CHAPTER
FOUR

Mechanics of earthquakes

Friction of faults is often unstable, and slip occurs rapidly as a rupture dynamically propagates over the fault surface. These sudden motions generate seismic waves, and this is the mechanism of the most common and important type of earthquake. Seismicity is thus the short-timescale phenomenon of brittle tectonics. In this chapter we discuss the dynamics of faulting and review the most important attributes of earthquakes from the point of view of the rupture process.

4.1 HISTORICAL DEVELOPMENT

During most of human history, people's notion of the origin of earthquakes lay within the realm of mythology. Several of the schools of ancient Greek philosophy considered earthquakes to be natural phenomena, although their speculations on the matter relied heavily on imagination and do not bear much relationship with modern theories (see Adams (1938) for an excellent historical account of thinking on this topic from the Greeks up through the Renaissance). It was not until the middle and latter part of the nineteenth century that instrumental measurements began to be made and note taken of the geological associations of earthquakes. Lyell (1868) considered earthquakes to be an important agent in Earth dynamism, and was aware of both faulting and permanent changes in elevation brought about by them. Although Lyell carefully described the faulting and deformation produced in several earthquakes, like his contemporary Mallet, Lyell believed that the immediate cause of earthquakes was thermal, either a consequence of volcanic activity or thermal expansion or contraction.

The first clear connection between earthquakes and dynamic faulting and its relationship to tectonic processes was made by G. K. Gilbert. He had seen the immediate aftereffects, including the surface faulting, of the 1872 Owens Valley earthquake in California, and in his extensive

Fig. 4.1. The famous photograph by Koto (1893) of the scarp of the 1891 Nobi (Mino–Owari) earthquake, Japan. A more recent photograph of the same scene maybe found in Bolt (1978, p. 41). Faulting in this location was mainly normal, with the north side up.

mapping in the Great Basin had also remarked on the fresh–appearing scarps that so often front the mountain ranges there. He concluded that the elevations of the mountains were produced by repeated sudden ruptures along these faults. Thus, he stated (Gilbert, 1884):

> The upthrust produces a local strain in the crust, involving a certain amount of compression and distortion, and this strain increases until it is sufficient to overcome the starting friction along the fractured surface. Suddenly, and almost instantaneously, there is an amount of motion sufficient to relieve the strain, and this is followed by a long period of quiet, during which the strain is gradually reimposed. The motion at the instant of yielding is so swift and so abruptly terminated as to cause a shock, and the shock vibrates through the crust with diminishing force in all directions. . . In this region a majority of the mountain ranges have been upraised by a fracture on one side or the other, and in numerous instances there is evidence that the last increase of height was somewhat recent.

Gilbert claimed only that this theory of earthquakes applied to the Great Basin, but similar connections between earthquakes and faulting soon were made elsewhere. McKay (1890) journeyed to the site of a large earthquake of two years previous in the South Island of New Zealand and discovered there a fresh strike–slip fault scarp on the Hope fault. Soon afterwards, a great oblique–slip scarp was found at the site of the 1891 Nobi earthquake in Japan (Koto, 1893; see Figure 4.1). Koto discussed at some length the debate among European geologists as to whether faulting was the cause or effect of earthquakes, quoting extensively from Lyell, and argued cogently for the faulting origin hypothesis. He was evidently unaware of Gilbert's views on the subject. The extensive rupturing of the San Andreas Fault during the 1906 San Francisco

earthquake, and the geodetic measurements that showed that this ground breakage was not superficial, finally led to the dominance of the faulting theory of earthquakes, as expressed in the analysis of that earthquake by Reid (1910).

That the vast majority of shallow tectonic earthquakes arise from faulting instabilities was proven eventually by seismological observations, but this occurred only after a long delay. Although the theory of radiation from a double–couple source was introduced first by Nakano in 1923, there was slow progress in implementing this development. The determination of an earthquake focal mechanism from its radiated field requires both substantial computation and a widely distributed network of standardized seismometers, so progress awaited developments in instrumentation and computers.

