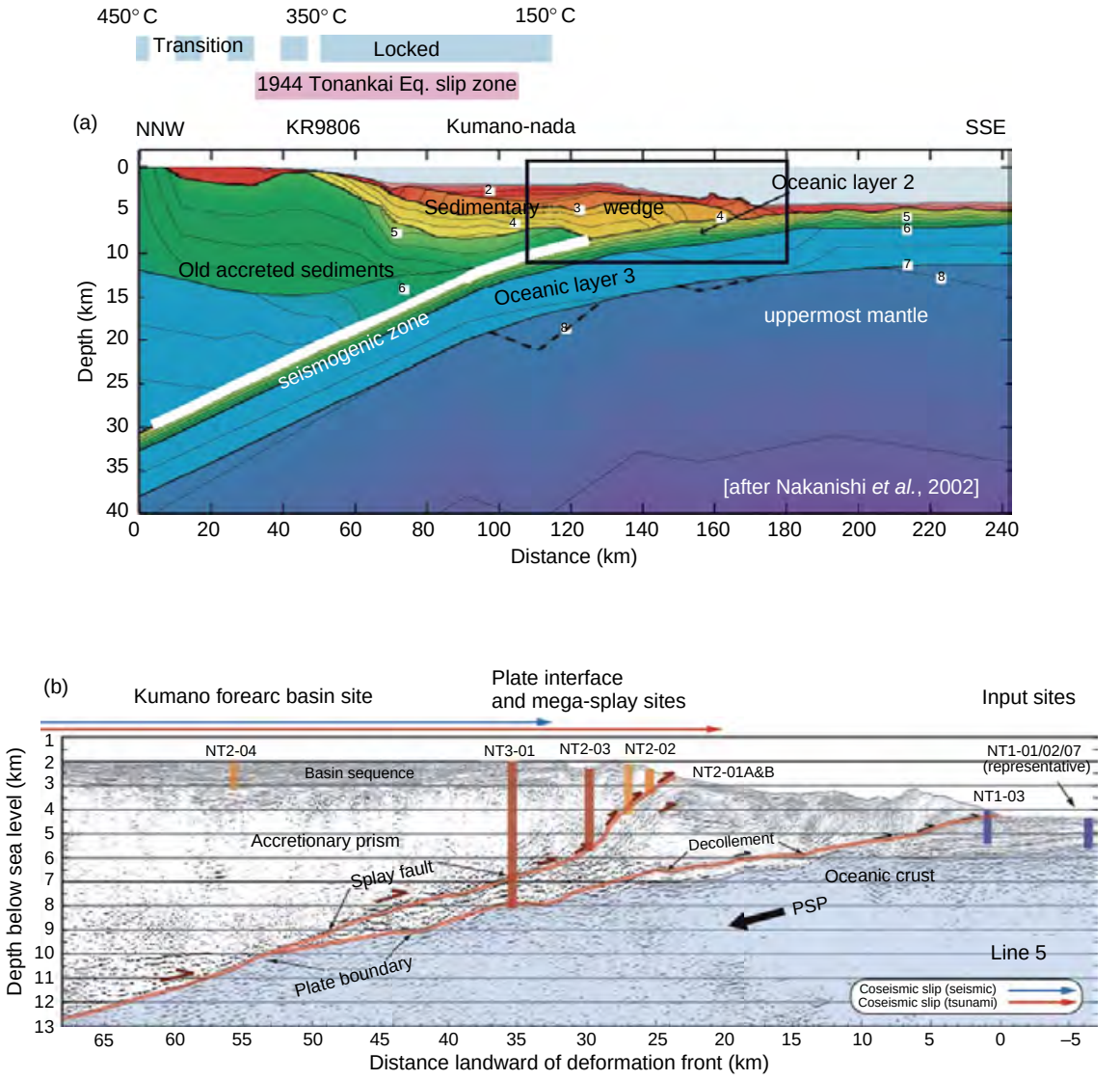


Figure 12 (a) Seismic velocity structure in the Nankai trough area reveals the nature of the downgoing Philippine Sea plate, the accretionary wedge and old accreted sediments (after Nakanishi *et al.*, 2002). The extent of the 1944 Tonankai earthquake slip zone, the presumed extent of the locked zone (indicated by the heavy white line) and the presumed seismogenic zone are also shown. (b) A detailed view of the proposed drilling area (see box in (a) where seismic reflection data reveals the detailed structure of the main décollement and splay faults in the accretionary prism). Note that a series of boreholes are planned to penetrate various parts of the main décollement and splay faults. Courtesy of H. Tobin.

lithology conditions, testing hypotheses for stable versus unstable frictional behavior;
- investigate partitioning between seismic versus aseismic processes on the main plate boundary, through monitoring of seismicity, borehole strain, and pore fluid pressure;
- test whether there are interseismic temporal changes in state along the plate-boundary thrust, including possible earthquake precursory signals; and
- use downhole measurements and sampling to calibrate observations made in the broader geophysical volume surrounding the boreholes.

In many ways, the specific science goals of this experiment are somewhat similar to those of the SAFOD project, although in a much different geologic setting. Specifically, the goals are to

- obtain samples of faults and surrounding environment
  - characterize lithology, structure, elastic and mechanical properties, porosity, permeability, pore fluid chemistry, microbiology;
  - use familiar ODP coring techniques, augmented by cuttings and sidewall coring;
- characterize the near-borehole environment;
  - geophysical logging: wireline and Logging While Drilling (LWD);
  - active testing for pore fluid pressure, stress, hydrogeologic properties (permeability, storage);
- monitor the borehole environment over time
  - passive and active source seismology;
  - strain and tilt;
  - pore fluid pressure;
  - temperature; and
  - EM field.

A detailed view of the proposed drilling area is shown in **Figure 12(b)** (see box in **Figure 12(a)**) where seismic reflection data reveal the detailed structure of the main décollement and splay faults in the accreationary prism. As shown, a series of relatively shallow holes will be drilled through splay faults in the accretionary wedge to study aseismic faulting. Following drilling, these holes will be used for downhole observatories. An eventual ultradeep borehole to be drilled into a seismogenic section of the décollement and possibly seismogenic deep part of the splay fault.

4.22.6 Summary

Drilling and downhole measurements in active fault zones provide critical tests of hypotheses arising from seismologic observations, laboratory rock deformation experiments, and geological observations of exhumed fault zones. Drilling provides the only direct means of measuring pore pressure, stress, permeability, and other important parameters within and near an active fault zone at depth. It is also the only way to collect fluid and rock samples from the fault zone and wall rocks at seismogenic depths and to monitor time-dependent changes in fluid pressure, fluid chemistry, deformation, temperature and electromagnetic properties at depth during the earthquake cycle. In the context of the processes and properties alluded to above, *in situ* observations and sampling through drilling perform two critical, and unique, functions. Sampling of fault rocks and fluids and downhole measurements provide essential constraints on mineralogy, grain size, fluid chemistry, temperature, stress, pore geometry, and other parameters that would allow laboratory investigations of fault-zone rheology and frictional behavior to be conducted under realistic *in situ* conditions. By *in situ* sampling, downhole measurement, and long-term monitoring in active fault zones, one is able to test and refine the broad range of current theoretical models for faulting and seismogenesis by providing realistic constraints on fault-zone physical properties, loading conditions, and mechanical behavior at depth. In particular, by comparing results of microstructural observations and rheological investigations on core with measurements of microseismicity, fluid pressure and deformation during the fault-zone monitoring phase of this experiment, we would be able to differentiate between fault-zone processes (e.g., fluid pressure fluctuations) associated with fault creep versus earthquakes.

Of course, no single drill hole into an active fault can address all of the questions listed above. Rather, like other avenues of earthquake research, the experiments being carried out in fault-related scientific drilling projects around the world (such as the ones noted above) cumulatively address many critical scientific questions about fault-zone structure, composition, and processes. Over time, the accumulation of data from these experiments will result in important advances in earthquake science.

References

Abercrombie RE (1995) Earthquake source scaling relationships from −1 to 5 M_1 using seismograms recorded at 2.5 km depth. *Journal of Geophysical Research* 100: 24015–24036.

Abercrombie RE and Leary P (1993) Source parameters of small earthquakes recorded at 2.5 km depth, Cajon pass, Southern California: Implications for earthquake scaling. *Geophysical Research Letters* 20: 1511–1514.

Ando M (2001) Geological and geophysical studies of the Nojima Fault from drilling: An outline of the Nojima Fault Zone Probe. *The Island Arc* 10: 206–214.

Andreani M, Boullier A-M, and Gratier J-P (2005) Development of schistosity by dissolution–crystallization in a California serpentinite gouge. *Journal of Structural Geology* 27: 2256–2267.

Andrews D (2002) A fluid constitutive relation accounting for thermal pressurization of pore fluid. *Journal of Geophysical Research* 107(B12): 2363 (doi:10.1029/2002JB001942).

Andrews DJ and Ben-Zion Y (1997) Wrinkle-like slip pulse on a fault between different materials. *Journal of Geophysical Research* 102: 553–571.

Angevine CL, Turcotte DL, and Furnish MD (1982) Pressure solution lithification as a mechanism for the stick-slip behavior of faults. *Tectonics* 1: 151–160.

Anooshehpoor A and Brune JN (1994) Frictional heat generation and seismic radiation in a foam rubber model of earthquakes. *Pure and Applied Geophysics* 142: 735–748.

Bangs NL, Shipley TH, Gulick SPS, Moore G, Kuromoto S, and Nakamura Y (2004) Evolution of the Nankai Trough decollement from the trench into the seismogenic zone; inferences from three-dimensional seismic reflection imaging. *Geology* 32: 273–276.

Barton CA, Zoback MD, and Burns KL (1988) *In-situ* stress orientation and magnitude at the Fenton Geothermal site, New Mexico, determined from wellbore breakouts. *Geophysical Research Letters* 15: 467–470.

Baumgärtner J and Zoback MD (1989) Interpretation of hydraulic fracturing pressure–time records using interactive analysis methods. *International Journal of Rock Mechanics, Mining Science and Geomechanical Abstracts* 26: 461–470.

Ben-Avraham Z and Zoback MD (1992) Transform-normal extension and asymmetric basins: An alternative to pull-apart models. *Geology* 20: 423–426.

Beroza GC and Ellsworth WL (1996) Properties of the seismic nucleation phase. *Tectonophysics* 261: 209–227.

Bird P (1984) Hydration-phase diagrams and friction of montmorillonite under laboratory and geologic conditions, with implications for shale compaction, slope stability, and strength of fault gouge. *Tectonophysics* 107: 235–260.

Blanpied ML, Lockner DA, and Byerlee JD (1991) Fault stability inferred from granite sliding experiments at hydrothermal conditions. *Geophysical Research Letters* 18: 609–612.

Boness N and Zoback MD (2006) A multi-scale study of the mechanisms controlling shear velocity anisotropy in the San Andreas Fault Observatory at Depth. *Geophysics* 7(5): F131–F146.

Bos B and Spiers CJ (2001) Experimental investigation into the microstructural and mechanical evolution of phyllosilicate-bearing fault rock under conditions favouring pressure solution. *Journal of Structural Geology* 23: 1187–1202.

Boullier A-M, Ohtani T, Fujimoto K, Ito H, and Dubois M (2001) Fluid inclusions in pseudotachylytes from the Nojima fault, Japan. *Journal of Geophysical Research* 106(B10): 21965–21978 (doi:10.1029/2000JB000043).

Boullier A-M and Robert F (1992) Paleoseismic events recorded in Archaean gold-quartz vein networks, Val d'Or, Albitibi, Quebec, Canada. *Journal of Structural Geology* 14: 161–179.

Brace WF and Kohlstedt DL (1980) Limits on lithospheric stress imposed by laboratory experiments. *Journal of Geophysical Research* 85: 6248–6252.

Brodsky EE and Kanamori H (2001) The elastohydrodynamic lubrication of faults. *Journal of Geophysical Research* 106: 16357–16374.

Brudy M, Zoback MD, Fuchs K, Rummel F, and Baumgärtner J (1997) Estimation of the complete stress tensor to 8 km depth in the KTB scientific drill holes: Implications for crustal strength. *Journal of Geophysical Research* 102: 18453–18475.

Bruhn RL, Parry WT, Yonkee WA, and Thompson T (1994) Fracturing and hydrothermal alteration in normal fault zones. *Pure and Applied Geophysics* 142: 609–644.

Bruhn RL, Yonkee WA, and Parry WT (1990) Structural and fluid-chemical properties of seismogenic normal faults. *Tectonophysics* 175: 139–157.

Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research* 75: 4997–5009.

Brune JN, Brown S, and Johnson PA (1993) Rupture mechanism and interface separation in foam rubber models of earthquakes: A possible solution to the heat flow paradox and the paradox of large overthrusts. *Tectonophysics* 218: 59–67.

Brune JN, Henyey TL, and Roy RF (1969) Heat flow, stress, and rate of slip along the San Andreas fault, California. *Journal of Geophysical Research* 74: 3821–3827.

Burford RO and Harsh PW (1980) Slip on the San Andreas fault in central California from alinement array surveys. *Bulletin of the Seismological Society of America* 70: 1233–1261.

Byerlee JD (1978) Friction of rocks. *Pure and Applied Geophysics* 116: 615–629.

Byerlee JD (1990) Friction, overpressure and fault normal compression. *Geophysical Research Letters* 17: 2109–2112.

Byerlee JD (1993) A model for episodic flow of high pressure water in fault zones before earthquakes. *Geology* 21: 303–306.

Byerlee JD and Summers R (1976) A note on the effect of fault gouge thickness on fault stability. *International Journal of Rock Mechanics, Mining Science and Geomechanical Abstracts* 13: 35–36.

Byrne T and Fisher D (1990) Evidence for a weak and overpressured décollement beneath sediment-dominated accretionary prisms. *Journal of Geophysical Research* 98: 9081–9097.

Chester FM, Evans JP, and Biegel RL (1993) Internal structure and weakening mechanisms of the San Andreas fault. *Journal of Geophysical Research* 98: 771–786.

Chester FM and Higgs NG (1992) Multimechanism friction constitutive model for ultrafine quartz gouge at hypocentral conditions. *Journal of Geophysical Research* 97: 1859–1870.

Cornet FH, Bernard P, and Moretti I (2004) The Corinth Rift Laboratory. *Comptes Rendus Geosciences* 336: 235–241.

Dahlen FA (1990) Critical taper model of fold-and-thrust belts and accretionary wedges. *Annual Review of Earth and Planetary Sciences* 18: 55–99.

Davis D, Suppe J, and Dahlen FA (1983) The mechanics of fold and thrust belts and accretionary wedges. *Journal of Geophysical Research* 88: 1153–1172.

Dewers T and Ortoleva P (1994) Nonlinear dynamical aspects of deep basin hydrology; fluid compartment formation and episodic fluid release. *American Journal of Science* 294: 713–755.

Dreger DS, Tkalcic H, and Johnston M (2000) Dilational processes accompanying earthquakes in the Long Valley Caldera. *Science* 288: 122–125.

Dieterich JH (1992) Earthquake nucleation on faults with rate- and state-dependent strength. *Tectonophysics* 211: 115–134.

Eichhubl P and Boles JR (2000) Rates of fluid flow in fault systems – Evidence for episodic rapid fluid flow in the Miocene Monterey Formation Coastal California. *American Journal of Science* 300: 571–600.

Ellsworth WL and Beroza GC (1995) Seismic evidence for an earthquake nucleation phase. *Science* 268: 851–855.

Ellsworth WL and Beroza GC (1998) Observation of the seismic nucleation phase in the Ridgecrest, California earthquake sequence. *Geophysical Research Letters* 25: 401–404.

Evans JP and Chester FM (1995) Fluid-rock interaction in rocks of the San Andreas system: Inferences from San Gabriel fault rock geochemistry and microstructures. *Journal of Geophysical Research* 100: 13007–13020.

Fehler MC (1989) Stress control of seismicity patterns observed during hydraulic fracturing experiments at the Fenton Hill hot dry rock geothermal energy site, New Mexico. *International Journal of Rock Mechanics and Mining Sciences* 26: 211–219.

Fisher DM (1996) Fabrics and veins in the forearc: A record of cyclic fluid flow at depths of <15 km. In: Bebout GE, Schou DW, Kirby SH, and Platt JJ (eds.) *American Geophysical Union Geophysical Monograph, Vol. 96: Subduction: Top to Bottom*, pp. 75–89. Washington, DC: American Geophysical Union.

Fitzenz DD and Miller SA (2003) Fault compaction and overpressured faults: Results from a 3D model of a ductile fault zone. *Geophysical Journal International* 155: 111–125.

Flesch LM, Holt WE, Haines AJ, and Shen-Tu B (2000) Dynamics of the Pacific-North American plate boundary in the Western United States. *Science* 287: 834–836.

Fournier RL (1996) Compressive and tensile failure at high fluid pressure where preexisting fractures have cohesive strength, with application to the San Andreas fault. *Journal of Geophysical Research* 101: 25499–25509.

Fujimoto K, Tanaka H, Higuchi T, Tomida N, Ohtani T, and Ito H (2001) Alteration and mass transfer inferred from the Hirabayashi GSJ drill penetrating the Nojima fault, Japan. *The Island Arc* 10: 401–410.

Fyfe S, Price NJ, and Thomson AB (1978) *Fluids in the Earth's Crust*. New York: Elsevier.

Hacker BR, Blanpied ML, Lockner DA, and Jové CF (1995) Dehydration and friction: Laumontite -> wairakite + H_2O. *Geological Society of America, Abstracts with Programs* 27: A282.

Hainzl S and Ogata Y (2005) Detecting fluid signals in seismicity data through statistical earthquake modeling. *Journal of Geophysical Research* 110: B05S07 (doi:10.1029/2004JB003247).

Harris RA (1998) Introduction to special section: Stress triggers, stress shadows and implications for seismic hazard. *Journal of Geophysical Research* 103: 24347–24358.

Harris RA (2004) Numerical simulations of large earthquakes: Dynamic rupture propagation on heterogeneous faults. *Pure and Applied Geophysics* 161: 2171–2181 (doi:10.1007/s00024-004-2556-8).

Healy JH, Rubey WW, Griggs DT, and Raleigh CB (1968) The Denver earthquakes. *Science* 161: 1301–1310.

Heaton TH (1990) Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors* 64: 1–20.

Heesakkers V, Murphy SK, van Aswegen G, *et al.* (2005) The rupture zone of the M = 2.2 earthquake that reactivated the ancient Pretorius Fault in TauTona Mine, South Africa. *EOS Transactions of the American Geophysical Union* 86: S31B-04.

Henyey TL and Wasserburg GJ (1971) Heat flow near major strike-slip faults in California. *Journal of Geophysical Research* 76: 7924–7946.

Hickman S (1991) Stress in the lithosphere and the strength of active faults, US National Report on International Union of Geodesy and Geophysics 1987–1990. *Reviews of Geophysics* 29: 759–775.

Hickman S, Barton CA, Zoback MD, Morin R, Sass J, and Benoit R (1997) *In-situ* stress and fracture permeability along the Stillwater fault zone, Dixie Valley, Nevada. *International Journal of Rock Mechanics and Mining Sciences* 34: 3–4 (Paper No. 126).

Hickman S and Evans B (1992) Growth of grain contacts in halite by solution transfer: Implications for diagenesis, lithification, and strength recovery. In: Evans B and Wong T-F (eds.) *Fault Mechanics and Transport Properties of Rocks*, pp. 253–280. San Diego, CA: Academic Press.

Hickman S, Sibson R, and Bruhn R (1995) Introduction to special section: Mechanical involvement of fluids in faulting. *Journal of Geophysical Research* 100: 12831–12840.

Hickman S and Zoback MD (2004) Stress measurements in the SAFOD pilot hole: Implications for the frictional strength of the San Andreas fault. *Geophysical Research Letters* 31: L15S12.

Hickman S, Zoback MD, and Ellsworth WE (2004) Introduction to special section: Preparing for the San Andreas Fault Observatory at Depth. *Geophysical Research Letters* 31: L12S01 (doi:10.1029/20040GL20688).

Hubbert MK and Rubey WW (1959) Role of fluid pressure in mechanics of overthrust faulting. *Geological Society of America, Bulletin* 70: 115–205.

Hunt JM (1990) Generation and migration of petroleum from abnormally pressured fluid compartments. *American Association of Petroleum Geologists Bulletin* 74: 1–12.

Iio Y (1992) Slow initial phase of the S-wave velocity pulse generated by microearthquakes. *Geophysical Research Letters* 19: 477–480.

Iio Y (1995) Observation of the slow initial phase of microearthquakes implications for earthquake nucleation and propagation. *Journal of Geophysical Research* 100: 15333–15349.

Imanishi K, Ellsworth WE, and Prejean SG (2004) Earthquake source parameters eterminded by the SAFOD pilot hole array. *Geophysical Research Letters* 31: L12S09 (doi:10.1029/2004GL019420).

Imanishi K and Ellsworth WL (2006) Source scaling relationships of microearthquakes at Parkfield, CA, determined using the SAFOD pilot hole seismic array. In: McGarr A, Kanamori H, and Abercrombie R (eds.) *AGU Monograph on Radiated Energy and the Physics of Earthquake Faulting*. Geophysical Monograph Series 170, pp. 81–90. American Geophysical Union.

Janecke SU and Evans JP (1988) Feldspar-influenced rock rheologies. *Geology* 16: 1064–1067.

Kanamori H (1980) State of stress in the Earth's lithosphere. In: Dziewonski AM and Boschi E (eds.) *Physics of the Earth's Interior*, pp. 531–554. Bologna, Italy: Societá Italiana di Fisica.

Kano H, Mori J, Fujio R, *et al.* (2006) Heat signature of the Chelungpu fault associated with the 1999 Chi-chi Taiwan earthquake. *Geophysical Research Letters* 33: L14306.

Kao H and Angelier J (2001) Stress tensor inversion for the Chi-chi earthquake sequence and its implications on regional collision. *Bulletin of the Seismological Society of America* 91: 1028–1040.

Karner S, Marone C, and Evans B (1997) Laboratory study of fault healing and lithification in simulated fault gouge under hydrothermal conditions. *Tectonophysics* 277: 41–55.

Kennedy BM, Kharaka YH, Evans WC, *et al.* (1997) Mantle fluids in the San Andreas fault system, California. *Science* 278: 1278–1281.

Kerrich R, La Tour TE, and Willmore L (1984) Fluid participation in deep fault zones: Evidence from geological, geochemical, and $^{18}O/^{16}O$ relations. *Journal of Geophysical Research* 89: 4331–4343.

Kirby SH (1983) Rheology of the lithosphere. *Reviews of Geophysics and Space Physics 21:* 1458–1487.

Ko S-Z, Olgaard DL, and Wong T-F (1997) Generation and maintenance of pore presure excess in a dehydrating system 1. Experimental and microstructural observations. *Journal of Geophysical Research* 102: 825–839.

Kostrov BV (1964) Self-similar problem of propagation of shear cracks. *Journal of Applied Mathematics and Mechanics* 28: 1077–1087.

Lachenbruch AH (1980) Frictional heating, fluid pressure, and the resistance to fault motion. *Journal of Geophysical Research* 85: 6097–6112.

Lachenbruch AH and McGarr A (1990) Stress and heat flow. In: Wallace RE (ed.) *US Geological Survey Professional Paper 1515: The San Andreas Fault System, California*, pp. 261–277. Menlo Park, CA: USGS.

Lachenbruch AH and Sass JH (1980) Heat flow and energetics of the San Andreas fault zone. *Journal of Geophysical Research* 85: 6185–6223.

Lachenbruch AH and Sass JH (1992) Heat flow from Cajon Pass, fault strength and tectonic implications. *Journal of Geophysical Research* 97: 4995–5015.

Lachenbruch AH and Thompson GA (1972) Oceanic ridges and transform faults: Their intersection angles and resistance to place motion. *Earth and Planetary Science Letters* 15: 116–122.

Lockner DA and Byerlee JD (1995) An earthquake instability model based on faults containing high fluid-pressure compartments. *Pure and Applied Geophysics* 145: 717–745.

Logan JM and Rauenzahn KA (1987) Frictional dependence of gouge mixtures of quartz and montmorillonite on velocity, composition and fabric. *Tectonophysics* 144: 87–108.

Lund B and Zoback MD (1999) State of stress to 7 km depth in the Siljan drill holes from observations of drilling-induced tensile wall fractures. *International Journal of Rock Mechanics* 36: 169–190.

Ma K-F, Brodsky EE, Mori J, Song C, Ji TR, and Kanamori H (2003) Evidence for fault lubrication during the 1999 Chi-Chi, Taiwan, earthquake (Mw 7.6). *Geophysical Research letters* 30: 1244 (doi:10.1029/2002GL015380).

Ma K-F, Tanaka H, Song S-R, *et al.* (2006) Slip zone and energetics of a large earthquake from the Taiwan Chelungpu-fault Drilling Project. *Nature* 444: 473–476.

Madariaga R (1976) Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America* 66: 639–666.

Magee ME and Zoback MD (1993) Evidence for a weak interpolate thrust fault along the northern Japan subduction zone and implications for the mechanics of thrust faulting and fluid expulsion. *Geology* 21: 809–812.

Marone C, Raleigh CB, and Scholz CH (1990) Frictional behavior and constitutive modeling of simulated fault gouge. *Journal of Geophysical Research* 95: 7007–7025.

Mase CW and Smith L (1987) Effects of frictional heating on the thermal, hydrological, and mechanical response of a fault. *Journal of Geophysical Research* 92: 6249–6272.

Martinsen RS (1997) Stratigraphic controls on the development and distribution of fluid-pressure compartments. *American Association of Petroleum Geologists Memoir* 67: 223–241.

McNutt M (1980) Implications of regional gravity for state of stress in the Earth's crust and upper mantle. *Journal of Geophysical Research* 85: 6377–6396.

McNutt MK and Menard HW (1982) Constraints on yield strength in the oceanic lithosphere derived from observations of flexure. *Geophysical Journal of the Royal Astronomical Society* 71: 363–394.

Melosh HJ (1979) Acoustic fluidization: A new geologic process?. *Journal of Geophysical Research* 84: 7513–7520.

Melosh HJ (1996) Dynamic weakening of faults by acoustic fluidization. *Nature* 379: 601–606.

Miller SA (1996) Fluid-mediated influence of adjacent thrusting on the seismic cycle at Parkfield. *Nature* 382: 799–802.

Moore DE, Lockner DA, Summers R, Shengli M, and Byerlee JD (1997) Strengths of serpentinite gouges at elevated temperatures. *Journal of Geophysical Research* 102: 14787–14801.

Moore JC, Shipley TH, Goldberg D, *et al.* (1995) Abnormal fluid pressure and fault-zone dilation in the Barbados acretionary prism: Evidence from logging while drilling. *Geology* 23: 605–608.

Moore JC and Vrolijk P (1992) Fluids in accretionary prisms. *Reviews of Geophysics* 30: 113–135.

Morrow CA, Shi LQ, and Byerlee JD (1982) Strain hardening and strength of clay-rich fault gouges. *Journal of Geophysical Research* 87: 6771–6780.

Morrow C, Radney B, and Byerlee J (1992) Frictional strength and the effective pressure law of montmorillonite and illite clays. In: Evans B and Wong T-F (eds.) *Fault Mechanics and Transport Properties of Rocks*, pp. 69–88. San Diego, CA: Academic Press.

Mount VS and Suppe J (1987) State of stress near the San Andreas fault: Implications for wrench tectonics. *Geology* 15: 1143–1146.

Mount VS and Suppe J (1992) Present-day stress orientations adjacent to active strike-slip faults: California and Sumatra. *Journal of Geophysical Research* 97: 11995–12013.

Mrotek KA, Quittmeyer RC, Naumhoff PG, and Stratton CT (1988) Observations of the earthquake swarm near Moodus, Connecticut. September/October, 1987. *EOS Transactions of the American Geophysical Union* 69: 495.

Murray MH and Segall P (2001) Modeling broadscale deformation in northern California and Nevada from plate motions and elastic strain accumulation. *Geophysical Research Letters* 28(22): 4315–4318.

Nadeau R, Antolik M, Johnson P, Foxall W, and McEvilly TV (1994) Seismological studies at Parkfield III: Microearthquake clusters in the study of fault-zone dynamics. *Bulletin of the Seismological Society of America* 84: 247–263.

Nadeau RM and McEvilly TV (1997) Seismological studies at Parkfield V: Characteristic microearthquake sequences as fault-zone drilling targets. *Bulletin of the Seismological Society of America* 87: 1463–1472.

Nadeau RM, Foxall W, and McEvilly TV (1995) Clustering and periodic recurrence of microearthquakes on the San Andreas fault at Parkfield, California. *Science* 267: 503–507.

Nadeau RM, Michelini A, Uhrhammer RA, Dolenc D, and McEvilly TV (2004) Detailed kinematics, structure and recurrence of micro-seismicity in the SAFOD target region. *Geophysical Research Letters* 31: L12S08 (doi:10.1029/2003GL019409).

Nakanishi A, Takahashi N, Park J-O, *et al.* (2002) Crustal structure across the coseismic rupture zone of the 1944 Tonankai earthquake, the central Nanki Trough seismogenic zone. *Geophysical Research* 107: doi:10.1029/2001JB00424.

Nur A and Booker JR (1972) Aftershocks caused by pore-fluid flow? *Science* 175: 885–887.

Ohnaka M (1992) Earthquake source nucleation: A physical model for short-term precursors. *Tectonophysics* 211: 149–178.

Ohtani T, Tanaka H, Fujimoto K, Higuchi T, Tomida N, and Ito H (2001) Internal structure of the Nojima fault zone from the Hirabayashi GSJ drill core. *The Island Arc* 10: 392–400.

Oppenheimer DH, Reasenberg PA, and Simpson RW (1988) Fault-plane solutions for the 1984 Morgan Hill, California earthquake sequence: Evidence for the state of stress on the Calaveras fault. *Journal of Geophysical Research* 93: 9007–9026.

Oshiman N, Shimamoto T, Takemura K, and Wibberley CAJ eds. (2001) Nojima Fault Zone probe. *The Island Arc* 10: 195–505.

Parry WT (1994) Fault fluid compositions from fluid inclusion observations. In: Hickman S, Sibson R, and Bruhn R (eds.) *Proceedings USGS Red Book Conference on the Mechanical Involvement of Fluids in Faulting*, pp. 334–348. Menlo Park, CA: USGS.

Peyrat S, Olsen KB, and Madariaga R (2004) Which dynamic rupture parameters can be estimated from strong ground motion and geodetic data? *Pure and Applied Geophysics* 161: 2155–2169 (doi:10.1007/s00024-004-2555-9).

Pinkston J, Stern L, and Kirby S (1987) Hydrothermal reactions on artificial fault surfaces in dunite: Fibrous mineral growth, slickensides and temperature sensitivity of reaction weakening. *EOS Transactions of the American Geophysical Union* 68: 405.

Power WL and Tullis TE (1989) The relationship between slickenside surfaces in fine-grained quartz and the seismic cycle. *Journal of Structural Geology* 11: 879–893.

Powley DE (1990) Pressures and hydrogeology in petroleum basins. *Earth, Science Reviews* 29: 215–226.

Prejean SG and Ellsworth WL (2001) Observations of earthquake source parameters at 2 km depth in the Long Valley Caldera, Eastern California. *EOS Transactions of the American Geophysical Union* 91: 165–177.

Raleigh CB, Healy JH, and Bredehoeft JD (1972) An experiment in earthquake control at Rangely, Colorado. *Science* 191: 1230–1237.

Reches Z, DAFSAM, and NELSAM teams (2006) Building a natural earthquake laboratory at focal depth (DAFSAM-NELSAM Project, South Africa). *Scientific Drilling* 3: 30–33.

Reinen LA and Tullis TE (1995) Microstructural evidence of strain localization and distributed strain in serpentine friction experiments. *EOS Transactions of the American Geophysical Union* 76: F560.

Reinen LA, Weeks JD, and Tullis TE (1994) The frictional behavior of lizardite and antigorite sepentinites: Experiments, constitutive models, and implications for natural faults. *Pure and Applied Geophysics* 143: 317–358.

Rice JR (1992) Fault stress states, pore pressure distributions, and the weakness of the San Andreas fault. In: Evans B and Wong T-F (eds.) *Fault Mechanics and Transport Properties of Rocks*, pp. 475–503. San Diego, CA: Academic Press.

Rice JR (2006) Heating and weakening of faults during earthquake slip. *Journal of Geophysical Research* 111: B05311 (doi:10.1029/2005JB004006, 29 pages).

Rice JR and Cleary MP (1976) Some basic stress diffusion solutions for fluid-saturated elastic porous media with compressible constituents. *Reviews of Geophysics* 14: 227–241.

Rudnicki J (1988) Physical models of earthquake instability and precursory processes. *Pure and Applied Geophysics* 126: 524–531.

Rutter EH and Mainprice DH (1979) On the possibility of slow fault slip controlled by diffusive mass transfer processes. *Gerlands Beitrage zur Geophysik* 88: 154–162.

Saffer DM and Bekins BA (2006) An evaluation of factors influencing pore pressure in accretionary complexes: Implications for taper angle and wedge mechanics. *Journal of Geophysical Research* 111: B04101 (doi:10.1029/2005JB003990).

Saffer DM, Bekins BA, and Hickman S (2003) Topographically driven groundwater flow and the San Andreas heat flow paradox revisited. *Journal of Geophysical Research* 108(B5): 2274.

Scholz CH (1989) Mechanics of faulting. *Annual Review of Earth and Planetary Sciences* 17: 309–334.

Scholz CH (2000) Evidence for a strong San Andreas fault. *Geology* 28(2): 163–166.

Schulz SE and Evans JP (2000) Mesoscopic structure of the Punchbowl fault, southern California and the geologic and geophysical structure of active strike-slip faults. *Journal of Structural Geology* 22: 913–930.

Schulz SS (1989) Catalog of creepmeter measurements in California from 1966 through 1988. *US Geological Survey Open-File Report 89–650*. Menlo Park, CA: USGC.

Segall P and Rice JR (1995) Dilatancy, compaction and slip instability of a fluid-infiltrated fault. *Journal of Geophysical Research* 100: 22155–22171.

Shimamoto T and Logan J (1981) Effects of simulated clay gouges on the sliding behavior of Tennessee sandstone. *Tectonophysics* 75: 243–255.

Shipley TH, Moore GF, Bangs NL, Moore JC, and Stoffa PL (1994) Seismically inferred dilatancy distribution, northern Barbados Ridge decollement: Implications for fluid migration and fault strength. *Geology* 22: 411–414.

Sibson RH (1973) Interactions between temperature and pore-fluid pressure during earthquake faulting: A mechanism for partial or total stress relief. *Nature, Physical Science* 243: 66–68.

Sibson RH (1985) Stopping of earthquake ruptures at dilational fault jogs. *Nature* 316: 248–251.

Sibson RH (1990) Rupture nucleation on unfavorably oriented faults. *Bulletin of the Seismological Society of America* 80: 1580–1604.

Sleep NH (1995) Ductile creep, compaction and rate and state dependent friction within major fault zones. *Journal of Geophysical Research* 100: 13065–13080.

Sleep NH and Blanpied ML (1992) Creep, compaction, and the weak rheology of major faults. *Nature* 359: 687–692.

Sleep NH and Blanpied ML (1994) Ductile creep and compaction: A mechanism for transiently increasing fluid pressure in mostly sealed fault zones. *Pure and Applied Geophysics* 143: 9–40.

Smith L, Forster CB and Evans JP (1990) Interaction of fault zones, fluid flow and heat transfer on a basin scale. In: Newman SP and Neretnieks I (eds.) *Hydrogeology of Low Permeability Environments*. International Association of Hydrological Sciences, Selected Papers, 2, pp. 41–67.

Spray JG (1987) Artificial generation of pseudotachylite using friction welding apparatus: Simulation of melting on a fault plane. *Journal of Structural Geology* 9: 49–60.

Stock JM, Healy JH, Hickman SH, and Zoback MD (1985) Hydraulic fracturing stress measurements at Yucca Mountain, Nevada, and relationship to regional stress field. *Journal of Geophysical Research* 90: 8691–8706.

Tanaka H, Chen WM, Kawabata K, and Urata N (2007) Thermal properties across the Chelungpu fault zone and evaluations of positive thermal anomaly on the slip zones: Are these residuals of heat from faulting. *Geophysical Research Letters* 34: L1309.

Tanaka H, Kujimoto K, Ohtani T, and Ito H (2001) Structural and chemical characterization of shear zones in the freshly activated Nojima fault, Awaji Island, southwest Japan. *Journal of Geophysical Research* 106: 8789–8810.

Tanaka H, Wang CY, Chen WM, *et al.* (2002) Initial science report of shallow drilling penetrating into the Chelungpu fault zone, Taiwan. *Terrestrial, Atmospheric and Oceanic Sciences* 13: 227–251.

Townend J and Zoback MD (2000) How faulting keeps the crust strong. *Geology* 28(5): 399–402.

Townend J and Zoback MD (2004) Regional tectonic stress near the San Andreas fault in Central and Northern California. *Geophysical Research Letters* 31(13): L15S11.

Townend J and Zoback MD (2001) Implications of earthquake focal mechanisms for the frictional strength of the San Andreas fault system. In: Holdsworth RE, Strachan RA, Macloughlin JJ, and Knipe RJ (eds.) *Special Publication of the Geological Society of London, Vol. 186: The Nature and Tectonic Significance of Fault Zone Weakening*, pp. 13–21. Geological Society of London: London.

Wald DJ and Heaton TH (1994) Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake. *Bulletin of the Seismological Society of America* 84: 668–691.

Wakita H and Sano Y (1987) High 3He emanation and seismic swarms observed in a nonvolcanic, forearc region. *Journal of Geophysical Research* 92: 12539–12546.

Wang CY, Mao N, and Wu FT (1980) Mechanical properties of clays at high pressure. *Journal of Geophysical Research* 85: 1462–1468.

Wang K, Mulder T, Rogers G, and Hyndman R (1995) Case for very low coupling stress on the Cascadia subduction fault. *Journal of Geophysical Research* 100: 12907–12918.

Wilcock WSD, Purdy GM, and Solomon SC (1990) Microearthquake evidence for extension across the Kane transform fault. *Journal of Geophysical Research* 95: 15439–15462.

Williams CF, Grubb FV, and Galanis SP (2004) Heat flow in the SAFOD pilot hole and implications for the strength of the San Andreas Fault. *Geophysical Research Letters* 31: L15S14 (doi:10.1029/2003GL019352).

Wintsch R, Christoffersen R, and Kronenberg A (1995) Fluid-rock reaction weakening of fault zones. *Journal of Geophysical Research* 100: 13021–13032.

Wu FT (1978) Mineralogy and physical nature of clay gouge. *Pure and Applied Geophysics* 116: 655–689.

Wu H-Y, Ma K-F, Zoback M, Boness N, Ito H, and Hung J-H (2007) Stress orientations of Taiwan Chelungpu-Fault Drilling Project (TCDP) hole-A as observed from geophysical logs. *Geophysical Research Letters* 34: L01303.

Yue LF, Suppe J, and Hung JH (2005) Structural geology of a classic thrust belt earthquake: The 1999 Chi-Chi earthquake Taiwan (M 7.6). *Journal of Structural Geology* 27: 2058–2083.

Zoback MD (2000) Strength of the San Andreas. *Nature* 405: 31–32.

Zoback MD, Barton CA, Brudy M, *et al.* (2003) Determination of stress orientation and magnitude in deep wells. *International Journal of Rock Mechanics* 40: 1049–1076.

Zoback MD and Beroza GC (1993) Evidence for near-frictionless faulting in the 1989 (M 6.9) Loma Prieta, California, earthquake and its aftershocks. *Geology* 21: 181–185.

Zoback MD and Healy JH (1984) Friction, faulting, and '*in situ*' stresses. *Annales Geophysicae* 2: 689–698.

Zoback MD and Healy JH (1992) *In situ* stress measurements to 3.5 km depth in the Cajon Pass Scientific Research Borehole: Implications for the mechanics of crustal faulting. *Journal of Geophysical Research* 97: 5039–5057.