Also, scientific opinion often was divided on the subject. Surface faulting can be observed for only a very small fraction of earthquakes, namely those large earthquakes that occur on land; even then, faulting often may be obscured by heavy vegetation, so it was quite possible to deny this as a general mechanism. There was also a great debate about whether the double–couple or the single–couple is the correct representation of earthquakes (see Kasahara [1981] for a recounting of this issue). In retrospect this argument seems futile because the single–couple does not connect two equilibrium states and hence is not physically possible.

The modern era of earthquake source studies began with the installation of the Worldwide Standardized Seismic Network in the early 1960s and with the widespread use of computers. It was only then that dynamic faulting gained widespread acceptance as the origin of the majority of seismic events.

4.2 EARTHQUAKE PHYSICS

In this section an account of the basic physics of earthquakes is presented. An excellent introduction to this topic is given by Kanamori and Brodsky (2004). Here we give an abbreviated treatment in which the rupture aspects are emphasized and less is said about the radiation of seismic waves.

4.2.1 The dynamic energy balance

An earthquake may be considered to be a dynamically running shear crack, so we begin our discussion, as in Chapter 1, with the energy balance for this process. The energy balance, Equation (1.6), may be rewritten by the inclusion of terms for the kinetic energy and the frictional work done on the crack surface behind the tip.

An earthquake produces a sudden slip of average amount $\Delta\overline{u}$ over a rupture area A. This results in a drop of stress from an initial value $\sigma_1$to a final value $\sigma_2$where $(\sigma_1 - \sigma_2) = \Delta\sigma_s$ is called the static stress drop and $(\sigma_1 + \sigma_2)/2 = \sigma_m$, the mean stress. All these quantities are quite variable over the rupture surface, perhaps fractally so, and so these values are the spatial averages.

The linear measure of an earthquake is the seismic moment (Maruyama, 1963),

$$M_o = \mu\Delta\overline{u}A. \tag{4.1a}$$

This is the scalar magnitude of the seismic moment tensor

$$M_{oij} = \mu(\Delta u_i n_j \Delta u_j n_i)A \tag{4.1b}$$

where Δu_i is the mean slip vector averaged over the rupture area A with unit normal n_i. We use μ in this case to denote shear modulus rather than friction coefficient. This usage should be apparent from its context. M_o is routinely measured from the seismic radiation or geodetic data (Lay and Wallace, 1995), and earthquake size is reported using a magnitude scale that is based on moment,

$$M_W = \frac{log_{10} M_o}{1.5} - 6.07 \quad (M_o \text{ in N m}) \tag{4.2}$$

Rupture area A can be estimated by various means: inversions of seismic or geodetic data, aftershock area for large earthquakes, the corner frequency of the body wave spectrum for smaller earthquakes, or the rupture length for large continental earthquakes. Stress drop is related to strain drop by

$$\Delta\sigma_s = C\mu \frac{\Delta\overline{u}}{\widetilde{L}} \tag{4.3}$$

where C is a geometric constant of order one and $\widetilde{L}$ is a characteristic length scale of the rupture. From measurements of A and M_o, $\Delta\overline{u}$ can be estimated. $\widetilde{L}$ can be approximated as $A^{1/2}$, and $\Delta\sigma_s$ can then be calculated. We can correspondingly express $\Delta\sigma_s$ as

$$\Delta\sigma_s = CM_o A^{-3/2} = CM_o \widetilde{L}^{-3} \tag{4.4}$$

The dynamic energy balance is

$$W_F = E_R + E_F + E_G \tag{4.5}$$

where the work of faulting, $W_F = \sigma_m \Delta\overline{u} A$, is the net potential energy drop. (This neglects gravitational work, which may be important in dip–slip faulting [Savage and Walsh, 1978; Walsh and Rice, 1979].) The frictional work $E_F = \sigma_F \Delta\overline{u} A$, where σ_F is the average frictional stress resisting sliding. E_R is the energy radiated in seismic waves and E_G is surface energy and the energy loss resulting from other types of permanent damage in forming the rupture. E_R can be estimated from seismic waves (e. g. Boatwright and Choy, 1986) or from an empirical relation to moment.