Zoback MD, Zoback ML, Mount VS, *et al.* (1987) New evidence on the state of stress of the San Andreas fault system. *Science* 238: 1105–1111.

4.23 Complexity and Earthquakes

D. L. Turcotte, R. Shcherbakov, and J. B. Rundle, University of California, Davis, CA, USA

Nomenclature

b	b-value
c	characteristic time (s)
e	energy density (J kg^{-1})
k_c	connecting spring constant (N m^{-1})
k_L	loader plate spring constant (N m^{-1})
m	earthquake magnitude
p	exponent in the Omori's law
r	range of interaction (m)
r	decay rate of aftershocks
s	power-law exponent
t	time (s)
t_f	time to failure (s)
t_d	characteristic damage time (s)
A	area (m^2)
C_v	coefficient of variation
D	fractal dimension
E	seismic energy release (J)
E	Young's modulus (Pa)
M	seismic moment (J)
N	number of events
N_{as}	number of aftershocks
P	cumulative distribution function
V	driver plate velocity (m s^{-1})
α	damage variable
β	modified b-value
ε	strain
ε_y	yield strain
ε_B	cumulative Benioff strain
μ	mean
ν	hazard rate (s^{-1})
ξ	correlation length (m)
σ	standard deviation
σ	stress (Pa)
σ_y	yield stress (Pa)
τ	characteristic time (s)
ρ	power-law exponent

Glossary

accelerated moment release (AMR) A systematic regional increase in seismic activity prior to an earthquake.
aftershocks Earthquakes that follow the largest earthquake in a sequence.
Båth's law The observation that the difference in magnitude between a main shock and the largest aftershock is approximately constant.
cellular automata model A model consisting of a grid of elements that interact through specified rules.
characteristic earthquake A sequence of earthquakes that occur quasi-periodically on a fault or fault segment with approximately equal magnitudes.
correlation length The length over which events are correlated.
critical point A system is at the critical point when the correlation length goes to infinity.
damage Nonelastic deformation of a solid.
deterministic chaos The behavior of a set of deterministic equations in which adjacent solutions diverge exponentially.
epidemic-type aftershock sequence (ETAS) model A branching model for aftershock generation that includes aftershocks of aftershocks, etc.
foreshocks Earthquakes that precede the largest earthquake in a sequence.
fractal A power-law dependence of the number of events on the characteristic length of the events as defined in [36].
Gutenberg–Richter frequency–magnitude distribution Relation between the number of earthquakes and their magnitudes as defined in [1].
interoccurrence times The time intervals between successive earthquakes in a region.
main shock The largest earthquake in an associated sequence of earthquakes.
Omori's law Dependence of the rate of aftershock occurrence on the time since the main shock occurred as defined in [7].
recurrence times The time intervals between characteristic earthquakes.
self-organized criticality A theory describing the behavior of complex systems consisting of many interacting elements. The self-organized critical system oscillates about a quasi-steady state, there is a steady input of energy into the system and a stochastic output through avalanches whose magnitudes are distributed according to a power law.

4.23.1 Introduction

Complex behavior manifests itself in various natural phenomena. Examples can be found at different levels of organization of natural systems ranging from molecules and microorganisms to the large-scale structure of the universe and the existence of life. All these systems are characterized by highly nonlinear interactions among their constituent parts and behavior far from equilibrium. As a result, complexity is achieved through the internal nonlinear dynamics and external driving.

Earthquakes represent one example of such complex behavior. They usually occur in the upper brittle layer of the Earth's crust as a result of instabilities generated by moving plates. Constant driving assisted by the convective movement of the mantle supplies the energy needed to sustain a seismogenic region in a state far from equilibrium. This state is reached through the processes of self-organization and is characterized by long-range temporal and spatial correlations and scale-invariant behaviors. It is also believed that correlations are inherently present in the Earth's crust and are signatures of the nonlinear dynamics of rheological, frictional, chemical, and other processes. From this point of view, the Earth's crust can be considered as a strongly correlated nonlinear dissipative system.

Theoretical studies of seismology are usually dominated by applications of classical physics. However, many aspects of seismicity exhibit complex nonlinear behaviors. These aspects are best addressed using concepts developed in modern statistical physics. Examples include scaling laws associated with phase changes and the chaotic behavior of dynamical systems. Related concepts include scale invariance and fractality, deterministic chaos in low-dimensional systems, self-organized criticality, etc. Relevant models are generally nonlinear and stochastic in nature and the standard approaches of mathematical physics are not applicable. Instead, numerical simulations are used.

The main aim of this chapter is to introduce the reader to different aspects of scaling properties of earthquakes which are the manifestation of the complex nature of earthquake processes and illustrate the

methods and techniques developed in statistical physics to study these properties. This chapter is organized as follows. In Section 4.23.2 we will discuss relevant observations concerning the occurrence and statistics of earthquakes. We will emphasize applicable scaling laws. Examples include:

1. Gutenberg–Richter (GR) frequency–magnitude relation. This is a universally applicable scaling law that relates the cumulative number of earthquakes to their magnitudes and rupture areas.
2. Omori's law for the decay of aftershock activity following a main shock. This is another universally applicable scaling law that gives the dependence of the rate of aftershock production on the time since the main shock occurred.
3. Båth's law. This law states that the magnitude of the largest aftershock has a universal scaling relative to the magnitude of the main shock.

Other less known, and in some cases more controversial, scaling laws include

1. Accelerated moment release (AMR) prior to large earthquakes. This is a systematic increase in the cumulative moment (energy) of smaller earthquakes prior to a large earthquake.
2. Recurrence times. This is a statistical distribution of the recurrence times between characteristic earthquakes on a fault, possibly satisfying the Weibull distribution.
3. Interoccurrence statistics in both time and space of all earthquakes in a region. For aftershock sequences following large main shocks, the distribution of interoccurrence times between aftershocks appears to satisfy the nonhomogeneous Poisson statistics to a good approximation.

Our presentations will emphasize aftershocks since the available data can be of very good quality (Kisslinger, 1996). The major problem with earthquakes in general is that the recurrence times are large compared with the period for which data are available. One approach used to overcome this problem is to utilize numerical simulations to produce synthetic earthquake catalogs.

In Section 4.23.3 we will discuss the general characteristics of self-organizing complex systems. These include

1. Fractality, the applicability of power laws, self-similar scaling laws.
2. Deterministic chaos. Low-order systems related to seismicity have been shown to exhibit deterministic chaos. The implications of this association will be discussed.
3. Criticality and self-organized criticality. The applicability of these concepts is certainly controversial but must be addressed.

In Section 4.23.4 relevant models will be discussed. We classify these models as

1. Cellular automata models. The classic example relevant to seismicity is the multiple slider-block model.
2. Numerical simulations. Numerical simulations are widely used to forecast weather. Initial attempts to apply these approaches to seismicity are being carried out.
3. Damage mechanics models. Damage mechanics is used to model the complex brittle behavior in engineering materials. Applications to seismicity will be discussed.

Finally, in Section 4.23.5, the material presented earlier will be summarized. Future directions will also be considered. A specific example is earthquake forecasting (prediction).

4.23.2 Relevant Observations

4.23.2.1 GR Frequency–Magnitude Scaling

Earthquakes are known to universally (with a few exceptions) satisfy a relatively simple frequency–magnitude scaling. In terms of the earthquake magnitude m this scaling satisfies the GR frequency–magnitude relation (Gutenberg and Richter, 1954)

$$\log N(\geq m) = a - bm \qquad [1]$$

where N is the cumulative number of earthquakes with a magnitude greater than m occurring in a specified area and time, and b and a are constants. This relation is valid for earthquakes both regionally and globally. The constant b or 'b-value' varies from region to region, but is generally in the range of $0.8 < b < 1.2$ (Frohlich and Davis, 1993). Although the earthquake magnitude m has an empirical basis, we will show that [1] is a power-law (fractal) relation.

The frequency–magnitude distributions of the regional seismicity in southern California on a yearly basis are plotted in **Figure 1** using data obtained from the *SCSN* catalog. For each individual year between 1983 and 2002, the cumulative number of earthquakes N with magnitudes greater than m is plotted as a function of m. The period 1983–2003 taken

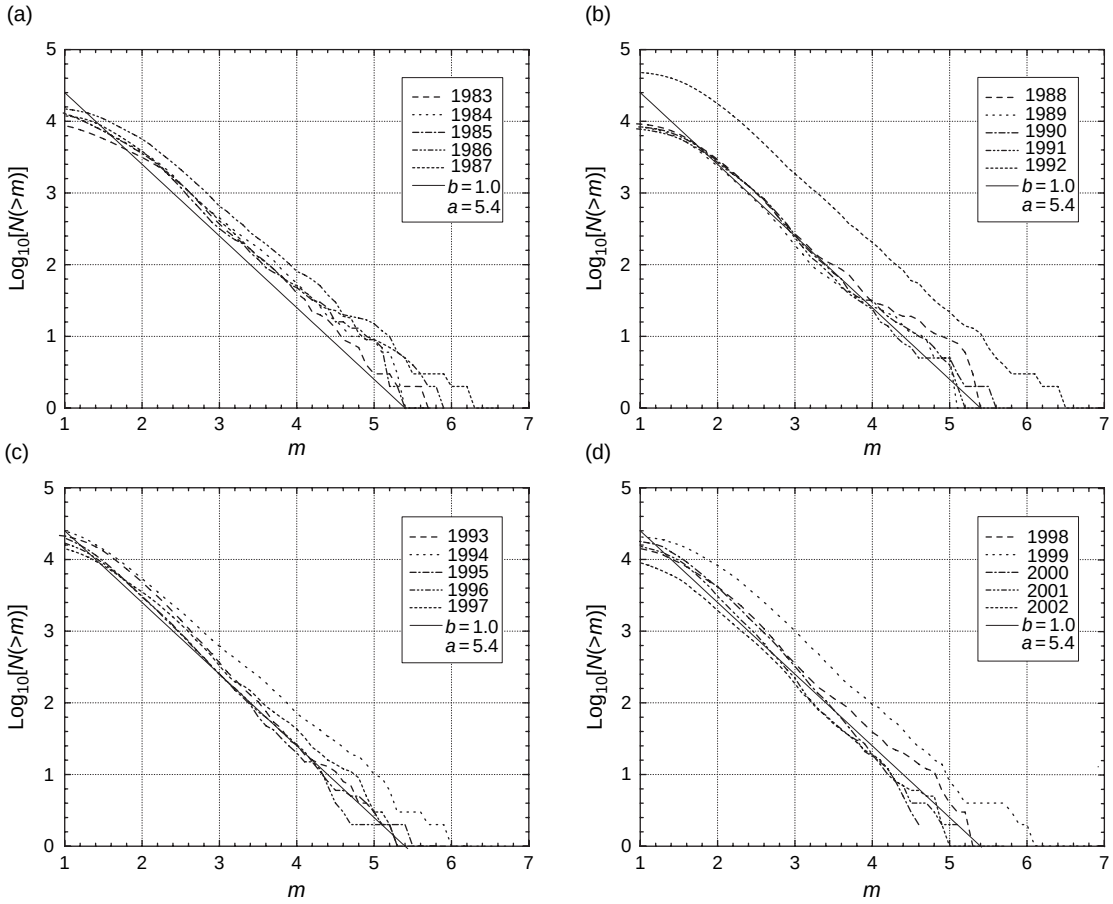


Figure 1 Cumulative number of earthquakes per year, N, occurring in southern California with magnitudes greater that m as a function of m. Twenty individual years are considered (SCSN catalog, http://www.data.scec.org): (a) 1983–87, (b) 1988–92, (c) 1993–97, (d) 1994–2002. The solid straight line in (a)–(c) is the Gutenberg–Richter relation [1] with $b = 1.0$ and $a = 5.4$. The larger number of earthquakes in 1987, 1992, 1994, and 1999 can be attributed to the aftershocks of the Whittier-Narrows, Landers, Northridge, and Hector Mine earthquakes, respectively. If aftershocks are excluded, the background seismicity in southern California is nearly uniform in time.

together results in the GR power-law relation [1] with $b = 1.0$ and $a = 5.4$, shown as the solid straight lines in **Figures 1(a)–1(d)**. In **Figure 1** there is generally good agreement between each individual year's data and the GR relation (solid straight line) for the period 1983–2003. The exceptions can be attributed to the aftershock sequences of the 1987 Whittier-Narrows, 1992 Landers, 1994 Northridge, and 1999 Hector Mine earthquakes.

For years without large numbers of aftershocks, the background seismicity in Southern California illustrated in **Figure 1** is nearly uniform from year to year, and is not a function of time. Small earthquakes behave like thermal background noise. There is observational evidence that the earth's crust is continuously on the brink of failure (Scholz, 1991). One example is induced seismicity. Whenever the crust is loaded, earthquakes are induced whether in a tectonically active area or not. Examples of nontectonic loading include the filling of a reservoir behind a newly completed dam or the high-pressure injection of fluids in a deep well.

Studies have shown that aftershocks also satisfy the GR relation [1] with 'b-values' that are not statistically different from the frequency–magnitude distribution of all earthquakes (Kisslinger, 1996). The Parkfield earthquake (28 Sep. 2004; $m_{\mathrm{ms}} = 6.0$; epicenter 35.818° N, −120.366° W) generated a well-documented aftershock sequence (Bakun *et al.*, 2005). We will consider as aftershocks all earthquakes that occurred in an elliptical region centered at 35.9° N and −120.5° W with the radii of sizes 0.4° and 0.15°

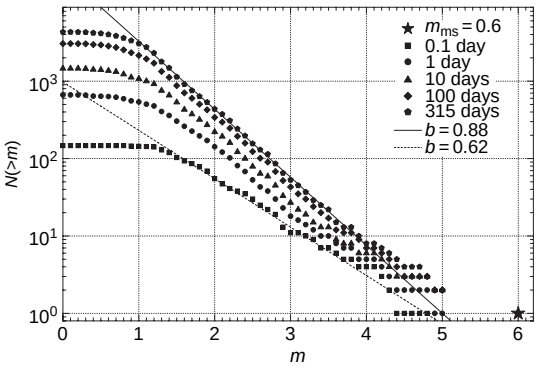


Figure 2 The frequency–magnitude distribution of aftershocks for several time intervals elapsed since the Parkfield main shock. The Gutenberg–Richter scaling relation [1] is fit to obtain the b-values for 0.1 day ($b = 0.62$) and 315 days ($b = 0.88$) after the main shock.

and oriented 137° NW during the 315-day period following the earthquake. The cumulative numbers of aftershocks with magnitudes greater than m, $N(\geq m)$, are given as a function of m in **Figure 2**. The fit of eqn [1] to the Parkfield data is obtained by taking $b = 0.88$ and $a = 4.4$.

Presumably aftershocks satisfy the GR scaling for the same reason that all earthquakes do. However, no generally accepted theory is available for the explanation of scale-invariant nature of this distribution. Generally, two end-member models have been proposed: (1) each fault has a GR distribution of earthquakes and (2) there is a power-law frequency–area distribution of faults and each fault has a reoccurrence of characteristic earthquakes. Observations generally favor the second explanation (Turcotte, 1997).

4.23.2.2 Båth's Law

Another important scaling law concerning aftershocks is Båth's law. This law states that the difference in magnitude between the main shock and its largest aftershock is approximately constant independent of the magnitude of the main shock (Båth, 1965). That is

$$\Delta m = m_{ms} - m_{as}^{max} \quad [2]$$

with m_{ms} the magnitude of the main shock, m_{as}^{max} the magnitude of the largest aftershock, and typically $\Delta m \approx 1.2$.

A number of extensive studies of the statistical variability of Δm have been carried out (Vere-Jones, 1969; Kisslinger and Jones, 1991; Tsapanos, 1990; Felzer *et al.*, 2002, 2003; Console *et al.*, 2003a; Helmstetter and Sornette, 2003a).

A modified version of Båth's law has been proposed by Shcherbakov and Turcotte (2004) and is based on an extrapolation of the GR statistics for aftershocks. The magnitude of the 'largest' aftershock consistent with GR scaling for aftershocks is obtained by formally setting $N(\geq m) = 1$ in [1] with the result

$$a = bm^* \quad [3]$$

where m^* is the inferred magnitude of the 'largest' aftershock for the given aftershock sequence. In general, this extrapolated value will differ from the mean value of the largest aftershock obtained by averaging over an ensemble of main shock–aftershock sequences having the same main shock magnitude. If Båth's law is applicable to the inferred values of m^* we can write

$$\Delta m^* = m_{ms} - m^* \quad [4]$$

where Δm^* is approximately a constant. Substitution of [3] and [4] into [1] gives

$$\log_{10} N(\geq m) = b(m^* - m) \quad [5]$$

with b, m_{ms}, and m^* specified; the frequency–magnitude distribution of aftershocks can be determined using [5].

The applicability of both forms of Båth's law is illustrated in **Figure 2**. For the Parkfield earthquake we have $m_{ms} = 6.0$ and $m_{as}^{max} = 5.0$, from the fit of the GR relation [1] to the data we obtained $a = 4.4$. From [3] we obtain $m^* = 5.0$. For this earthquake the magnitudes of the largest actual aftershock and largest inferred aftershock are equal. Shcherbakov and Turcotte (2004) studied 10 large earthquakes that occurred in California between 1987 and 2003 with magnitudes equal to or greater than $m_{ms} \geq 5.5$. The mean difference in magnitudes between these main shocks and their largest detected aftershocks was 1.16 with a standard deviation $\sigma = 0.46$ (Båth's law). The mean difference in magnitudes between these main shocks and their inferred 'largest' aftershocks was 1.11 with $\sigma = 0.29$ (modified Båth's law). The modified version of Båth's law is a better approximation for these earthquakes than the original one.

Shcherbakov and Turcotte (2004) explained the general applicability of the modified form of Båth's law in terms of energy partitioning. The applicability of the law implies that the average ratio of the total energy radiated in an aftershock sequence to the

energy radiated by the main shock E_{as} is approximately a constant given by

$$\frac{E_{as}}{E_{ms}} = \frac{2b}{3-2b} 10^{-(3/2)\Delta m^*} \quad [6]$$

Taking $\Delta m^* = 1.0$ and $b = 0.88$ for the Parkfield earthquake we find from [6] that $E_{as}/E_{ms} = 0.045$.

4.23.2.3 Omori's Law

The temporal decay of aftershock activity following a main shock also obeys a universal scaling law to a good approximation. This is known as the modified Omori's law and can be written in the form (Utsu *et al.*, 1995; Scholz, 2002)

$$r(t) \equiv \frac{dN}{dt} = \frac{1}{\tau[1 + t/c]^p} \quad [7]$$

where $r(t) \equiv dN/dt$ is the rate of occurrence of aftershocks with magnitudes greater than m, t is the time that has elapsed since the main shock, τ and c are characteristic times, and the exponent p has a value near unity. The total number of aftershocks with magnitudes greater than m, $N_{tot}(\geq m)$, is obtained by integrating [7] with the result (Shcherbakov *et al.*, 2004)

$$\begin{aligned} N_{tot}(\geq m) &= \int_0^\infty r(t)\ dt \\ &= \int_0^\infty \frac{dt}{\tau(1 + t/c)^p} = \frac{c}{\tau(p-1)} \end{aligned} \quad [8]$$

This is only valid for $p > 1$. Combining [7] and [8] gives

$$\frac{1}{N_{tot}} \frac{dN}{dt} = \frac{p-1}{c} \frac{1}{(1 + t/c)^p} \quad [9]$$

Substitution of [8] into [5] gives

$$\frac{c}{\tau} = (p-1)10^{b(m^* - m)} \quad [10]$$

This is an important relationship between the two characteristic times c and τ that appear in the modified Omori's law [7].

In general both c and τ can be functions of the lower cutoff magnitude m, $c(m)$ and $\tau(m)$. To preserve GR scaling we assume

$$c(m) = c^* 10^{\gamma(m^* - m)} \quad [11]$$

$$\tau(m) = \tau^* 10^{-\alpha(m^* - m)} \quad [12]$$

where c^* and τ^* correspond to $c(m^*)$ and $\tau(m^*)$, respectively and γ and α are obtained by fitting the data. Substitution of [11] and [12] into [10] gives the requirements that

$$c^* = \tau^*(p-1) \quad [13]$$

$$b = \gamma + \alpha \quad [14]$$

We next consider the rates of aftershock occurrence at times $t \gg c$. In this limit [7] becomes

$$r(t \gg c,\ m) = \frac{1}{\tau(m)} \left[\frac{c(m)}{t}\right]^p \quad [15]$$

We further assume that these asymptotic rates satisfy GR scaling given by

$$r(t \gg c,\ m) = r^* 10^{\beta(m^* - m)} \quad [16]$$

The quantity β is a b-value for aftershocks at large times. Substitution of [11], [12], and [16] into [15] gives the requirements that

$$r^* = \frac{1}{\tau^*} \left(\frac{c^*}{t}\right)^p \quad [17]$$

$$\beta = \alpha + p\gamma \quad [18]$$

Combining [13] and [18] gives

$$\beta = b + (p-1)\gamma \quad [19]$$

In general, the b-value for all aftershocks is not equal to the b-value for aftershocks at times $t \gg c$ after the main shock, that is, $\beta \neq b$. A relationship between scaling exponents such as that given in [13] is often found near critical points (Ma, 1976; Stauffer and Aharony, 1992).

There are two limiting cases for the behavior of the modified Omori's law. The first is to assume that the characteristic time c is a constant c_0, independent of the lower magnitude cutoff m of aftershocks that are considered. In this case from [11] we have $\gamma = 0$ and from [14] and [19] we have

$$b = \beta = \alpha \quad [20]$$

In this case the two b-values, b and β, are equal. Also in this limit the substitution of [10] into [7] gives

$$\frac{dN}{dt} = \frac{(p-1)10^{b(m^* - m)}}{c_0[1 + t/c_0]^p} \quad [21]$$

Similar results have been derived by Reasenberg and Jones (1989) and by Yamanaka and Shimazaki (1990).

In the second case, it is assumed that the characteristic time τ is a constant $\tau = \tau_0$, independent of the

lower magnitude cutoff m. In this case from [12] we have $\alpha = 0$ and from [14] and [19] we obtain $\gamma = b$ and

$$\beta = pb \qquad [22]$$

As a representative example we consider the rate of occurrence of aftershocks following the Parkfield earthquakes. The numbers of aftershocks per unit time $r \equiv dN/dt$ are given in **Figure 3** as a function of the time t after the main shocks for several aftershock magnitude cutoffs. In **Figure 3(a)** Omori's law [7] is fit to the data assuming that both c and τ are functions of m as given by [11] and [12] with $b = 0.88$, $m^* = 5.0$, $p = 1.09$, $\gamma = 0.76$, $\alpha = 0.12$, $\tau^* = 115$ s, and $c^* = 10.26$ s. The corresponding values of $c(m)$ and $\tau(m)$ are given in **Figure 3(a)**. It is seen that $c(m)$ is a strong function of m, whereas $\tau(m)$ has only a weak dependence on m. From [19] we obtain $\beta = 0.95$, this compares with $b = 0.88$. Thus β, the b-value for aftershocks at large times, is 8% greater than the b-value for the whole sequence.

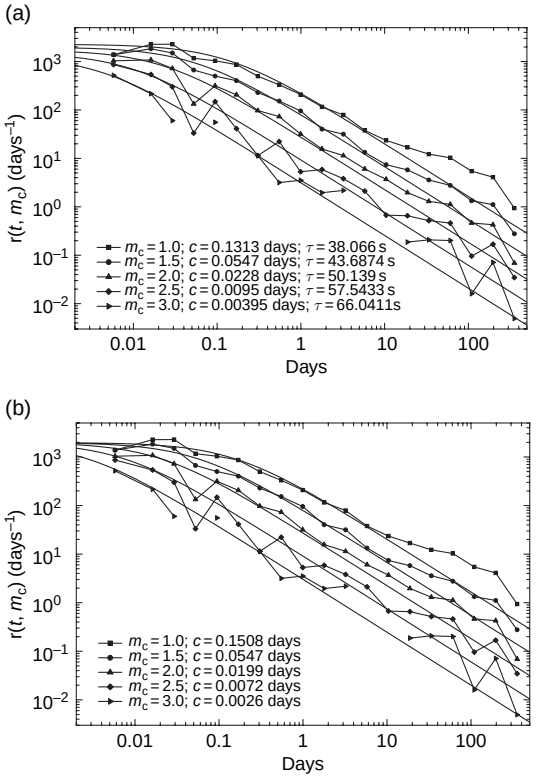


Figure 3 The rates of occurrence of aftershocks with magnitudes greater than m_c in numbers per day, $r(t,m_c)$, are given for 315 days after the main shock. Lower magnitude cutoffs were taken to be $m_c = 1.0$, 1.5, 2.0, 2.5, and 3.0. (a) The value of $c(m_c = 2.5)$ has been obtained from the best fit. The rest of c's were calculated using assuming $\beta = 0.96$. (b) It is assumed that $\tau = 43.7095$ s is a constant for all magnitude cutoffs.

We next fit the modified Omori's law [7] to the aftershock data for the Parkfield earthquake assuming that τ is constant, $\tau = \tau_0$. This requires that $\alpha = 0$ in [12] and the results are given in **Figure 3(b)** for the same values of b, m^*, and p and $\gamma = 0.88$, $\tau_0 = 43.7$ s, and $c^* = 3.94$ s. From [19] we find again that $\beta = 0.96$. In both cases considered, **Figures 3(a)** and **3(b)**, the fit of the modified Omori's law to the data is quite good and a large range of values of $c(m)$ is required. It is not possible to fit the data with a constant value of c. For the fit in **Figure 3(b)** we have used constant τ. This implies that there are no small earthquakes at short times. We note here that the catalog is not complete in the vicinity of the main shock especially for small events. However, our form of generalized Omori's law allows us to recover the decay rate for small times after the main shock. Some authors have argued that the entire rollover of aftershock rates at small times is an artifact of under-recording of early aftershocks (Kagan, 2004). However, our fit of the modified Omori's law to the data can be accomplished using only the power-law decay curves. Confirmation of the validity of modified Omori's law comes from systematic differences between $b = 0.88$ obtained from **Figure 2** and $\beta = 0.96$ and $\beta/b = 1.09$.

Results for eight earthquakes with well-documented aftershock sequences are given in **Table 1**. In terms of magnitude they range from the $m = 9.3$ Sumatra earthquake to the $m = 6.0$ Parkfield earthquake. The b-values range from 0.78 to 1.20, the β-values range from 0.90 to 1.32. Values of Δm^* for the modified Båth's law range from 0.75 to 1.6. Values of the power-law exponent p in the modified Omori's law range from 1.09 to 1.22. Values of τ_0 range from 36.3 to 155.0 s and seem to increase somewhat with earthquake magnitude.

We next consider the implications of the hypothesis that $\tau = \tau_0$ independent of the aftershock cutoff magnitude m. One implication of this hypothesis is that the GR frequency–magnitude scaling of aftershocks is not established instantaneously at the time of the main shock. At short times after the main shock with $t/c \ll 1$, [7] reduces to

$$r \equiv \frac{dN}{dt} = \frac{1}{\tau_0} \qquad [23]$$

Thus τ_0 is the mean time delay before the first aftershock, also there are very few small aftershocks at

Table 1 Results for eight earthquakes

Earthquake	m_{ms}	b	Δm^*	p	β	β/p	τ_0, (s)
Landers	7.3	0.98	1.10	1.22	1.20	1.22	107.3
Northridge	6.7	0.91	0.75	1.17	1.06	1.16	53.1
Hector Mine	7.1	1.02	1.35	1.22	1.23	1.20	83.7
San Simeon	6.5	1.0	1.10	1.12	1.09	1.09	50.2
Parkfield	6.0	0.88	1.0	1.09	0.96	1.09	43.7
Sumatra	9.3	1.20	1.6	1.10	1.32	1.10	155.5
Kobe	7.3	0.78	1.1	1.16	0.90	1.15	44.1
Tottori	7.3	0.92	1.6	1.16	1.02	1.11	36.3

early times. At long times after the main shock with $t/c \gg 1$, from [7] we obtain

$$r \equiv \frac{dN}{dt} = \frac{1}{\tau_0}\left[\frac{c(m)}{t}\right]^p \qquad [24]$$

This is the power-law decay of aftershock activity. With $\tau = \tau_0$ the dependence of $c(m)$ on m is obtained from [10] with the result

$$c(m) = \tau_0(p-1)\left(\frac{10^{m^*}}{10^m}\right)^b \qquad [25]$$

The equilibration time t_e for the establishment of GR distribution of aftershock magnitudes can be approximated by the time $t = t_e$ when the aftershock decay rates from [23] and [24] are equal. The result is

$$t_e = c(m) \qquad [26]$$

The equilibration time t_e corresponds to the characteristic time $c(m)$. Since $c(m)$ is a function of the cutoff magnitude m, so is the equilibration time $t_e(m)$.

We see from **Figure 3** that $c(m)$ systematically increases with decreasing lower cutoff magnitudes m. Thus the equilibration time t_e also increases for the smaller magnitude aftershock cutoffs. The GR distribution is first established for the larger aftershocks and is subsequently established for smaller aftershocks.

The moment M of an aftershock can be related to its magnitude m by

$$\log_{10} M = d_1 m + d_2 \qquad [27]$$

where d_1 and d_2 are constants. Kanamori and Anderson (1975) have established a theoretical basis for taking $d_1 = 3/2$. Substitution of [27] into [25] gives

$$c(m) = \tau_0(p-1)\left(\frac{M^*}{M}\right)^{2b/3} \qquad [28]$$

where M^* is the moment of the 'inferred' largest aftershock with magnitude m^* and M is the moment of the cutoff magnitude m earthquake. Kanamori and Anderson (1975) have also shown that it is a good approximation to relate the moment M of an earthquake to the area A_r of the earthquake rupture by

$$M = \eta A^{3/2} \qquad [29]$$

where η is a constant. This result is valid for earthquakes with rupture dimensions less than the seismogenic depth. Substitution of [29] into [28] with [27] gives

$$t_e(A_r) = c(A_r) = \tau_0(p-1)\left(\frac{A_r^*}{A_r}\right)^b \qquad [30]$$

where A_r^* is the rupture area of the 'inferred' largest aftershock with magnitude m^* and A_r is the rupture area of the cutoff magnitude m earthquake.

Setting $A_r = A_r^*$ in [30] we have $t_e(m^*) = (p-1)\tau_0$. Since p is generally close to but greater than one, the equilibration time of the largest aftershock is shorter than the time delay τ_0 until the largest aftershock occurs. Thus there is a time delay t_e for the establishment of the GR distribution only for small aftershocks.

We have shown that the time delays associated with aftershocks are well approximated by the modified Omori's law, but why do the delays occur? When an earthquake occurs, there are regions where the stress is increased. This increase in stress is the fundamental cause of aftershocks. However, the systematic time delay before the occurrence of aftershocks requires an explanation. Das and Scholz (1981), Main (2000), and Ojala *et al.* (2003) have attributed this delay to stress corrosion combined with a critical stress intensity factor. Shaw (1993) has utilized a phenomenological approach to the dynamics of subcritical crack growth. A time

delay is implicit in the empirically derived rate and state friction law. Dieterich (1994) has related the power-law decrease in aftershock activity to this law. There appears to be fundamental similarities between aftershock delays and the nucleation of bubbles in a superheated liquid. These similarities led Rundle (1989), Rundle and Klein (1993), and Rundle *et al.* (1999) to relate aftershock sequences to the power-law scaling in the vicinity of a spinodal line. This association is also supported by the relationship between the three-dimensional (3-D) spatial distributions of aftershocks and the 'backbone' of a 3-D percolation cluster given by Robertson *et al.* (1995).

4.23.2.4 Accelerated Moment Release

There is accumulating evidence that there may be an increase in the number of intermediate-sized earthquakes prior to a large earthquake. The occurrence of a relatively large number of intermediate-sized earthquakes in northern California prior to the 1906 San Francisco earthquake was noted by Sykes and Jaumé (1990). Other studies have been carried out by Knopoff (1996) and by Jaumé and Sykes (1999).

It has also been observed that an increase in the seismic activity prior to large earthquakes takes the form of a power law. This was first proposed by Bufe and Varnes (1993). They considered the cumulative Benioff strain in a region defined as

$$\varepsilon_{\mathrm{B}}(t) = \sum_{i=1}^{N(t)} \sqrt{e_i} \qquad [31]$$

where e_i is the seismic energy release in the ith precursory earthquake and $N(t)$ is the number of precursory earthquakes considered up to time t. It should be emphasized that the Benioff strain defined to be the square root of the seismic energy release is not an actual strain.

The precursory increase in seismicity is referred to as AMR. In terms of the cumulative Benioff strain it is quantified as

$$\varepsilon_{\mathrm{B}}(t) = \varepsilon_{\mathrm{B}}(t_{\mathrm{f}}) - B\left(1 - \frac{t}{t_{\mathrm{f}}}\right)^{s} \qquad [32]$$

where $\varepsilon_{\mathrm{B}}(t_{\mathrm{f}})$ is the final cumulative Benioff strain at the time t_{f} when the large earthquake occurs, t is time measured forward from the beginning of AMR, B is a constant, and s is the exponent. Again it should be emphasized that the cumulative Benioff strain is not proportional to cumulative moment release since the earthquake moment is proportional to the earthquake energy release.

Examples of AMR have been given by Bowman *et al.* (1998), Bufe *et al.* (1994), Varnes and Bufe (1996), Brehm and Braile (1998, 1999b, 1999a), Robinson (2000), Zöller *et al.* (2001), Main (1999), Bowman and King (2001), Yang *et al.* (2001), King and Bowman (2003), Bowman and Sammis (2004), and Sammis *et al.* (2004). Rundle *et al.* (2000) found that the distribution of values of the power-law exponent s for 12 earthquakes was $s = 0.26 \pm 0.15$. A systematic study of the optimal spatial region and magnitude range to obtain the power-law seismic activation has been carried out by Bowman *et al.* (1998). Four examples of their results are given in **Figure 4**, where the cumulative Benioff strain $\varepsilon_{\mathrm{B}}(t)$ has been correlated (solid line) with [32]. Clear increases in seismic activity prior to these earthquakes are illustrated.

In obtaining the data given in **Figure 4**, Bowman *et al.* (1998) investigated circular regions about each of the characteristic earthquakes they considered. They determined the optimal radius for activation. Sornette and Sammis (1995) associated this precursory activation with the approach to a critical phase transition. Saleur *et al.* (1996a, 1996b) interpreted this behavior in terms of correlation lengths that increase prior to the characteristic earthquake. Thus we will refer to this radius of the optimal activation region as an activation correlation length (ACL).

Values of ACL as obtained by several authors are given in **Figure 5** as a function of earthquake magnitude and the characteristic rupture length. Although there is considerable scatter, there is a clear increase in the ACL with increasing earthquake magnitude. Also included for comparison is the relation

$$\xi = 10A_{\mathrm{r}}^{1/2} \qquad [33]$$

Many of the ACLs are greater than 10 times the characteristic rupture length.

Dobrovolsky *et al.* (1979) and Keilis-Borok and Kossobokov (1990) reported a similar scaling for the maximum distance between an earthquake and its precursors using pattern recognition techniques. Large ACLs are also suggested by the remotely triggered aftershocks following the Landers, California, earthquake (Hill *et al.*, 1993).

Zöller *et al.* (2001) directly estimated the ACL from earthquake catalogs using single-link cluster analysis (Frohlich and Davis, 1990). They studied 11 earthquakes in California with $m \geq 6.5$ since 1952

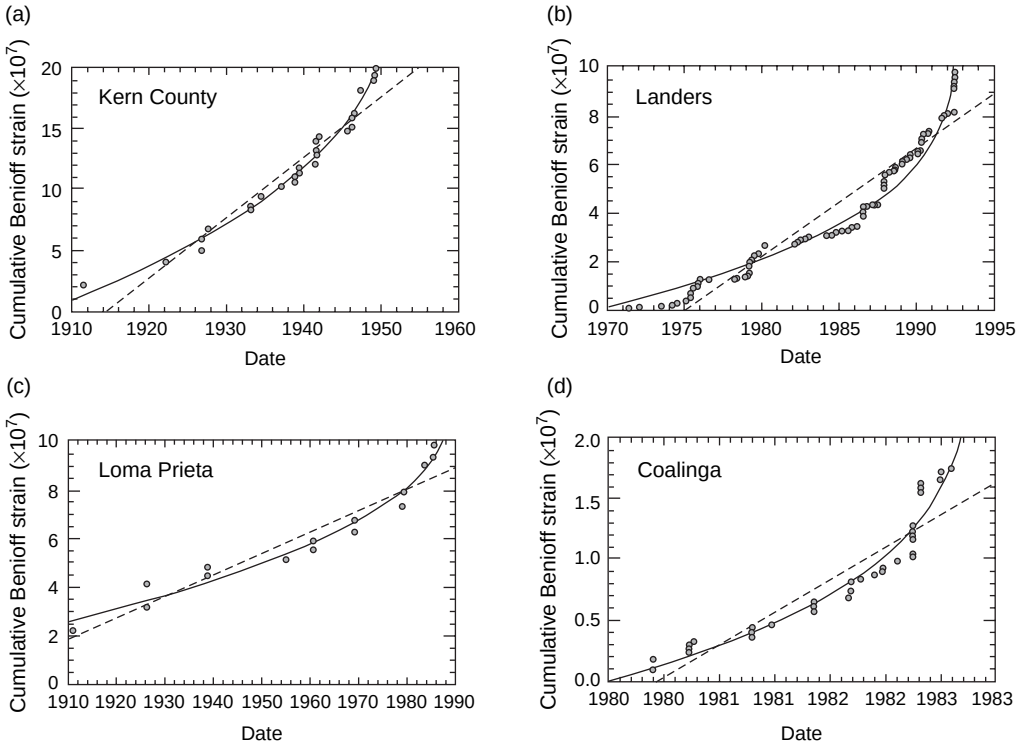


Figure 4 Data points are cumulative Benioff strains $\varepsilon_B(t)$ determined from [31] prior to four major earthquakes in California (Bowman *et al.*, 1998). Clear increases in seismic activity prior to the 1952 Kern County, 1989 Loma Prieta, 1992 Landers, and 1983 Coalinga earthquakes are illustrated. In each of the four examples the data have been correlated (solid lines) with the power-law relation given in [32]. The values of the power-law exponent s used in [32] were 0.30 for Kern County, 0.18 for Landers, 0.28 for Loma Prieta, and 0.18 for Coalinga. Dashed straight lines represent a best-fit constant rate of seismicity.

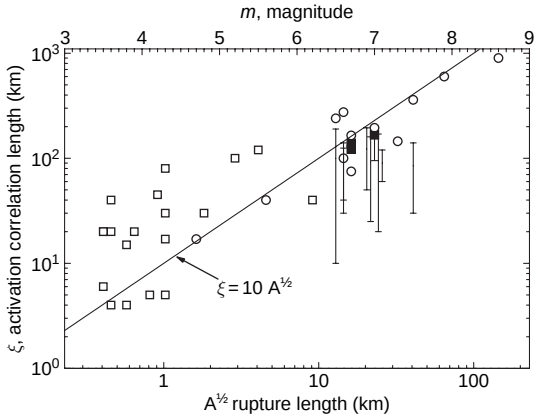


Figure 5 Correlation lengths ξ for precursory seismic activation are given as a function of the square-root of the rupture area and the earthquake magnitude m. The open circles are the values given by Bowman *et al.* (1998), the open squares are values obtained for earthquakes in the New Madrid seismic zone by Brehm and Braile (1998), the error bars are limits obtained for earthquakes in the western United States by Brehm and Braile (1999a), and the solid squares are values obtained for earthquakes in New Zealand by Robinson (2000).

and found systematic increases in the correlation lengths prior to most of these earthquakes. Goltz and Bose (2002) have studied precursory seismic activation using configurational entropy.