The work of faulting can be expressed as

$$W_F = \frac{\sigma_1 + \sigma_2}{2} \Delta\overline{u} A = \frac{\sigma_1 - \sigma_2}{2} \Delta\overline{u} A + \sigma_2 \Delta\overline{u} A \tag{4.6}$$

The first term,$(((\sigma_1 - \sigma_2)/2)\Delta\overline{u}A)$, can be determined by seismological means as outlined above, whereas the second term, $\sigma_2 \Delta\overline{u} A$, cannot because it requires knowledge of the absolute value of stress.

Seismic waves are responsible for the damage caused by earthquakes and hence it is of primary importance to characterize the relative importance of E_R to other quantities. We define the seismic efficiency η as

$$\eta = \frac{E_R}{W_F} \tag{4.7}$$

which cannot be determined because, as mentioned above, W_F depends on the absolute value of stress. We can also define

$$\zeta = \frac{E_R}{M_o} = \frac{1}{\mu}\frac{E_R}{\Delta\overline{u}A} \tag{4.8}$$

which defines a scaled energy per unit slip and area. The radiation efficiency, η_R is defined as

$$\eta_R = \frac{E_R}{E_R + E_S} \tag{4.9}$$

The quantity $\sigma_A = \mu\zeta = \sigma_m\eta$ is called the apparent stress (Aki, 1966; Wyss and Brune, 1968). It may be shown (Savage and Wood, 1971), neglecting surface energy losses, that

$$\sigma_A = \sigma_m - \sigma_F = \frac{(\sigma_1 - \sigma_2)}{2} - (\sigma_F - \sigma_2) \tag{4.10}$$

and, because we assume that friction stops the fault sliding, $\sigma_F > \sigma_2$. It is also clear that $\sigma_F < \sigma_m$, therefore

$$0 < \sigma_A < \frac{(\sigma_1 - \sigma_2)}{2} \tag{4.11}$$

The minimum value occurs when $(\sigma_1 - \sigma_2) = 2(\sigma_1 - \sigma_F)$, and $E_R = 0$, which corresponds to the simple case we explored in Section 2.3.5. At the other extreme, $\sigma_2 = \sigma_F$. This represents full radiation damping and yields the maximum E_R. The apparent stress can thus be seen as a measure of radiation resistance. In the maximum case the driving stress does not fall below the friction level in order to stop sliding, contrary to what was illustrated in Figure 2.16 – the motion is instead damped by the radiation. Inserting this result in the energy balance (Equation (4.4)) we find a simplified version

$$E_R = \frac{(\sigma_1 - \sigma_2)}{2}\Delta\overline{u}A - E_G \tag{4.12}$$

Savage and Wood showed that the inequality (Equation (4.11)) is generally true – that is, the apparent stress is always less than half the stress drop, i.e. less than the maximum value. In a dynamic model of a circular crack with friction, Madariaga (1976) found that the driving stress overshot the frictional stress by 20%.

Because Equation (4.12) reflects the maximum value of E_R, the radiation efficiency is

$$\eta_R < \frac{\frac{(\sigma_1-\sigma_2)}{2}\Delta\overline{u}A - E_G}{\frac{(\sigma_1-\sigma_2)}{2}\Delta\overline{u}A} \tag{4.13}$$

The relative value of E_G is still under debate (see Section 3.2.2). Estimated values, obtained from analyzing fault gouge, range from negligible to major contributions to the frictional work.

In illustrative models it is often assumed, for simplicity, that $\sigma_2 = \sigma_F$, despite the issues just raised. A simple model of that type is shown in Figure 4.2a, representing some point within the earthquake rupture. Stress arising from the stress concentration of the approaching rupture front increases to a value σ_y, above the initial value σ_1. The frictional resisting stress on the fault follows the curved friction weakening path from σ_y to B, where it reaches $\mu_D\sigma_n = \sigma_2$. The shaded area under that curve represents E_G, which includes surface energy as well as other permanent deformation processes. Note that the energy balance implicitly dictates that this scales with rupture area, so that it must include permanent deformation occurring during