Although there is considerable observational evidence for AMR, this phenomenon is certainly not universal for all earthquakes and systematic attempts to use AMR for earthquake predictions have not been successful. Critical evaluations of AMR have been given by Gross and Rundle (1998), Jaumé (2000), Vere-Jones *et al.* (2001), Zöller *et al.* (2001), Zöller and Hainzl (2001), Ben-Zion and Lyakhovsky (2002), and Robinson *et al.* (2005). The concept of ACL is also controversial. Huc and Main (2003) and McKernon and Main (2005) have systematically looked for correlated seismicity in earthquake catalogs and have not found results consistent with ACL. There is also the question of the physical mechanisms associated with ACL. The lengths required are certainly greater than classic elastic length scales. However, Felzer and Brodsky (2006) have found

that aftershocks systematically extend beyond elastic length scales. As a result, ACL may be associated with complex interactions between faults.

4.23.2.5 Recurrence Times

A property of seismicity that is important in seismic hazard assessment is the statistical distribution of recurrence times of characteristic earthquakes. Examples of characteristic earthquakes include great earthquakes on the northern section of the San Andreas Fault (the 1906 San Francisco earthquake), the southern section of the San Andreas Fault (the great 1857 earthquake), and the Parkfield section of the San Andreas Fault.

Characteristic earthquakes are associated with quasi-periodicity but can also have considerable variability. A measure of the variability of recurrence times on a fault or fault segment is the coefficient of variation C_v (the ratio of the standard deviation σ to the mean μ). For strictly periodic earthquakes on a fault or fault segment, we would have $\sigma = C_v = 0$. For the random (i.e., exponential with no memory) distribution of recurrence times, we would have $C_v = 1$, (i.e., $\sigma = \mu$). Ellsworth *et al.* (1999) analyzed 37 series of recurrent earthquakes and suggested a provisional generic value of the coefficient of variation $C_v \approx 0.5$. A number of alternative distributions have been proposed for this purpose. These include the exponential (Poisson), the log-normal, Brownian passage-time (inverse Gaussian), and Weibull distributions (Davis *et al.*, 1989; Sornette and Knopoff, 1997; Ogata, 1999a; Matthews *et al.*, 2002). We will primarily consider the Weibull distribution. The cumulative distribution function (cdf) for this distribution is given by

$$P(t) = 1 - \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right] \qquad [34]$$

where $P(t)$ is the fraction of the recurrence times that are shorter than t, and β and τ are fitting parameters. If $\beta = 1$ this is the Poisson (random) distribution.

Probably the best-studied sequence of characteristic earthquakes is the sequence that occurred on the Parkfield, California section of the San Andreas Fault between 1857 and 2004 (Bakun *et al.*, 2005). This is because the slip rate is relatively high ($\approx$30 mm yr^{-1}) and the earthquakes are relatively small ($m \approx 6.0$), thus the recurrence times are relatively short ($\approx$25 yr).

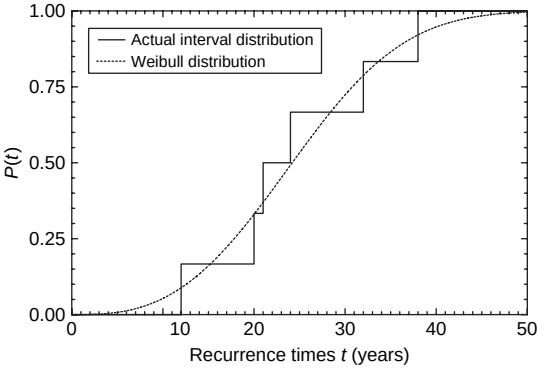


Figure 6 Cumulative distribution of recurrence times of Parkfield earthquakes. The discontinuous line is the distribution of the six actual recurrence times. The continuous line is the best-fit Weibull distribution ($\mu = 24.5$ yr, $\sigma = 9.25$ yr, $C_v = 0.378$, $\tau = 27.5$ yr, and $\beta = 2.88$).

Slip on the Parkfield section of the San Andreas Fault occurred during $m \approx 6$ earthquakes that occurred in 1857, 1881, 1901, 1922, 1934, 1966, and 2004. The cumulative probability distribution (cdf) of Parkfield recurrence times ($t = 12$, 20, 21, 24, 32, 38 years) is given in **Figure 6**. The mean, standard deviation, and coefficient of variation of these recurrence times are $\mu = 24.5$ years, $\sigma = 9.25$ years, and $C_v = \sigma/\mu = 0.378$, respectively. Taking these values, the corresponding fitting parameters for the Weibull distribution are $\tau = 27.49$ years and $\beta = 2.88$. Using these values, the cdf from [34] is also shown in **Figure 6**. Quite good agreement is observed.

Another important question is whether there is a scaling of seismicity that requires the validity of the Weibull distribution. The Weibull distribution is widely used in engineering to model the statistical distribution of failure times. Its applicability has been demonstrated by many actual tests (Meeker and Escobar, 1998).

It should be also pointed out that the stretched exponential distribution (a Weibull distribution with $0 < \beta < 1$) is found to be widely applicable in heterogeneous and homogeneous nucleation. In this context, its applicability is known as Avrami's law (Avrami, 1940). The similarities between nucleation problems (i.e., droplet formation in supercooled steam) and the nucleation of earthquakes have been discussed previously (Rundle *et al.*, 2003). In Avrami's law, however, β is found to be less than 1.

The sequence of recurrence times on a fault is a time series. Bunde *et al.* (2003), Bunde *et al.* (2004), Altmann *et al.* (2004), Altmann and Kantz

(2005), and Sanchez *et al.* (2002) have shown that time series that exhibit long-range correlations have recurrence time statistics that satisfy stretched exponential distributions. Studies have been carried out of the distribution of interval times between peaks over threshold for fractional Gaussian noises. These noises have a power-law dependence of the two-point correlation function C on the delay time s, $C(s) \approx s^{-\gamma}$. It was shown that the cumulative distribution of interval times satisfied the stretched exponential given by [34] with $\beta = \gamma$. For fractional noises $0 < \gamma < 1$ so that the distribution with $\beta < 1$ has a fat tail. Bunde *et al.* (2003) argue against distributions in which the extreme event statistics become Poisson, that is, like the Brownian passage time distribution.

A number of authors have related the spatial distribution of earthquakes to scale-free networks (Abe and Suzuki, 2003, 2004a, 2004b; Baiesi and Paczuski, 2004, 2005; Console and Murru, 2001; Console *et al.*, 2003b).

4.23.2.6 Interoccurrence Times

Another property of seismicity that has been shown to satisfy a universal scaling law is the statistical distribution of interoccurrence times. All earthquakes in a specified region and specified time window T with magnitudes greater than m are considered to be point events. Based on studies of properties of California seismicity, a unified scaling law for the temporal distribution of earthquakes was proposed by Bak *et al.* (2002). The distribution of interoccurrence times between successive earthquakes was obtained by using as scaling parameters both a grid size over which the region was subdivided, and a lower magnitude cutoff. Two distinct scaling regimes were found. For short times, aftershocks dominate the scaling properties of the distribution, decaying according to the modified Omori's law. For long times, an exponential scaling regime was found that can be associated with the occurrence of main shocks. To take into account the spatial heterogeneity of seismic activity, it has been argued that the second regime is not an exponential but another power law (Corral, 2003). An alternative approach to describe a unified scaling of earthquake recurrence times was suggested by Corral (2004a, 2004b, 2005a, 2005b), where the distributions computed for different spatial areas and magnitude ranges were rescaled with the rate of seismic activity for each area considered. An analysis of the change in behavior between these two regimes based on a nonstationary Poisson sequence of events was carried out by Lindman *et al.* (2005). Related studies have been carried out by Carbone *et al.* (2005), Davidsen and Goltz (2004), and Livina *et al.* (2005a, 2005b).

We illustrate this scaling law using the interoccurrence time distribution of Parkfield aftershocks. We define interoccurrence times between successive aftershocks as $\Delta t_i = t_i - t_{i-1}$, $i = 1, 2, \ldots$ and study their statistical distribution over a finite time interval T; we use a lower magnitude cutoff m as a scaling parameter and study sequences with different m's. The normalized distribution (histogram) of interoccurrence times is given in **Figure 7**. The distributions with lower-magnitude cutoffs have a shorter power-law scaling regime and start to roll over more quickly. This can be explained by the presence of more events in the sequences with lower-magnitude cutoffs and as a result the shortening of the mean time intervals between events. Another scaling parameter which affects the roll over is the time interval T following the main hock. An increase in T leads to the occurrence of longer time intervals Δt between events. In our analysis of aftershock interoccurrence times, we have used a time interval of $T = 90$ days to try to eliminate the contribution from the background seismicity which starts to influence the sequence in later times.

Shcherbakov *et al.* (2005b) showed that the observed distribution of interoccurrence times can be well approximated by a distribution derived from assuming that aftershocks follow

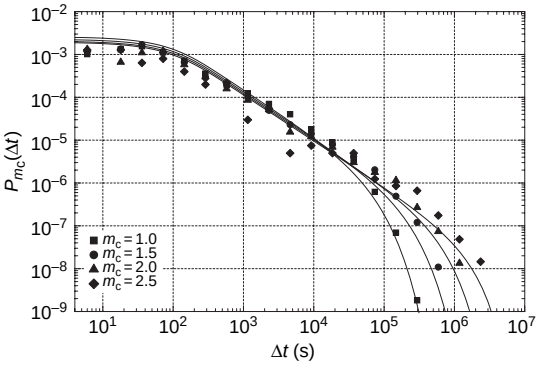


Figure 7 Interoccurrence time distributions for the Parkfield aftershock sequence for different magnitude cutoffs $m_c = 1.0$, 1.5, 2.0, and 2.5. A time interval of $T = 90$ days following the main shock has been used. The solid lines have been computed using eqn [35].

nonhomogeneous Poisson statistics with the rate given by the modified Omori's law [7] over a finite time interval T

$$P_T(\Delta t) = \frac{1}{N}\left[\int_0^{T-\Delta t} r(s)r(s+\Delta t) \times \mathrm{e}^{-\int_s^{s+\Delta t} r(u)\mathrm{d}u}\mathrm{d}s + r(\Delta t)\mathrm{e}^{-\int_0^{\Delta t} r(u)\mathrm{d}u}\right] \quad [35]$$

where $N = \int_0^T r(u)\mathrm{d}u$ is the total number of events during a time period T. The plots of this equation for different magnitude cutoffs m are also given in **Figure 7** as solid curves. Davidsen and Paczuski (2005) have also found a scaling for the distances between successive earthquake distribution in a region.

4.23.3 Self-Organizing Complex Systems

4.23.3.1 Fractality

Mandelbrot (1967) introduced the concept of fractality in terms of the length of a rocky coastline. Fractality implies power-law scaling and self-similarity. The basic definition of a fractal set is given by Turcotte (1997)

$$N_i = \frac{C}{r_i^D} \quad [36]$$

where N_i is the number of objects with a characteristic linear dimension r_i, C is a constant of proportionality, and D is the fractal dimension.

Aki (1981) showed that the empirical GR frequency–magnitude scaling given in [1] can be interpreted as a fractal relation. He used the empirical relation between earthquake magnitude m and the earthquake rupture area A_r obtained from [27] and [29] to write

$$\log\left(\alpha A_\mathrm{r}^{3/2}\right) = \frac{3}{2}m + d \quad [37]$$

where α and d are constants. Combining [1] and [37] gives

$$N = \frac{C'}{A_\mathrm{r}^b} \quad [38]$$

Comparing [36] and [38] we can conclude that

$$D = 2b \quad [39]$$

It must be emphasized that N is the cumulative number of earthquakes with characteristic linear dimension $A_\mathrm{r}^{1/2}$ greater than the value specified whereas N_i in [36] is noncumulative. It should also be emphasized that the integral of [38] from 0 to ∞ diverges, cutoffs must be specified. For any natural process such as earthquakes upper and lower cutoffs are certainly appropriate, for example the validity of [38] is certainly greater than the atomic scale and less than the planetary scale.

4.23.3.2 Deterministic Chaos

Lorenz (1963) discovered deterministic chaos. He derived three coupled total differential equations as an approximation for thermal convection in a fluid layer heated from below. He showed that solutions in a parameter range had an exponentially diverging sensitivity to initial conditions. This divergent behavior is a characteristic of chaotic behavior. The chaotic behavior of these equations has been taken as evidence that the behavior of the atmosphere and oceans is chaotic.

Burridge and Knopoff (1967) introduced the slider-block model as a simple model for the behavior of a fault. One can show that a pair of connected slider blocks can exhibit deterministic chaos (Huang and Turcotte, 1990). Consider the behavior of a pair of slider blocks, a constant-velocity driver plate drags the blocks over a surface at a mean velocity, v. The two blocks are coupled to each other by a spring with spring constant k_c and to a driver plate by springs with spring constant k_L. Other model parameters are the block masses m and the coefficients of static and dynamic friction μ_s and μ_d. Solutions are governed by two parameters, the stiffness of the system $k_\mathrm{c}/k_\mathrm{L}$ and the ratio of static to dynamic friction coefficients $\mu_\mathrm{s}/\mu_\mathrm{d}$. Classic chaotic behavior is obtained for a wide range of parameter values as long as the symmetry is broken.

Slider-block models are a simple analogy to the behavior of faults in the Earth's crust. An important question is whether the chaotic behavior of a pair of slider blocks implies that seismicity is also an example of chaotic behavior. This question will be addressed in the next section. If seismicity is chaotic it implies that the prediction of earthquakes is not possible in a deterministic sense. Only a probabilistic approach to the occurrence of earthquakes will be possible (Geller, 1997; Geller *et al.*, 1997).

4.23.3.3 Criticality and Self-Organized Criticality

The concept of self-organized criticality (Turcotte, 1997, 1999b; Jensen, 1998) evolved from the 'sandpile' model proposed by Bak *et al.* (1988). In this model, there is a square grid of boxes, and at each time step a particle is dropped into a randomly selected box. When a box accumulates four particles, they are redistributed to the four adjacent boxes, or in the case of edge boxes they are lost from the grid. Since only nearest-neighbor boxes are involved in the redistribution, this is a cellular automata model. Redistributions can lead to further instabilities and avalanches of particles in which many particles may be lost from the edges of the grid. This model was called a 'sandpile' model because of the resemblance to an actual sandpile. The randomly dropped particles in the model are analogous to the addition of sand particles to a sandpile. The model avalanches are analogous to sand avalanches down the sides of the sandpile. Extensive analytical and numerical studies of the 'sandpile' model were carried out by Kadanoff *et al.* (1989); Dhar (1990); Priezzhev *et al.* (1996); Ivashkevich and Priezzhev (1998); Dhar (1999). It was found that the noncumulative frequency–area distribution of avalanches to have a power-law distribution with a slope near 5/4 (Priezzhev *et al.*, 1996). A modification of the sandpile model that utilizes a fractal distribution of box sizes reproduces GR statistics for aftershocks (Barriere and Turcotte, 1991, 1994; Henderson *et al.*, 1994; Huang *et al.*, 1998).

A second example of 'self-organized criticality' is the behavior of large arrays of slider blocks. The two-block model was considered in the last section. The standard multiple slider-block model consists of a square array of slider blocks as illustrated in **Figure 8**. Each block with mass m is attached to the loader plate with a loader spring, spring constant k_L. Adjacent blocks are attached to each other with a connector spring, spring constant k_c. The constant-velocity driver plate is pulled forward with velocity v. A block remains stationary as long as the net force on the block is less than the static resisting force. Again, two parameters determine the behavior of the system, the ratio of static and dynamic coefficients of friction, μ_s/μ_d and the stiffness of the system, k_c/k_L.

Carlson and Langer (1989) considered long linear arrays of slider blocks with each block connected by springs to the two neighboring blocks and to a constant-velocity driver. They used a velocity-weakening friction law and considered up to 400 blocks. Slip events involving large numbers of blocks were observed, the motions of all blocks involved in a slip event were coupled, and the applicable equations of motion had to be solved simultaneously. The system was completely deterministic. Frequency–area statistics were obtained for slip events. The events fell into two groups: smaller events obeyed a power-law (fractal) relationship with a slope near unity, but there were an anomalously large number of large events that included all the slider blocks. The observed behavior was characteristic of self-organized criticality. The motion of the driver plate is the steady input. The slip events are the 'avalanches' with a power-law (fractal) frequency–size distribution.

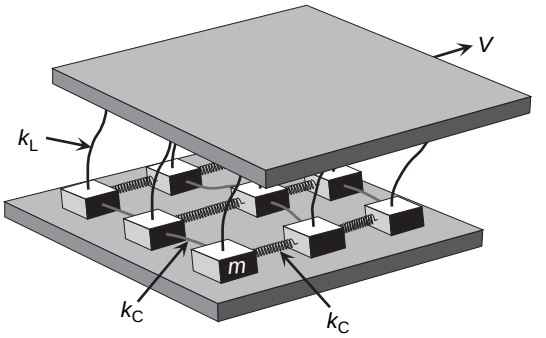


Figure 8 Illustration of the 2-D slider-block model. An array of blocks, each with mass m, is pulled across a surface by a driver plate at a constant velocity, V. Each block is coupled to the adjacent blocks with either leaf or coil springs (spring constant k_c), and to the driver plate with leaf springs (spring constant k_L).

Nakanishi (1990) studied multiple slider-block models using the cellular automata approach. A linear array of slider blocks was considered but only one block was allowed to move in a slip event. The slip of one block could lead to the instability of either or both of the adjacent blocks, which would then be allowed to slip in a subsequent step or steps, until all blocks were again stable. Brown *et al.* (1991) proposed a modification of this model involving a two-dimensional (2-D) array of blocks. The use of this approach greatly reduces the complexity of the calculations and the results using the two approaches are generally very similar. A wide variety of slider-block models have been proposed and studied; these have been reviewed by Rundle *et al.* (2003).

Huang and Turcotte (1992) carried out a large number of simulations on a square array of blocks

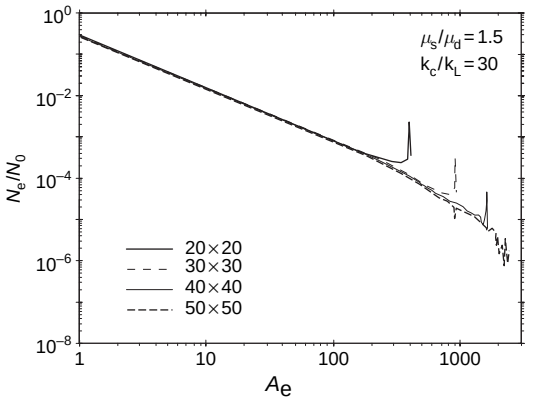


Figure 9 Results of simulations for a 2-D slider-block model with multiple blocks (Huang and Turcotte, 1992). The ratio of the number of slip events, N_e, with area A_e, to the total number of slip events N_0, is plotted against A_e – the number of blocks involved in an event. Results are given for systems with stiffness $k_c/k_L = 30$, ratio of coefficients of friction $\mu_s/\mu_d = 1.5$, and grid sizes 20×20, 30×30, 40×40, and 50×50. The peaks at $A_e = 400$, 900, and 1600, correspond to catastrophic slip events involving the entire system.

using stick-slip friction and the cellular automata approach. Their noncumulative frequency–area statistics for model slip events are given in **Figure 9**. The number of slip events per time step N_e/N_T with area A_e is given as a function of A_e. The simulation results are well fit by the power-law scaling relation [36] with $D = 2.72$. For stiff systems (k_c/k_L large), the entire grid of slider blocks is strongly correlated and large slip events including all the blocks occur regularly. These are the peaks for $A_e = 400$, 900, and 1600 blocks, illustrated in **Figure 9**. For soft systems (k_c/k_L relatively small), no large events occur.

A simple inverse-cascade model can explain the behavior of slider-block arrays (Turcotte, 1999a, 1999b). A metastable cluster is the region over which a slip event propagates once it has been initiated. At any one time, there can be many metastable clusters on the grid. The metastable regions grow by coalescence which is self-similar and gives power-law scaling. Avalanches sample the distribution of smaller clusters and, at the same time, remove the largest clusters. Gabrielov *et al.* (1999) and Yakovlev *et al.* (2005) have shown that the coalescence of clusters forms an inverse cascade that is identical to the formation of fractal drainage networks. The primary mechanism for the coalescence of clusters is bridging of gaps between clusters. The inverse cascade is further quantified by introducing the concept of a cluster order which is identical to the classification of stream networks. The joining of streams as stream networks evolve is analogous to the coalescence of clusters as they grow. Streams can be classified according to both the primary branching and side or secondary branching. Clusters can be similarly defined. When two clusters of equal order coalesce it is analogous to the joining of two streams of the same order. When a cluster of lower order coalesces with a cluster of higher order it is analogous to the side branching between a stream of lower and higher order. In the slider-block model, slip events sample the population of smaller clusters but do not deplete their number. Metastable blocks cascade from smaller clusters to larger clusters until the cascade is terminated in the largest slip events. These largest slip events define the upper limit of the power-law frequency–area scaling of both metastable clusters and slip events. The cascade is self-similar and the populations of clusters in terms of primary and secondary coalescences are identical to the populations of branches in self-similar drainage networks.

4.23.4 Relevant Models

4.23.4.1 Cellular Automata Models

Cellular automata models involve a grid of sites. At each time step, the model is updated by a prescribed set of rules. Interactions among prescribed sites are limited to a neighborhood, usually nearest neighbors. The two models discussed in the previous section, the sandpile model and the slider-block model are examples of cellular automata models. In this section we will consider a third example, the forest-fire model (Drossel and Schwabl, 1992). The forest-fire model consists of a square grid of sites, at each time step either an attempt is made to plant a tree on a randomly selected site or a model match is dropped on the site. A tree is planted if a site is not occupied by a tree. If a match is dropped on a site occupied by a tree, that tree and all adjacent trees are destroyed in a model fire. The sparking frequency is the inverse of the number of attempted tree drops before a model match is dropped. It is found that the frequency–area distribution of fires over a range of sizes is power-law with a slope near unity. Thus the forest-fire model behaves very much like the sandpile and the slider-block models. Like the sandpile model, sites are selected randomly so that it is a stochastic model. Like the slider-block model there is a tuning

parameter, the sparking frequency. If the sparking frequency is high, fires prevent the formation of large clusters. The power-law cascade of planted trees from small to large clusters is terminated with a maximum cluster size that is a fraction of the grid size. If the sparking frequency is low, large system size clusters occur. The critical point is the value of the firing frequency for the transition between these two behaviors.

Another model that is widely studied in statistical physics is the site-percolation model (Stauffer and Aharony, 1992). The site-percolation model can be regarded as the forest-fire model without fires. This is a transient problem as the number of trees on the grid increases monotonically until every site is occupied by a tree. However, there is a well-defined critical point during this evolution. This critical point occurs when a cluster of trees spans the grid. Monte Carlo simulations with large square grids show that a spanning cluster is formed when 59.275% of the grid sites are occupied by trees. At this critical point, the frequency–area distribution of tree clusters is power-law with a slope of 2.055. This similarity to the forest-fire model only occurs at the critical point. At other times during the transient planting of trees, the distribution is not power law (Turcotte, 1999b). The tuning parameter in this critical-point problem is the fraction of sites occupied by the spanning cluster. The critical value of this tuning parameter for site percolation on a square grid is $p_c = 0.59275$.

In order to study the behavior of characteristic earthquakes, a model has been developed (Newman and Turcotte, 2002) that combines aspects of the forest-fire and site-percolation models. In this model, fires only burn spanning (percolating) clusters. Trees are planted on unoccupied sites until a cluster spans the grid, at this time step the entire cluster is removed and tree planting continues. The inverse cascade of trees from smaller to larger clusters gives a power-law distribution of cluster sizes at all times.

The fires in this model are somewhat analogous to characteristic earthquakes on a major fault or fault segment. The planting of tree corresponds to the tectonic increase in stress. The variations in cluster structure correspond to stress variations in the crust. The variability of the sizes is analogous to the variability in characteristic earthquakes and the associated variability in the recurrence times between characteristic earthquakes. In this model, there is a memory of the structure of previous fires for any given fire,

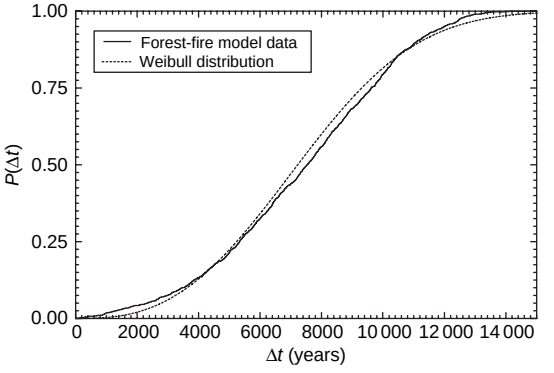


Figure 10 Cumulative distribution of recurrence intervals (in time steps) between model fires. The dashed line is the best-fit stretched exponential distribution ($\mu = 7385$, $\sigma = 2908$, $C_v = 0.3938$, $\tau = 8300$, and $\beta = 2.74$).

each characteristic earthquake is influenced by the structure of previous characteristic earthquakes. The cdf of recurrence intervals (in time steps) between fires is given in **Figure 10**. The mean number of time steps between fires is $\mu = 7385$, the standard deviation is $\sigma = 2908$, and the coefficient of variation is $C_v = 0.3938$. Also included is the cdf for the Weibull distribution [34] with the same μ and coefficient of variation C_v as the model results, this required $\tau = 8300$ and $\beta = 2.74$. It is also interesting to note that the $C_v = 0.394$ for the recurrence times of the model fires is very similar to the $C_v = 0.378$ the Parkfield earthquakes.

The applicability of the Weibull distribution to the recurrence times for earthquakes was discussed in Section 4.23.2.5. It was pointed out that the Weibull distribution is widely used for failure-time statistics in engineering (Meeker and Escobar, 1998).

A standard failure problem in engineering is the fiber bundle or stranded cable. The generally accepted approach to the dynamic time-dependent failure of a fiber bundle is to specify an expression for the rate of failure of fibers (Coleman, 1958). The form of this breakdown rule is given by

$$\frac{dN}{dt} = -v_0 N \left(\frac{\sigma}{\sigma_0}\right)^p \qquad [40]$$

where N is the number of unbroken fibers at time t, σ is the applied stress that can be a function of time, v_0 is the reference hazard rate at the reference stress σ_0, and p is the power-law exponent.

Since in the fiber-bundle model we assume that the stress σ on all fibers is equal, the fibers do not interact. Thus, it is appropriate to assume an ergodic

hypothesis. Instead of applying our model to a fiber bundle with N_0 fibers, we apply it to a single fiber that is replaced N_0 times. The single fiber is analogous to a fault, and the failure of the fiber is analogous to an earthquake. The distribution of failure times of the fiber is analogous to the distribution of earthquake recurrence times on a fault. The cdf of failure times of fibers P is related to N by

$$P(t) = 1 - \frac{N(t)}{N_0} \quad [41]$$

where N_0 is the total number of earthquakes. We further assume that after each fiber failure the stress on the fiber is reset to zero and increases linearly with time

$$\sigma(t) = \sigma_0 \frac{t}{t_0} \quad [42]$$

This represents the linear increase in the tectonic stress on a fault after an earthquake. In order to relate the constants v_0 and p in the fiber-bundle model to the constants β and τ in the Weibull distribution, we write

$$p = \beta - 1 \quad [43]$$

$$v_0 = \frac{\beta}{\tau} \quad [44]$$

Substitution of [43] to [44] into [41] gives

$$\frac{dP(t)}{dt} = \frac{\beta}{\tau}\left(\frac{t}{\tau}\right)^{\beta-1}[1 - P(t)] \quad [45]$$

Integrating with the initial condition that $P(0) = 0$ at $t = 0$ gives

$$P(t) = 1 - \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right] \quad [46]$$

This is the cumulative form of the Weibull distribution given in [34].

4.23.4.2 Numerical Simulations

Simulation-based approaches to forecasting and prediction of natural phenomena have been used with considerable success for weather and climate. Turbulent phenomena are represented by parametrizations of the fluid dynamics, and the equations are typically solved over spatial grids having length scales of tens to hundreds of kilometers. Even though the behavior of the atmosphere and oceans is recognized to be chaotic, numerical weather forecasts are generally quite accurate for periods of 48–72 h. A particular example is the probabilistic forecasts of hurricane tracks.

Ideally, similar simulations can be carried out to assess the earthquake hazard. However, tectonic models involving finite strain are not feasible at this time. Such models would require both block motions (displacements on faults) and continuum deformation (granulation). It is not possible to evolve block motions due to geometrical incompatibilities, there will be either overlaps or gaps.

A more limited simulation model for distributed seismicity on the San Andreas and adjacent faults in southern California was given by Rundle (1988). This model included stress accumulation and release as well as stress interactions including the San Andreas and adjacent faults. The model was based on a set of mapped faults with estimated slip rates, a prescribed plate tectonic motion, earthquakes on all faults, and elastic interactions.

An updated version of the model has been developed (Rundle *et al.*, 2001, 2002a, 2004, 2005). This version, called Virtual California, is a 3-D earthquake simulation code characterized by

1. Rectangular, vertically dipping strike-slip fault segments embedded in an elastic half space. The faults interact elastically utilizing dislocation theory.
2. A friction law based on laboratory studies.
3. Stress accumulation by means of 'backslip', so that the long-term rate of slip is matched, on average, by the observed rate of stress accumulation on the faults comprising the fault system.

A similar model, Standard Physical Earth Model (SPEM), was developed by Ward (1992) and applied to seismicity associated with subduction at the Middle America trench. This model was further developed and applied to the entire San Andreas system by Goes and Ward (1994), to the San Andreas system in southern California by Ward (1996), and to the San Andreas system in northern California by Ward (2000).

Although the statistics of the simulated earthquakes produced by SPEM are similar to those produced by Virtual California, there are important differences between the two codes. Whereas Virtual California involves rectangular fault elements in an elastic half space, SPEM is a plane strain (or plane stress) computation in an elastic plate. Even though there are significant differences between the details of the physics of the two codes, SPEM and Virtual California, the coefficients of variation, inter-event

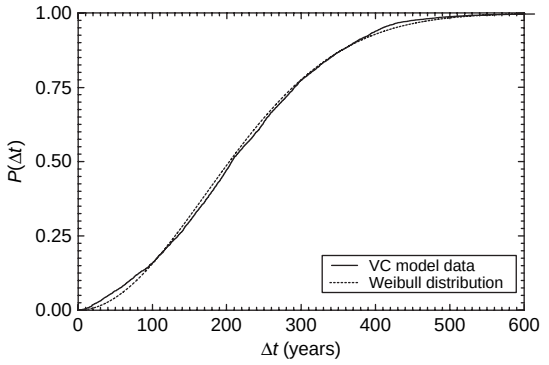


Figure 11 Cumulative distribution of recurrence times for 4096 simulated earthquakes with $m > 7.5$ on the northern San Andreas Fault. The dashed line is the best-fit stretched exponential distribution ($\mu = 217$ yr, $\sigma = 115$ yr, $C_v = 0.528$, $\tau = 245$ yr, and $\beta = 1.97$).

statistics, and probabilities of occurrence are roughly similar, implying that the fault interactions lead to a large-scale organization of the system. The statistical distribution of recurrence times on the northern San Andreas Fault (site of the 1906 San Francisco earthquake) was obtained from a 1 000 000 year simulation. The cdf for 4606 simulated earthquakes with $M > 7.5$ on this section is given in **Figure 11**. The mean recurrence time is $\mu = 217$ years, the standard deviation is $\sigma = 114.7$ years, and the coefficient of variation is $C_v = 0.528$. Also included is the cdf for the Weibull distribution with the same mean μ and standard deviation σ as the simulation results. This required $\tau = 245$ yr and $\beta = 1.97$. It is seen that the Weibull distribution is in quite good agreement with the results of the simulation.

4.23.4.3 Damage Mechanics Models

One approach to the complex brittle deformation of a solid is to use damage mechanics. This approach was originally developed to quantify the irreversible brittle deformation of solids associated with microcracking. Several empirical continuum damage mechanics models were introduced and are used in civil and mechanical engineering (Kachanov, 1986; Krajcinovic, 1996; Shcherbakov and Turcotte, 2003). Damage mechanics has also been applied to the brittle deformation of the Earth's crust by a number of authors (Lyakhovsky *et al.*, 1997, 2001; Ben-Zion and Lyakhovsky, 2002; Turcotte *et al.*, 2003). In this application, the deformation and damage is associated with earthquakes.

The application of continuum damage mechanics can be illustrated by considering a rod in a state of uniaxial stress. For an elastic material Hooke's law is applicable and is written in the form

$$\sigma = E_0 \varepsilon \qquad [47]$$

where ε is a strain and E_0 is the Young's modulus of the undamaged material.

In this paper we will consider a model of continuum damage mechanics as introduced by Shcherbakov *et al.* (2005a). If the stress is less than the yield stress, [47] is assumed to be valid. If the stress is greater than the yield stress, $\sigma > \sigma_y$, a damage variable α is introduced according to

$$\sigma - \sigma_y = E_0(1-\alpha)(\varepsilon - \varepsilon_y) \qquad [48]$$

where $\sigma_y = E_0 \varepsilon_y$. When $\alpha = 0$, [48] reduces to [47] and linear elasticity is applicable, as failure occurs. Increasing values of α in the range $0 \leq \alpha < 1$ quantify the weakening (decreasing E) associated with the increase in the number and size of microcracks in the material. Several authors (Krajcinovic, 1996; Turcotte *et al.*, 2003) have shown a direct correspondence between the damage variable α in a continuum material and the number of surviving fibers N in a fiber bundle that originally had N_0 fibers, $\alpha = 1 - N/N_0$.

To complete the formulation of the damage problem it is necessary to specify the kinetic equation for the damage variable. In analogy to (Lyakhovsky *et al.*, 1997) we take

$$\frac{d\alpha(t)}{dt} = 0, \quad \text{if} \quad 0 \leq \sigma \leq \sigma_y \qquad [49]$$

$$\frac{d\alpha(t)}{dt} = \frac{1}{t_d}\left[\frac{\sigma(t)}{\sigma_y} - 1\right]^p \left[\frac{\varepsilon(t)}{\varepsilon_y} - 1\right]^2, \quad \text{if} \quad \sigma > \sigma_y \qquad [50]$$

where t_d is a characteristic time scale for damage and p is a constant to be determined from experiments. The power-law dependence of $d\alpha/dt$ on stress (and strain) given above must be considered empirical in nature. However, a similar power dependence of dN/dt on stress is widely used in the analysis of fiber failures in fiber bundles (Newman and Phoenix, 2001).

As a first example we considered a rod to which a constant uniaxial tensional stress $\sigma_0 > \sigma_y$ is applied instantaneously at $t = 0$ and held constant until the sample fails. The applicable kinetic equation for the rate of increase of damage with time is obtained from [48]–[50] with the result

$$\frac{d\alpha}{dt} = \frac{1}{t_d}\left(\frac{\sigma_0}{\sigma_y} - 1\right)^{\rho+2} \frac{1}{[1-\alpha(t)]^2} \quad [51]$$

Integrating with the initial condition $\alpha(0) = 0$ gives

$$\alpha(t) = 1 - \left[1 - 3\frac{t}{t_d}\left(\frac{\sigma_0}{\sigma_y} - 1\right)^{\rho+2}\right]^{1/3} \quad [52]$$

Substituting [52] into [48] gives the strain in the sample as a function of time t

$$\frac{\varepsilon(t)}{\varepsilon_y} = 1 + \frac{(\sigma_0/\sigma_y - 1)}{\left[1 - 3\frac{t}{t_d}(\sigma_0/\sigma_y - 1)^{\rho+2}\right]^{1/3}} \quad [53]$$

Failure occurs at the time t_f when $\alpha \to 1$ ($\varepsilon \to \infty$). From [53] this failure time is given by

$$t_f = \frac{t_d}{3}\left(\frac{\sigma_0}{\sigma_y} - 1\right)^{-(\rho+2)} \quad [54]$$

The time to failure tends to infinity as a power-law as the applied stress approaches the yield stress.

A particularly interesting set of experiments on brittle failure were carried out by Guarino *et al.* (1999). These authors studied the failure of circular panels (222 mm diameter, 3–5 mm thickness) of chipboard panels. A differential pressure was applied rapidly across a panel and it was held constant until failure occurred. For these relatively thin panels, bending stresses were negligible and the panels failed under tension (a mode I fracture). Initially the microcracks appeared to be randomly distributed across the panel, but as the number of microcracks increased they tended to localize and coalesce in the region where the final rupture occurred. The times to failure t_f of these chipboard panels were found to depend systematically on the applied differential pressure P. Taking a yield pressure (stress) $P_y = 0.038$ MPa their results correlate very well with [53] taking $p = 0.25$ and $t_d = 168$ s. Shcherbakov *et al.* (2005a) and Ben-Zion and Lyakhovsky (2002) showed that the damage mechanic solution given in [53] provides an explanation for the power-law increase in Benioff strain associated with AMR.

As our final example we will consider a rod to which a constant uniaxial tensional strain $\varepsilon_0 > \varepsilon_y$ has been applied instantaneously at $t = 0$ and is held constant. The applicable equation for the rate of increase of damage with time is obtained from [48]–[51] with the result

$$\frac{d\alpha}{dt} = \frac{1}{t_d}\left(\frac{\varepsilon_0}{\varepsilon_y} - 1\right)^{\rho+2} 1 - \alpha(t)]^{\rho} \quad [55]$$

Integrating with the initial condition $\alpha(0) = 0$, we find

$$\alpha(t) = 1 - \left[1 + (\rho - 1)\frac{t}{t_d}\left(\frac{\varepsilon_0}{\varepsilon_y} - 1\right)^{\rho+2}\right]^{-1/(\rho-1)} \quad [56]$$

This result is valid as long as $p > 1$. The damage increases monotonically with time and as $t \to \infty$ the maximum damage is $\alpha(\infty) = 1$. Using [56] with [48] one gets the stress relaxation in the material as a function of time t

$$\frac{\sigma(t)}{\sigma_y} = 1 + \left(\frac{\varepsilon_0}{\varepsilon_y} - 1\right)\left[1 + (\rho - 1)\frac{t}{t_d}\left(\frac{\varepsilon_0}{\varepsilon_y} - 1\right)^{\rho+2}\right]^{-1/(\rho-1)} \quad [57]$$

At $t = 0$ we have linear elasticity corresponding to $\alpha = 0$. In the limit $t \to \infty$ the stress relaxes to the yield stress $\sigma(\infty) = \sigma_y$ below which no further damage can occur.

In order to relate our continuum damage mechanics model to aftershocks we determine the rate of energy release in the relaxation process considered above. Since the strain is constant during the stress relaxation, no work is done on the sample. We hypothesize that if the applied strain (stress) is instantaneously removed during the relaxation process the sample will return to a state of zero stress and strain following a linear stress–strain path with slope $E_0(1 - \alpha)$ to stress σ_y and a path with slope E_0 to zero stress. We assume that the difference between the energy added ε_{YA} and the energy recovered e_{aY} is lost in aftershocks and find that this energy e_{as} is given by

$$e_{as} = e_{YA} - e_{aY} = \frac{1}{2}E_0(\varepsilon_0 - \varepsilon_y)^2\alpha(t) \quad [58]$$

The rate of energy release is obtained by substituting [55] into [58] and taking the time derivative with the result

$$\frac{de_{as}}{dt} = \frac{E_0\varepsilon_y^2}{2t_d}\left(\frac{\varepsilon_0}{\varepsilon_y} - 1\right)^{\rho+4} \times \left[1 + (\rho - 1)\left(\frac{\varepsilon_0}{\varepsilon_y} - 1\right)^{\rho+2}\left(\frac{t}{t_d}\right)\right]^{-\rho/(\rho-1)} \quad [59]$$

The total energy of aftershocks e_{ast} is obtained by substituting $\alpha = 1$ into [58] with the result

$$e_{ast} = \frac{1}{2}E_0(\varepsilon_0 - \varepsilon_y)^2 \quad [60]$$

Combining [59] and [60] gives

$$\frac{1}{e_{\text{ast}}}\frac{\mathrm{d}e_{\text{as}}}{\mathrm{d}t}=\frac{\frac{1}{t_{\text{d}}}(\varepsilon_0/\varepsilon_{\text{y}}-1)^{\rho+2}}{\left[1+(\rho-1)(\varepsilon_0/\varepsilon_{\text{y}}-1)^{\rho+2}\left(\frac{t}{t_{\text{d}}}\right)\right]^{\rho/(\rho-1)}} \quad [61]$$

This result is clearly similar to the aftershock relation given in [9]. To demonstrate this further let us make the substitutions

$$\rho=\frac{p}{p-1} \quad [62]$$

$$c=\frac{t_{\text{d}}}{(\rho-1)(\varepsilon_0/\varepsilon_{\text{y}}-1)^{\rho+2}} \quad [63]$$

Substitution of [62] and [63] into [61] gives

$$\frac{1}{e_{\text{ast}}}\frac{\mathrm{d}e_{\text{as}}}{\mathrm{d}t}=\frac{p-1}{c}\frac{1}{(1+t/c)^{p}} \quad [64]$$

This damage mechanics result is identical in form to the modified Omori's law for aftershocks given in [9].