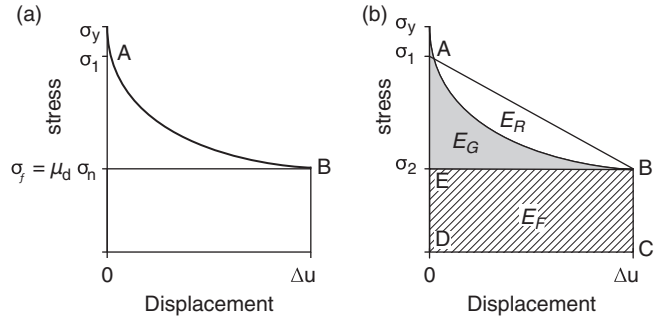

Fig. 4.2. A diagram showing the energy balance for a simple illustrative case. (a): a slip-weakening model in which the friction follows the path AFB. (b): a diagram showing the energy balance for a slip event with slip-weakening friction.

sliding, such as plastic deformation at the contact scale, chemical reactions such as calcining in the case of carbonate faults, and damage over the fault surface (wear).

The diagram for a slip event is shown in Figure 4.2(b). The area of the entire figure, ABCDA, is W_F. The applied stress follows the elastic unloading line AB. ABE is the first term on the RHS of Equation (4.6) and is composed of E_G plus E_R, as indicated in Equation (4.12). The rectangle EBCDE is the second term in Equation (4.6) and represents the frictional work that is lost in heat.

More realistic cases are discussed by Kanamori and Rivera (2006). For example, overshoot is likely, so that $\sigma_2 < \sigma_F$. The assumption that all damage occurs in the crack–tip region is also oversimplified, as was discussed in Section 3.2.2.

4.2.2 Dynamic rupture propagation

Earthquakes are often modeled with kinematic models in which the displacement history of motion is prescribed with some suitably few parameters. Often–used models of this type are the propagating dislocation model of Haskell (1964) and the Brune (1970) model, the latter of which, though employing an infinite rupture velocity, has the advantage of being rationalized in terms of the dynamic properties of the source. While such models may provide quite detailed descriptions of earthquakes they do not yield a physical insight into the rupture process itself, which is of prime concern here. For this purpose, we need to examine some of the results of dynamic models, by which is meant models that satisfy the dynamic equations of elasticity with the only prescription being that of a fracture criterion. We only discuss some pertinent results of the simplest models of this type. For a more rigorous and detailed mathematical treatment see, for example, Aki and Richards (2002), Kostrov and Das (1988), and Freund (1990).

The mechanism of rupture in unstable slip has so far been described in two different ways: as brittle fracture in Chapter 1, and as a stick–slip friction instability in Chapter 2. The two are mathematically equivalent in relating motion in the medium to a drop in shear stress on the fault surface, but have traditionally differed in the way in which the rupture process is considered. In theoretical fracture mechanics it is assumed that a characteristic fracture energy per unit area, a material property, is required for the crack to propagate. In the stick–slip model, on the other hand, rupture is assumed to occur when the stress on the fault reaches the static friction value and the condition for dynamic instability exists. In the fracture model, stresses at the crack tip may be arbitrarily high, but in the stick–slip model no energy is dissipated in the

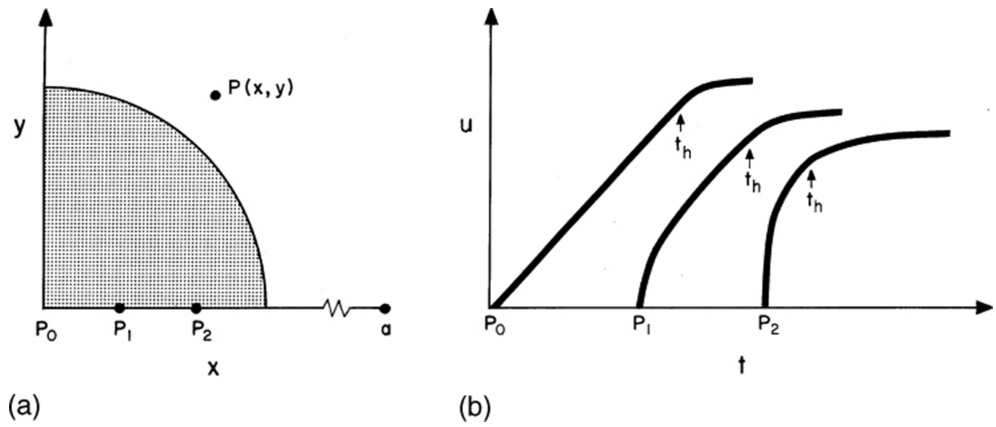

Fig. 4.3. Slip history at a point P: (a) The rupture (stippled) approaches P. (b) Slip history at three points P_0, P_1, and P_2 at successively greater distance from the point of rupture initiation. T_h denotes the arrival of a healing wave propagating back from the final perimeter of the rupture.

crack tip and the stresses there must remain finite. These differences mainly have arisen historically out of the idealizations involved in the two approaches. Experimental studies show that rupture in stick–slip friction propagates in the same manner as an elastic crack (Svetlizky and Fineberg, 2014). Here we summarize some of the features of simple dynamical rupture models.

Rupture initiates when the applied stress reaches the yield strength σ_Y at a point on the fault surface called the hypocenter. The rupture tip then propagates outwards at a rupture velocity V_R. At all other points the yield strength is greater than the applied stress, but can be overcome by the stress concentration in front of the rupture tip, which grows with rupture length x according to the stress–intensity factor $K = \Delta\sigma_D x^{1/2}$, where $\Delta\sigma_D = \sigma_1 - \sigma_F$. At a point at a distance P_1 from the hypocenter, as shown in Figure 4.3, slip initiates at a high particle velocity $v_{max} \propto K$, which decreases rapidly with K (as the rupture tip recedes from the point) to an asymptotic value proportional to $\Delta\sigma_D$. In a crack model, when the crack stops propagating at the final rupture perimeter, the information on the finiteness of the rupture propagates back into the rupture interior in a healing phase. When this information arrives at each point at time t_h, slip at that point begins to decelerate to rest. The stress history of such a point is shown in Figure 4.4(a). The behavior shown in Figure 4.4(a) may be considered typical of a point where the friction is velocity weakening.

If, on the other hand, the rupture encounters a point characterized by velocity strengthening, the behavior shown in Figure 4.4(b) would occur. Stress would rise after slip is initiated, resulting in a negative stress drop. This would produce an energy sink that will tend to impede further rupture propagation. Following cessation of slip, the stress would be higher than initially, and quasistatic recovery would occur in the form of afterslip.

The rupture velocity V_R depends upon the energy expended in propagating the crack. In the simplest case of pure shear, from an energy balance approach due initially from Mott (1948) (see also Lawn [2010] and Kanamori and Brodsky [2004]) we have,

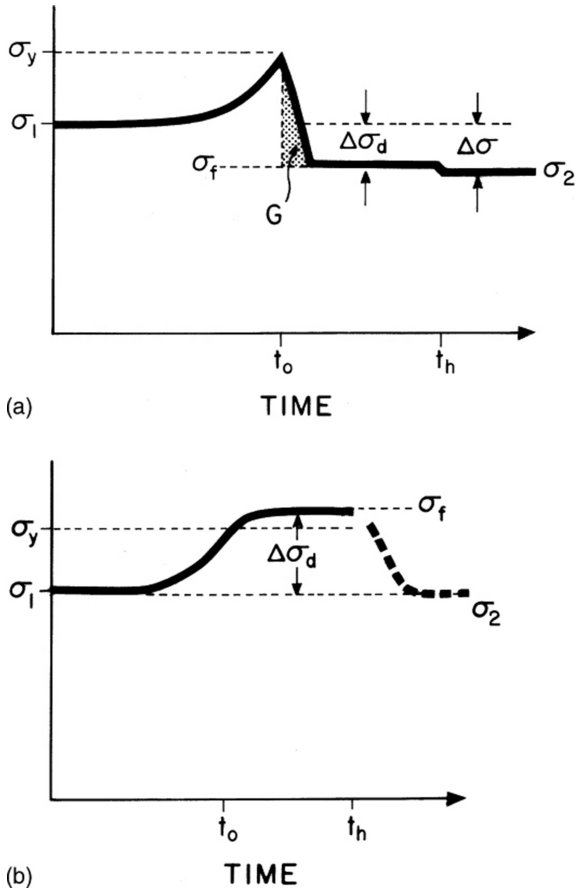

Fig. 4.4. Stress history during rupture at points that exhibit: (a) velocity weakening, (b) velocity strengthening. The dashed curve is postseismic relaxation (afterslip).