We next consider a specific example. In **Figure 3** the rate of occurrence of aftershocks with $m\geq 2.0$ following the $m_{\text{ms}}=6.0$ Parkfield (California) earthquake, 28 Sep. 2004, is given as a function of the time after the earthquake. Also shown is the correlation with [7] taking $\tau=43.7$ s, $c=0.0199$ days, and $p=1.09$. From [62] we find that the corresponding power-law exponent is $\rho=12.0$. Further assuming that $\varepsilon_0/\varepsilon_{\text{y}}=1.5$ we find from [63] that the damage time is $t_{\text{d}}=0.01$ s.

4.23.4.4 Epidemic-Type Aftershock Sequence (ETAS) Models

It is clear that aftershocks play an important role in regional seismicity. One approach to assessing this role, that has received considerable recent attention, is the epidemic-type aftershock sequence (ETAS) branching model. This model recognizes that each primary aftershock will generate its own family of secondary aftershocks, the second-order aftershocks will generate third-order aftershocks, and so forth. This is a stochastic model that utilizes the GR frequency–magnitude distribution, the modified Omori's law temporal decay distribution, and a prescribed spatial distribution to create synthetic catalogs. Of particular importance is the prescribed branching ratio.

This approach was first given by Kagan and Knopoff (1981) and in a more comprehensive form by Ogata (1988). Further refinement have been given by Ogata (1992, 1999b, 2001, 2004); Ogata *et al.* (1993, 2003); Guo and Ogata (1997); Helmstetter and Sornette (2002a, 2002b); Helmstetter and Sornette (2003a, 2003b); Helmstetter *et al.* (2003a, 2003b); Sornette and Helmstetter (2002); Saichev and Sornette (2004, 2005); Saichev *et al.* (2005); Sornette and Werner (2005). In this model each event is capable of producing secondary aftershock sequences and can be considered simultaneously as a foreshock, main shock, or aftershock. The resulting aftershock sequence is a combined effect of many aftershock sequences produced by each aftershock.

ETAS is an interesting way to utilize accepted scaling laws to generate synthetic earthquake catalogs. But the approach does not take account of the tectonic addition of stress associated with 'characteristic' earthquakes.

4.23.5 Discussion

4.23.5.1 Overview

Earthquakes and seismicity are clearly complex phenomena. Yet, within this complexity, there are robust scaling laws. Both GR frequency–magnitude scaling and Omori's law for the temporal decay of aftershocks were introduced empirically. It can be shown that GR scaling is equivalent to fractal scaling between the number of earthquakes and their rupture area. This scaling is scale invariant, it is robust, but do we understand it? One approach is to directly associate this scaling with the power-law slip-event scaling obtained in slider-block models. But in these models an individual slider block can participate in events of all sizes. This does not seem to be the case for earthquakes on faults, big faults appear to have large earthquakes and small faults small earthquakes. Thus, the GR scaling may be the consequence of a fractal distribution of fault sizes. This in turn can be attributed to scale-invariant comminution of the earth's brittle crust in active tectonic regions.

Omori's law also appears to be universally applicable, but why? Why do aftershocks occur? A simple explanation is that some regions experience an increase in stress during a main shock, although the average stress level decreases. Aftershocks relieve the excess stress in these regions. But why the time delay? A number of explanations have been given as discussed in Section 4.23.2.3. Both damage mechanics and fiber-bundle models incorporate time delays for brittle failure on an empirical basis. But the fundamental physics of these time delays remains controversial. There are certainly similarities

between these times and those associated with phase changes, for example the explosive boiling of superheated water.

Another important question concerns the relative roles of aftershocks and 'characteristic earthquakes'. The ETAS model is an interesting approach to seismicity. GR and Omori's laws are assumed and earthquakes are generated randomly from these distributions. Because of this randomness, later earthquakes can be larger than the initial 'main shock'. The original main shock becomes a foreshock. But there is certainly a role for characteristic earthquakes and a tectonic drive. It is clearly desirable to obtain a better understanding of the relative roles of characteristic earthquakes and aftershocks.

4.23.5.2 Earthquake Forecasting (Prediction)

A fundamental question is whether forecasts of the time and location of future earthquakes can be accurately made. It is accepted that long-term hazard maps of the expected rate of occurrence of earthquakes are reasonably accurate. But is it possible to do better? Are there precursory phenomena that will allow earthquakes to be forecast?

It is actually quite surprising that short-term local precursory phenomena are not seen. Prior to a volcanic eruption, increases in regional seismicity and surface movements are generally observed (Chastin and Main, 2003). For a fault system, the stress gradually increases until it reaches the frictional strength of the fault and a rupture is initiated. It is certainly reasonable to hypothesize that the stress increase would cause increases in background seismicity and aseismic slip. In order to test this hypothesis the Parkfield Earthquake Prediction Experiment was initiated in 1985. The expected Parkfield earthquake occurred beneath the heavily instrumented region on 28 Sep. 2004. No local precursory changes were observed (Bakun *et al.*, 2005).

In the absence of local precursory signals, the next question is whether broader anomalies develop, and in particular whether there is anomalous seismic activity. One approach to earthquake forecasting is to use the rate of occurrence of small earthquakes in the past. This type of forecast is referred to as a 'relative intensity' (RI) forecast (Holliday *et al.*, 2005). In such a forecast, the study region is tiled with boxes, that is, $0.1^\circ \times 0.1^\circ$. The number of earthquakes with magnitudes greater than a specified value in each box is determined over a specified time period. The RI score for each box is then computed as the total number of earthquakes in the box in the time period divided by the value for the box having the largest value. A threshold value in the interval [0–1] is then selected. Larger earthquakes are then forecast to occur only in boxes having an RI value larger than the threshold. The remaining boxes with RI scores smaller than the threshold represent sites at which larger earthquakes are forecast to not occur. The physical justification for this type of forecast is that large earthquakes are considered most likely to occur at sites of high seismic activity of small events in the past, that is, large '*a*'-values in the GR scaling given in II.1.

A second approach to earthquake forecasting is based on pattern informatics (PI) method, which quantifies temporal variations in seismicity (Rundle *et al.*, 2002b; Tiampo *et al.*, 2002a, 2002b, 2002c; Holliday *et al.*, 2005). This approach is similar to the RI method. Again the study region is tiled with boxes. Three time intervals are considered. The number of earthquakes per year in a box with magnitudes greater than a specified value is determined first for a reference interval t_0–t_2. Then the earthquake rates for a change interval t_1–t_2 are determined. The change in rates between the two intervals are quantified and the boxes with maximum changes are determined. These boxes ('hotspots') are the regions where earthquakes are forecast to occur during a forecast period t_2–t_3. The PI method was first applied to seismicity in southern California and adjacent regions (32°–37° N lat, 238°–245° E long). This region was divided into a grid of 3500 boxes with $\Delta x = 0.1^\circ$ (11 km). Consistent with the sensitivity of the southern California seismic network, the lower magnitude cutoff was taken to be $m = 3.0$. The initial time was $t_0 = 1932$, the change interval was from $t_1 = 1990$ to $t_2 = 2000$, and the forecast interval was from $t_2 = 2000$ to $t_3 = 2010$. The initial studies for California were published in 2002 (Rundle *et al.*, 2002b). This forecast was considered to be valid for the interval from 2000 to 2010 and would be applicable for earthquakes with $m = 5.0$ and larger. Since 1 Jan. 2000 18 earthquakes with $m \geq 5.0$ have occurred in the test region. We consider the forecast to be successful if the epicenter of the earthquake lies within a hotspot box or in one of the eight adjoining boxes. Of the 18 earthquakes, 16 were successfully forecast.

A series of statistical algorithms to make intermediate-term earthquake predictions have been developed by a Russian group under the direction of V. I. Keilis-Borok using pattern recognition

techniques (Keilis-Borok, 1990, 2002). Seismicity in circular regions with diameters of 500 km was analyzed. Based primarily on seismic activation, earthquake alarms were issued for one or more regions, with the alarms generally lasting for 3 years. Alarms have been issued regularly since the mid 1980s and scored two notable successes: the prediction of the 1988 Armenian earthquake and the 1989 Loma Prieta earthquake. While a reasonably high success rate has been achieved, there have been some notable misses including the recent $m = 9.3$ Sumatra and $m = 8.1$ Macquerie Island earthquakes.

More recently, this group has used chains of premonitory earthquakes as the basis for issuing alarms (Shebalin *et al.*, 2004; Keilis-Borok *et al.*, 2004). This method successfully predicted the $m = 6.5$, 22 Dec. 2003 San Simeon (California) earthquake and the $m = 8.1$, 25 Sep. 2003 Tokachi-oki, (Japan) earthquake with lead times of 6 and 7 months, respectively. However, an alarm issued for southern California, valid during the spring and summer of 2004, was a false alarm. Both the PI method and the Russian forecast algorithms depend heavily on variation in seismicity prior to an earthquake. Thus they may be related to AMR as discussed in Section 4.23.2.4.

4.23.5.3 Future Directions

Many aspects of seismology exhibit complexity as has been discussed in previous sections. This is an active area of research in both the geophysical and statistical physics communities. Although much progress has been made, many questions remain. Relevant areas include scaling laws, temporal and spatial correlations, critical phenomena, and nucleation.

Several scaling laws are now widely accepted. These include GR frequency–magnitude scaling, Omori's law for the decay of aftershock activity, and Båth's law relating the magnitude of the largest aftershock to the magnitude of the main shock. Other scaling laws are more controversial and require further studies. These include the power-law behavior of AMR, the Weibull distribution of recurrence times between characteristic earthquakes, and the nonhomogeneous Poisson distribution of interoccurrence times between aftershocks. The relative role of aftershocks and characteristic earthquakes needs to be clarified.

Models relevant to earthquakes and complexity are at an early stage of development. Slider-block models have certainly played a role but are clearly only weakly related to distributed seismicity. Damage mechanics also plays a role in understanding the complex behavior of brittle materials. Realistic simulations of distributed seismicity are just beginning to be developed. A major objective of these models is to provide estimates of the seismic hazard. A fundamental question is whether the future occurrence of earthquakes can be predicted. Certainly, the occurrence of small earthquakes is routinely utilized to forecast the risk of large earthquakes. This extrapolation is based on the applicability of GR frequency–magnitude statistics (Kossobokov *et al.*, 2000). But can seismicity patterns be used to reliably forecast future earthquakes in a timely manner? There is accumulating evidence that this may be possible.

Acknowledgment

This work has been supported by National Science Foundation Grant ATM-0327571.

References

Abe S and Suzuki N (2003) Law for the distance between successive earthquakes. *Journal of Geophysical Research* 108(B2): 2113.

Abe S and Suzuki N (2004a) Scale-free network of earthquakes. *Europhysics Letters* 65(4): 581–586.

Abe S and Suzuki N (2004b) Small-world structure of earthquake network. *Physica A* 337(1–2): 357–362.

Aki K (1981) A probabilistic synthesis of precursory phenomena. In: Simpson DW and Richards PG (eds.) *Earthquake Prediction: An International Review, Maurice Ewing Series*, Vol. 4, pp. 566–574. Washington, DC: American Geophysical Union.

Altmann EG, da Silva EC, and Caldas IL (2004) Recurrence time statistics for finite size intervals. *Chaos* 14(4): 975–981.

Altmann EG and Kantz H (2005) Recurrence time analysis, long-term correlations, and extreme events. *Physical Review E* 71(5): 056106.

Avrami M (1940) Kinetics of phase change. II: Transformation-time relations for random distribution of nuclei. *Journal of Chemical Physics* 8(2): 212–224.

Baiesi M and Paczuski M (2004) Scale-free networks of earthquakes and aftershocks. *Physical Review E* 69(6): 066106.

Baiesi M and Paczuski M (2005) Complex networks of earthquakes and aftershocks. *Nonlinear Processes in Geophysics* 12(1): 1–11.

Bak P, Christensen K, Danon L, and Scanlon T (2002) Unified scaling law for earthquakes. *Physical Review Letters* 88(17): 178501.

Bak P, Tang C, and Wiesenfeld K (1988) Self-organized criticality. *Physical Review A* 38(1): 364–374.

Bakun WH, Aagaard B, Dost B, *et al.* (2005) Implications for prediction and hazard assessment from the 2004 Parkfield earthquake. *Nature* 437(7061): 969–974.

Barriere B and Turcotte DL (1991) A scale-invariant cellular-automata model for distributed seismicity. *Geophysical Research Letters* 18(11): 2011–2014.

Barriere B and Turcotte DL (1994) Seismicity and self-organized criticality. *Physical Review E* 49(2): 1151–1160.

Båth M (1965) Lateral inhomogeneities in the upper mantle. *Tectonophysics* 2: 483–514.

Ben-Zion Y and Lyakhovsky V (2002) Accelerated seismic release and related aspects of seismicity patterns on earthquake faults. *Pure and Applied Geophysics* 159(10): 2385–2412.

Bowman DD and King GCP (2001) Accelerating seismicity and stress accumulation before large earthquakes. *Geophysical Research Letters* 28(21): 4039–4042.

Bowman DD, Ouillon G, Sammis CG, Sornette A, and Sornette D (1998) An observational test of the critical earthquake concept. *Journal of Geophysical Research* 103(B10): 24359–24372.

Bowman DD and Sammis CG (2004) Intermittent criticality and the Gutenberg-Richter distribution. *Pure and Applied Geophysics* 161(9–10): 1945–1956.

Brehm DJ and Braile LW (1998) Intermediate-term earthquake prediction using precursory events in the New Madrid seismic zone. *Bulletin of the Seismological Society of America* 88(2): 564–580.

Brehm DJ and Braile LW (1999a) Intermediate-term earthquake prediction using the modified time-to-failure method in southern California. *Bulletin of the Seismological Society of America* 89(1): 275–293.

Brehm DJ and Braile LW (1999b) Refinement of the modified time-to-failure method for intermediate-term earthquake prediction. *Journal of Seismology* 3(2): 121–138.

Brown SR, Scholz CH, and Rundle JB (1991) A simplified spring-block model of earthquakes. *Geophysical Research Letters* 18(2): 215–218.

Bufe CG, Nishenko SP, and Varnes DJ (1994) Seismicity trends and potential for large earthquakes in the Alaska–Aleutian region. *Pure and Applied Geophysics* 142(1): 83–99.

Bufe CG and Varnes DJ (1993) Predictive modeling of the seismic cycle of the greater San-Francisco bay-region. *Journal of Geophysical Research* 98(B6): 9871–9883.

Bunde A, Eichner JF, Havlin S, and Kantelhardt JW (2003) The effect of long-term correlations on the return periods of rare events. *Physica A* 330(1–2): 1–7.

Bunde A, Eichner JF, Havlin S, and Kantelhardt JW (2004) Return intervals of rare events in records with long-term persistence. *Physica A* 342(1–2): 308–314.

Burridge R and Knopoff L (1967) Model and theoretical seismicity. *Bulletin of the Seismological Society of America* 57: 341–371.

Carbone V, Sorriso-Valvo L, Harabaglia P, and Guerra I (2005) Unified scaling law for waiting times between seismic events. *Europhysics Letters* 71(6): 1036–1042.

Carlson JM and Langer JS (1989) Mechanical model of an earthquake fault. *Physical Review E* 40(11): 6470–6484.

Chastin SFM and Main IG (2003) Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions. *Geophysical Research Letters* 30(13): 1671.

Coleman BD (1958) Statistics and time dependence of mechanical breakdown in fibers. *Journal of Applied Physics* 29: 968–983.

Console R, Lombardi AM, Murru M, and Rhoades D (2003a) Båth's law and the self-similarity of earthquakes. *Journal of Geophysical Research* 108(B2): 2128.

Console R and Murru M (2001) A simple and testable model for earthquake clustering. *Journal of Geophysical Research* 106(B5): 8699–8711.

Console R, Murru M, and Lombardi AM (2003b) Refining earthquake clustering models. *Journal of Geophysical Research* 108(B10): 2468.

Corral Á (2003) Local distributions and rate fluctuations in a unified scaling law for earthquakes. *Physical Review E* 68(3): 035102.

Corral Á (2004a) Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Physical Review Letters* 92(10): 108501.

Corral Á (2004b) Universal local versus unified global scaling laws in the statistics of seismicity. *Physica A* 340(4): 590–597.

Corral Á (2005a) Mixing of rescaled data and Bayesian inference for earthquake recurrence times. *Nonlinear Processes in Geophysics* 12(1): 89–100.

Corral Á (2005b) Time-decreasing hazard and increasing time until the next earthquake. *Physical Review E* 71(1): 017101.

Das S and Scholz CH (1981) Off-fault aftershock clusters caused by shear-stress increase. *Bulletin of the Seismological Society of America* 71(5): 1669–1675.

Davidsen J and Goltz C (2004) Are seismic waiting time distributions universal? *Geophysical Research Letters* 31(21): L21612.

Davidsen J and Paczuski M (2005) Analysis of the spatial distribution between successive earthquakes. *Physical Review Letters* 94(4): 048501.

Davis PM, Jackson DD, and Kagan YY (1989) The longer it has been since the last earthquake, the longer the expected time till the next. *Bulletin of the Seismological Society of America* 79(5): 1439–1456.

Dhar D (1990) Self-organized critical state of sandpile automaton models. *Physical Review Letters* 64(14): 1613–1616.

Dhar D (1999) The Abelian sandpile and related models. *Physica A* 263(1–4): 4–25.

Dieterich J (1994) A constitutive law for rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research* 99(B2): 2601–2618.

Dobrovolsky IP, Zubkov SI, and Miachkin VI (1979) Estimation of the size of earthquake preparation zones. *Pure and Applied Geophysics* 117(5): 1025–1044.

Drossel B and Schwabl F (1992) Self-organized critical forest-fire model. *Physical Review Letters* 69(11): 1629–1632.

Ellsworth WL, Matthews MV, Nadeau RM, Nishenko SP, Reasenberg PA, and Simpson RW (1999) A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities. Technical report. *US Geological Survey Open-File Report 99–522*.

Felzer KR, Abercrombie RE, and Ekstrom G (2003) Secondary aftershocks and their importance for aftershock forecasting. *Bulletin of the Seismological Society of America* 93(4): 1433–1448.

Felzer KR, Becker TW, Abercrombie RE, Ekstrom G, and Rice JR (2002) Triggering of the 1999 m_W 7.1 Hector Mine earthquake by aftershocks of the 1992 m_W 7.3 Landers earthquake. *Journal of Geophysical Research* 107(B9): 2190.

Felzer KR and Brodsky EE (2006) Decay of aftershock density with distance indicates triggering by dynamic stress. *Nature* 441(7094): 735–738.

Frohlich C and Davis SD (1990) Single-link cluster-analysis as a method to evaluate spatial and temporal properties of earthquake catalogs. *Geophysical Journal International* 100(1): 19–32.

Frohlich C and Davis SD (1993) Teleseismic b-values – Or, much ado about 1.0. *Journal of Geophysical Research* 98(B1): 631–644.

Gabrielov A, Newman WI, and Turcotte DL (1999) Exactly soluble hierarchical clustering model: Inverse cascades, self-similarity, and scaling. *Physical Review E* 60(5): 5293–5300.