$$\eta_R \propto \frac{V_R^2}{\beta^2} \tag{4.14}$$

so that V_R approaches the limiting value of the shear velocity β as E_G approaches zero. An increase in the relative value of E_G results in a slower rupture velocity. A typical value for earthquakes is $\sim 0.75\beta$ (Kanamori and Brodsky, 2004).

Rupture velocity is limited by the speed at which stresses can propagate, which in the case of Mode III cracks, where the crack–tip stresses are pure shear, this is β, the shear wave velocity, although, since the propagation is along preexisting surfaces, it is C_R, the Rayleigh wave velocity. For Mode II cracks, normal stresses are present in the crack–tip stress field. In cases where the rupture resistance, defined as a dimensionless strength

$$S = \frac{(\sigma_y - \sigma_1)}{(\sigma_1 - \sigma_F)} \tag{4.15}$$

is lower than about 1.2 (Dunham, 2007), slip may be induced by those components, resulting in a transonic rupture velocity between β and the P wave velocity α. An example, from a numerical model, is shown in Figure 4.5. Notice the steepening of the rupture front with time due to the

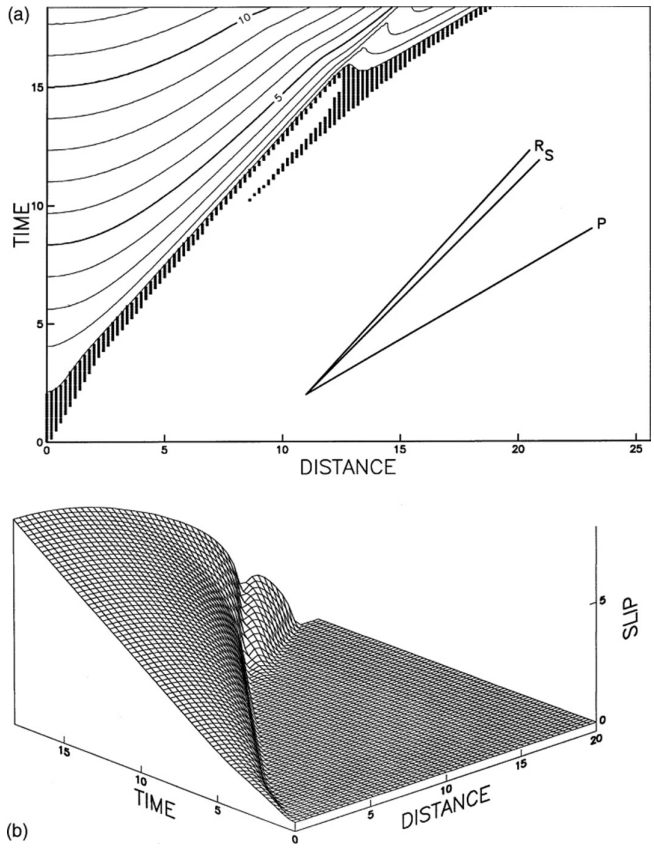

Fig. 4.5. Propagation of a Mode II crack with slip weakening and low rupture resistance. (a) Slip contours of dimensionless slip during propagation of the crack. The shaded region indicates the breakdown area in which slip is less than d_0. Lines indicate the P, S, and Rayleigh velocities. (b) Mesh perspective of the dimensionless slip in the same coordinates. Azimuth of view is in the Rayleigh wave direction. (From Andrews, 1985.)