Geller RJ (1997) Earthquake prediction: A critical review. *Geophysical Journal International* 131(3): 425–450.
Geller RJ, Jackson DD, Kagan YY, and Mulargia F (1997) Earthquakes cannot be predicted. *Science* 275(5306): 1616–1617.
Goes SDB and Ward SN (1994) Synthetic seismicity for the San Andreas fault. *Annali Di Geofisica* 37: 1495–1513.
Goltz C and Bose M (2002) Configurational entropy of critical earthquake populations. *Geophysical Research Letters* 29(20): 1990.
Gross S and Rundle J (1998) A systematic test of time-to-failure analysis. *Geophysical Journal International* 133(1): 57–64.
Guarino A, Ciliberto S, and Garcimartin A (1999) Failure time and microcrack nucleation. *Europhysics Letters* 47(4): 456–461.
Guo ZQ and Ogata Y (1997) Statistical relations between the parameters of aftershocks in time, space, and magnitude. *Journal of Geophysical Research* 102(B2): 2857–2873.
Gutenberg B and Richter CF (1954) *Seismicity of the Earth and Associated Phenomenon 2nd edn. Princeton: Princeton University Press.*
Helmstetter A and Sornette D (2002a) Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models. *Physical Review E* 66(6): 061104.
Helmstetter A and Sornette D (2002b) Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research* 107(B10): 2237.
Helmstetter A and Sornette D (2003a) Båth's law derived from the Gutenberg–Richter law and from aftershock properties. *Geophysical Research Letters* 30(20): 2069.
Helmstetter A and Sornette D (2003b) Predictability in the epidemic-type aftershock sequence model of interacting triggered seismicity. *Journal of Geophysical Research* 108(B10): 2482.
Helmstetter A, Sornette D, and Grasso JR (2003a) Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws. *Journal of Geophysical Research* 108(B1): 2046.
Helmstetter AS, Ouillon G, and Sornette D (2003b) Are aftershocks of large Californian earthquakes diffusing? *Journal of Geophysical Research* 108(B10): 2483.
Henderson JR, Main IG, Maclean C, and Norman MG (1994) A fracture-mechanical cellular-automaton model of seismicity. *Pure and Applied Geophysics* 142(3–4): 545–565.
Hill DP, Reasenberg PA, Michael A, *et al.* (1993) Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* 260(5114): 1617–1623.
Holliday JR, Nanjo KZ, Tiampo KF, Rundle JB, and Turcotte DL (2005) Earthquake forecasting and its verification. *Nonlinear Processes in Geophysics* 12: 965–977.
Huang J and Turcotte DL (1990) Are earthquakes an example of deterministic chaos. *Geophysical Research Letters* 17(3): 223–226.
Huang J and Turcotte DL (1992) Chaotic seismic faulting with a mass-spring model and velocity-weakening friction. *Pure and Applied Geophysics* 138(4): 569–589.
Huang Y, Saleur H, Sammis C, and Sornette D (1998) Precursors, aftershocks, criticality and self-organized criticality. *Europhysics Letters* 41(1): 43–48.
Huc M and Main IG (2003) Anomalous stress diffusion in earthquake triggering: Correlation length, time dependence, and directionality. *Journal of Geophysical Research* 108(B7): 2324.
Ivashkevich EV and Priezzhev VB (1998) Introduction to the sandpile model. *Physica A* 254(1–2): 97–116.
Jaumé SC (2000) Changes in earthquake size-frequency distributions underlying accelerating moment/energy release. In: Rundle J, Lawson D, and Klein W (eds.) *Geocomplexity and the Physics of Earthquakes*, vol. 120 of Geophysical Monograph Series, pp. 199–210. Washington, DC: American Geophysical Union.
Jaumé SC and Sykes LR (1999) Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes. *Pure and Applied Geophysics* 155(2–4): 279–305.
Jensen HJ (1998) *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge: Cambridge University Press.
Kachanov LM (1986) *Introduction to Continuum Damage Mechanics*. Dordrecht: Martinus Nijhoff.
Kadanoff LP, Nagel SR, Wu L, and Zhou SM (1989) Scaling and universality in avalanches. *Physical Review A* 39(12): 6524–6537.
Kagan YY (2004) Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of the Seismological Society of America* 94(4): 1207–1228.
Kagan YY and Knopoff L (1981) Stochastic synthesis of earthquake catalogs. *Journal of Geophysical Research* 86(NB4): 2853–2862.
Kanamori H and Anderson DL (1975) Theoretical basis of some empirical relations in seismology. *Bulletin of the Seismological Society of America* 65(5): 1073–1095.
Keilis-Borok V (2002) Earthquake prediction: State-of-the-art and emerging possibilities. *Annual Review of Earth and Planetary Sciences* 30: 1–33.
Keilis-Borok V, Shebalin P, Gabrielov A, and Turcotte D (2004) Reverse tracing of short-term earthquake precursors. *Physics of the Earth and Planetary Interiors* 145(1–4): 75–85.
Keilis-Borok VI (1990) The lithosphere of the earth as a nonlinear-system with implications for earthquake prediction. *Reviews of Geophysics* 28(1): 19–34.
Keilis-Borok VI and Kossobokov VG (1990) Premonitory activation of earthquake flow – Algorithm M8. *Physics of the Earth and Planetary Interiors* 61(1–2): 73–83.
King GCP and Bowman DD (2003) The evolution of regional seismicity between large earthquakes. *Journal of Geophysical Research* 108(B2): 2096.
Kisslinger C (1996) Aftershocks and fault-zone properties. *Advances in Geophysics*, 38: 1–36. San Diego: Academic Press.
Kisslinger C and Jones LM (1991) Properties of aftershock sequences in southern California. *Journal of Geophysical Research* 96(B7): 11947–11958.
Knopoff L (1996) Earthquake prediction: The scientific challenge. *Proceedings of the National Academy of Sciences of the United States of America* 93(9): 3719–3720.
Kossobokov VG, Keilis-Borok VI, Turcotte DL, and Malamud BD (2000) Implications of a statistical physics approach for earthquake hazard assessment and forecasting. *Pure and Applied Geophysics* 157(11–12): 2323–2349.
Krajcinovic D (1996) *Damage Mechanics*. Amsterdam: Elsevier.
Lindman M, Jonsdottir K, Roberts R, Lund B, and Bodvarsson R (2005) Earthquakes descaled: On waiting time distributions and scaling laws. *Physical Review Letters* 94(10): 108501.
Livina V, Tuzov S, Havlin S, and Bunde A (2005a) Recurrence intervals between earthquakes strongly depend on history. *Physica A* 348: 591–595.
Livina VN, Havlin S, and Bunde A (2005b) Memory in the occurrence of earthquakes. *Physical Review Letters* 95(20): 208501.
Lorenz EN (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20: 130–141.
Lyakhovsky V, Ben-Zion Y, and Agnon A (2001) Earthquake cycle, fault zones, and seismicity patterns in a rheologically layered lithosphere. *Journal of Geophysical Research* 106(B3): 4103–4120.
Lyakhovsky V, BenZion Y, and Agnon A (1997) Distributed damage, faulting, and friction. *Journal of Geophysical Research* 102(B12): 27635–27649.

Ma S-K (1976) *Modern Theory of Critical Phenomena*. Reading: W.A. Benjamin.
Main IG (1999) Applicability of time-to-failure analysis to accelerated strain before earthquakes and volcanic eruptions. *Geophysical Journal International* 139(3): F1–F6.
Main IG (2000) A damage mechanics model for power-law creep and earthquake aftershock and foreshock sequences. *Geophysical Journal International* 142(1): 151–161.
Mandelbrot BB (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* 156(3775): 636–638.
Matthews MV, Ellsworth WL, and Reasenberg PA (2002) A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America* 92(6): 2233–2250.
McKernon C and Main IG (2005) Regional variations in the diffusion of triggered seismicity. *Journal of Geophysical Research* 110(B5): B05S05.
Meeker WA and Escobar LA (1998) *Statistical Methods for Reliability Data*. New York: Wiley.
Nakanishi H (1990) Cellular-automaton model of earthquakes with deterministic dynamics. *Physical Review A* 41(12): 7086–7089.
Newman WI and Phoenix SL (2001) Time-dependent fiber bundles with local load sharing. *Physical Review E* 63(2): 021507.
Newman WI and Turcotte DL (2002) A simple model for the earthquake cycle combining self-organized complexity with critical point behavior. *Nonlinear Processes in Geophysics* 9(5–6): 453–461.
Ogata Y (1988) Statistical-models for earthquake occurrences and residual analysis for point-processes. *Journal of the American Statistical Association* 83(401): 9–27.
Ogata Y (1992) Detection of precursory relative quiescence before great earthquakes through a statistical-model. *Journal of Geophysical Research* 97(B13): 19845–19871.
Ogata Y (1999a) Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes. *Journal of Geophysical Research* 104(B8): 17995–18014.
Ogata Y (1999b) Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics* 155(2–4): 471–507.
Ogata Y (2001) Increased probability of large earthquakes near aftershock regions with relative quiescence. *Journal of Geophysical Research* 106(B5): 8729–8744.
Ogata Y (2004) Space-time model for regional seismicity and detection of crustal stress changes. *Journal of Geophysical Research* 109(B6): B06308.
Ogata Y, Jones LM, and Toda S (2003) When and where the aftershock activity was depressed: Contrasting decay patterns of the proximate large earthquakes in southern California. *Journal of Geophysical Research* 108(B6): 2318.
Ogata Y, Matsuura RS, and Katsura K (1993) Fast likelihood computation of epidemic type aftershock-sequence model. *Geophysical Research Letters* 20(19): 2143–2146.
Ojala IO, Ngwenya BT, Main IG, and Elphick SC (2003) Correlation of microseismic and chemical properties of brittle deformation in locharbriggs sandstone. *Journal of Geophysical Research* 108(B5): 2268.
Priezzhev VB, Ktitarev DV, and Ivashkevich EV (1996) Formation of avalanches and critical exponents in an Abelian sandpile model. *Physical Review Letters* 76(12): 2093–2096.
Reasenberg PA and Jones LM (1989) Earthquake hazard after a mainshock in California. *Science* 243(4895): 1173–1176.
Robertson MC, Sammis CG, Sahimi M, and Martin AJ (1995) Fractal analysis of 3-dimensional spatial distributions of earthquakes with a percolation interpretation. *Journal of Geophysical Research* 100(B1): 609–620.
Robinson R (2000) A test of the precursory accelerating moment release model on some recent New Zealand earthquakes. *Geophysical Journal International* 140(3): 568–576.
Robinson R, Zhou SY, Johnston S, and Vere-Jones D (2005) Precursory accelerating seismic moment release (amr) in a synthetic seismicity catalog: A preliminary study. *Geophysical Research Letters* 32(7): L07309.
Rundle J, Klein W, Turcotte DL, and Malamud BD (2000) Precursory seismic activation and critical-point phenomena. *Pure and Applied Geophysics* 157(11–12): 2165–2182.
Rundle JB (1988) A physical model for earthquakes. 1: Fluctuations and interactions. *Journal of Geophysical Research* 93(B6): 6237–6254.
Rundle JB (1989) A physical model for earthquakes. 3: Thermodynamical approach and its relation to nonclassical theories of nucleation. *Journal of Geophysical Research* 94(B3): 2839–2855.
Rundle JB and Klein W (1993) Scaling and critical phenomena in a cellular automation slider-block model for earthquakes. *Journal of Statistical Physics* 72(1–2): 405–412.
Rundle JB, Klein W, and Gross S (1999) Physical basis for statistical patterns in complex earthquake populations: Models, predictions and tests. *Pure and Applied Geophysics* 155(2–4): 575–607.
Rundle JB, Rundle PB, Donnellan A, and Fox G (2004) Gutenberg–Richter statistics in topologically realistic system-level earthquake stress-evolution simulations. *Earth, Planets and Space* 56(8): 761–771.
Rundle JB, Rundle PB, Donnellan A, *et al.* (2005) A simulation-based approach to forecasting the next great San Francisco earthquake. *Proceedings of the National Academy of Sciences of the United States of America* 102(43): 15363–15367.
Rundle JB, Rundle PB, Klein W, *et al.* (2002a) GEM plate boundary simulations for the plate boundary observatory: A program for understanding the physics of earthquakes on complex fault networks via observations, theory and numerical simulation. *Pure and Applied Geophysics* 159(10): 2357–2381.
Rundle JB, Tiampo KF, Klein W, and Martins JSS (2002b) Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting. *Proceedings of the National Academy of Sciences of the United States of America* 99(1): 2514–2521.
Rundle JB, Turcotte DL, Shcherbakov R, Klein W, and Sammis C (2003) Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. *Reviews of Geophysics* 41(4): 1019.
Rundle PB, Rundle JB, Tiampo KF, Martins JSS, McGinnis S, and Klein W (2001) Nonlinear network dynamics on earthquake fault systems. *Physical Review Letters* 8714(14): 148501.
Saichev A, Helmstetter A, and Sornette D (2005) Power-law distributions of offspring and generation numbers in branching models of earthquake triggering. *Pure and Applied Geophysics* 162(6–7): 1113–1134.
Saichev A and Sornette D (2004) Anomalous power law distribution of total lifetimes of branching processes: Application to earthquake aftershock sequences. *Physical Review E* 70(4): 046123.
Saichev A and Sornette D (2005) Distribution of the largest aftershocks in branching models of triggered seismicity: Theory of the universal Båth law. *Physical Review E* 71(5): 056127.
Saleur H, Sammis CG, and Sornette D (1996a) Discrete scale invariance, complex fractal dimensions, and log-periodic fluctuations in seismicity. *Journal of Geophysical Research* 101(B8): 17661–17677.
Saleur H, Sammis CG, and Sornette D (1996b) Renormalization group theory of earthquakes. *Nonlinear Processes in Geophysics* 3: 102–109.

Sammis CG, Bowman DD, and King G (2004) Anomalous seismicity and accelerating moment release preceding the 2001 and 2002 earthquakes in northern Baja California, Mexico. *Pure and Applied Geophysics* 161(11–12): 2369–2378.
Sanchez R, Newman DE, and Carreras BA (2002) Waiting-time statistics of self-organized-criticality systems. *Physical Review Letters* 88(6): 068302.
Scholz CH (1991) Earthquakes and faulting: Self-organized critical phenomena with a characteristic dimension. In: Riste T and Sherrington D (eds.) *Spontaneous Formation of Space Time Structure and Criticality*, pp. 41–56. Norwell, MA: Kluwer.
Scholz CH (2002) *The Mechanics of Earthquakes and Faulting*, 2nd edn. Cambridge: Cambridge University Press.
Shaw BE (1993) Generalized Omori law for aftershocks and foreshocks from a simple dynamics. *Geophysical Research Letters* 20(10): 907–910.
Shcherbakov R and Turcotte DL (2003) Damage and self-similarity in fracture. *Theoritical and Applied Fracture Mechanics* 39(3): 245–258.
Shcherbakov R and Turcotte DL (2004) A modified form of Båth's law. *Bulletin of the Seismological Society of America* 94(5): 1968–1975.
Shcherbakov R, Turcotte DL, and Rundle JB (2004) A generalized Omori's law for earthquake aftershock decay. *Geophysical Research Letters* 31(11): L11613.
Shcherbakov R, Turcotte DL, and Rundle JB (2005a) Aftershock statistics. *Pure and Applied Geophysics* 162(6–7): 1051–1076.
Shcherbakov R, Yakovlev G, Turcotte DL, and Rundle JB (2005b) Model for the distribution of aftershock interoccurrence times. *Physical Review Letters* 95: 218501.
Shebalin P, Keilis-Borok V, Zaliapin I, Uyeda S, Nagao T, and Tsybin N (2004) Advance short-term prediction of the large Tokachi-oki earthquake, September 25, 2003, M = 8.1: A case history. *Planets and Space* 56(8): 715–724.
Sornette D and Helmstetter A (2002) Occurrence of finite-time singularities in epidemic models of rupture, earthquakes, and starquakes. *Physical Review Letters* 89(15): 158501.
Sornette D and Knopoff L (1997) The paradox of the expected time until the next earthquake. *Bulletin of the Seismological Society of America* 87(4): 789–798.
Sornette D and Sammis CG (1995) Complex critical exponents from renormalization-group theory of earthquakes - Implications for earthquake predictions. *Journal de Physique I* 5(5): 607–619.
Sornette D and Werner MJ (2005) Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, Båth's law, and observed aftershock sequences. *Journal of Geophysical Research* 110(B8): B08304.
Stauffer D and Aharony A (1992) *Introduction to Percolation Theory,* 2nd edn. London: Taylor and Francis.
Sykes LR and Jaumé SC (1990) Seismic activity on neighboring faults as a long-term precursor to large earthquakes in the San-Francisco bay area. *Nature* 348(6302): 595–599.
Tiampo KF, Rundle JB, McGinnis S, Gross SJ, and Klein W (2002a) Eigenpatterns in southern California seismicity. *Journal of Geophysical Research* 107(B12): 2354.
Tiampo KF, Rundle JB, McGinnis S, Gross SJ, and Klein W (2002b) Meanfield threshold systems and phase dynamics: An application to earthquake fault systems. *Europhysics Letters* 60(3): 481–487.
Tiampo KF, Rundle JB, McGinnis SA, and Klein W (2002c) Pattern dynamics and forecast methods in seismically active regions. *Pure and Applied Geophysics* 159(10): 2429–2467.
Tsapanos TM (1990) Spatial-distribution of the difference between the magnitudes of the main shock and the largest aftershock in the Circum-Pacific belt. *Bulletin of the Seismological Society of America* 80(5): 1180–1189.
Turcotte DL (1997) *Fractals and Chaos in Geology and Geophysics,* 2nd edn. Cambridge: Cambridge University Press.
Turcotte DL (1999a) Seismicity and self-organized criticality. *Physics of the Earth and Planetary Interiors* 111(3–4): 275–293.
Turcotte DL (1999b) Self-organized criticality. *Reports on Progress in Physics* 62(10): 1377–1429.
Turcotte DL, Newman WI, and Shcherbakov R (2003) Micro and macroscopic models of rock fracture. *Geophysical Journal International* 152(3): 718–728.
Utsu T, Ogata Y, and Matsuura RS (1995) The centenary of the Omori formula for a decay law of aftershock activity. *Journal of the Physics of the Earth* 43(1): 1–33.
Varnes DJ and Bufe CG (1996) The cyclic and fractal seismic series preceding an m(b) 4.8 earthquake on 1980 February 14 near the Virgin Islands. *Geophysical Journal International* 124(1): 149–158.
Vere-Jones D (1969) A note on the statistical interpretation of Båth's law. *Bulletin of the Seismological Society of America* 59: 1535–1541.
Vere-Jones D, Robinson R, and Yang WZ (2001) Remarks on the accelerated moment release model: Problems of model formulation, simulation and estimation. *Geophysical Journal International* 144(3): 517–531.
Ward SN (1992) An application of synthetic seismicity in earthquake statistics: The Middle America Trench. *Journal of Geophysical Research* 97(B5): 6675–6682.
Ward SN (1996) A synthetic seismicity model for southern California: Cycles, probabilities, and hazard. *Journal of Geophysical Research* 101(B10): 22393–22418.
Ward SN (2000) San Francisco Bay Area earthquake simulations: A step toward a standard physical earthquake model. *Bulletin of the Seismological Society of America* 90(2): 370–386.
Yakovlev G, Newman WI, Turcotte DL, and Gabrielov A (2005) An inverse cascade model for self-organized complexity and natural hazards. *Geophysical Journal International* 163(2): 433–442.
Yamanaka Y and Shimazaki K (1990) Scaling relationship between the number of aftershocks and the size of the main shock. *Journal of the Physics of the Earth* 38(4): 305–324.
Yang WZ, Vere-Jones D, and Li M (2001) A proposed method for locating the critical region of a future earthquake using the critical earthquake concept. *Journal of Geophysical Research* 106(B3): 4121–4128.
Zöller G and Hainzl S (2001) Detecting premonitory seismicity patterns based on critical point dynamics. *Natural Hazards and Earth System Sciences* 1(1): 93–98.
Zöller G, Hainzl S, and Kurths J (2001) Observation of growing correlation length as an indicator for critical point behavior prior to large earthquakes. *Journal of Geophysical Research* 106(B2): 2167–2175.

Relevant Website

http://www.data.scec.org – Southern California Earthquake Data Center.