Lorenz contraction of the stress field. Rupture propagates at the Rayleigh velocity until the stress–intensity factor becomes large enough that the normal stresses can initiate slip, which produces a precursor that propagates at the P wave velocity. Transonic rupture velocities have been observed in laboratory experiments (Rosakis et al., 1999; Samudrala et al., 2002) and for earthquakes (e.g. Bouchon et al., 2001). This so–called "supershear" earthquake propagation appears to be restricted to large strike–slip (Mode II) earthquakes, on relatively smooth and straight fault sections, as expected from the theory (Bouchon et al., 2010).

Pulse-like, crack-like, and avalanche-like rupture propagation

Heaton (1990) pointed out, based on kinematic models available at the time, that the rise time (the slip duration at a point) in earthquakes was considerably shorter than either the total source duration or the time required for the arrival of a healing phase to terminate slip, as sketched in Figure 4.3. He proposed that the rupture propagated as a self–healing pulse. An

alternative explanation is fault heterogeneity, in which local asperities control the local slip duration (Beroza and Mikumo, 1996). There are several ways that pulse–like propagation can be produced in models: with extreme velocity–weakening (Cochard and Madariaga, 1996), with strong weakening from narrow shear–band formation (Daub et al., 2010), or with thermal weakening of the type discussed in Section 2.2.5 (Bizzarri, 2010; Noda et al., 2009). They can also arise as "wrinkle pulses" in sliding on a bimaterial interface when a Mode II rupture propagates in the slip direction of the more compliant material (Ampuero and Ben–Zion, 2008; Shi and Ben–Zion, 2006; Shlomai and Fineberg, 2016; Weertman, 1980). The difference in the rupture modes is illustrated in Figure 4.6. Supershear crack propagation, in the top panel, is characterized by the forerunning rupture initiation pulse, propagating at the *P* wave velocity, as was seen in Figure 4.5. The pulse–like mode, shown in the lower panel, differs from the crack modes in that slip stops in the interior some distance behind the rupture front, as shown in the lower right figure. The slip history at a point is shown in Figure 4.7 for laboratory models of crack–like propagation (upper curve) and pulse–like propagation for the two lower curves. In the latter two curves, the cessation of slip is marked by circles. Kinematic inversion models can now distinguish pulse–like from crack–like propagation in simple cases (Konca et al., 2013), but are defeated in the presence of strong heterogeneity. A problem with the pulse–like mode interpretation is that earthquake scaling laws indicate that slip scales with rupture dimensions, which is expected from crack models and difficult to explain with pulse models.

Rupture models in which strong heterogeneity in strength and/or stress dominates rupture propagation in which the failures of nearby elements are triggered by elastic interactions are called avalanche models (Ben–Zion and Rice, 1993; Dahmen et al., 2011; Fisher, 1998; Mehta et al., 2006). These models predict a triangular moment rate–time function with a symmetrical peaked form that scales with the event duration. This most closely resembles the source time function of large subduction earthquakes (Meier et al., 2017).

4.3 EARTHQUAKE PHENOMENOLOGY

4.3.1 Earthquake scaling relations

The analysis of dynamic rupture models and of the block–slider model (Section 2.3.5) shows that the dynamic characteristics of rupture, slip, velocity, and acceleration all scale linearly with stress drop, which is therefore the single most fundamental scaling parameter. In principle, measurements of any of these observables can be inverted for stress drop, but this turns out to be difficult in practice.

Geodetic and/or fault slip and length data provide information on static stress drop, and seismic measurements of slip, velocity, and acceleration on dynamic stress drop. Stress drop is defined at a point, but the inversions always yield integral averages over some ill–defined region of the fault. There are several reasons for this. First, in the general case of a heterogeneous distribution of stress drop on the fault, the slip history of any one point will be a function of the distribution of stress drop over all nearby points, rather than of one point, and both the functional dependence and the definition of "nearby" are not known. Second, the observed seismic waveform is an integral property of the slip distribution, and measurements of the slip parameters are themselves averages: temporal in the case of dynamic, and spatial in the case of static measurements. This will result, in the case of dynamic measurements, in the determination being a function of the period and type of seismic waves utilized. Finally, any such inversion is highly model–dependent, because a model must be used to relate radiation to

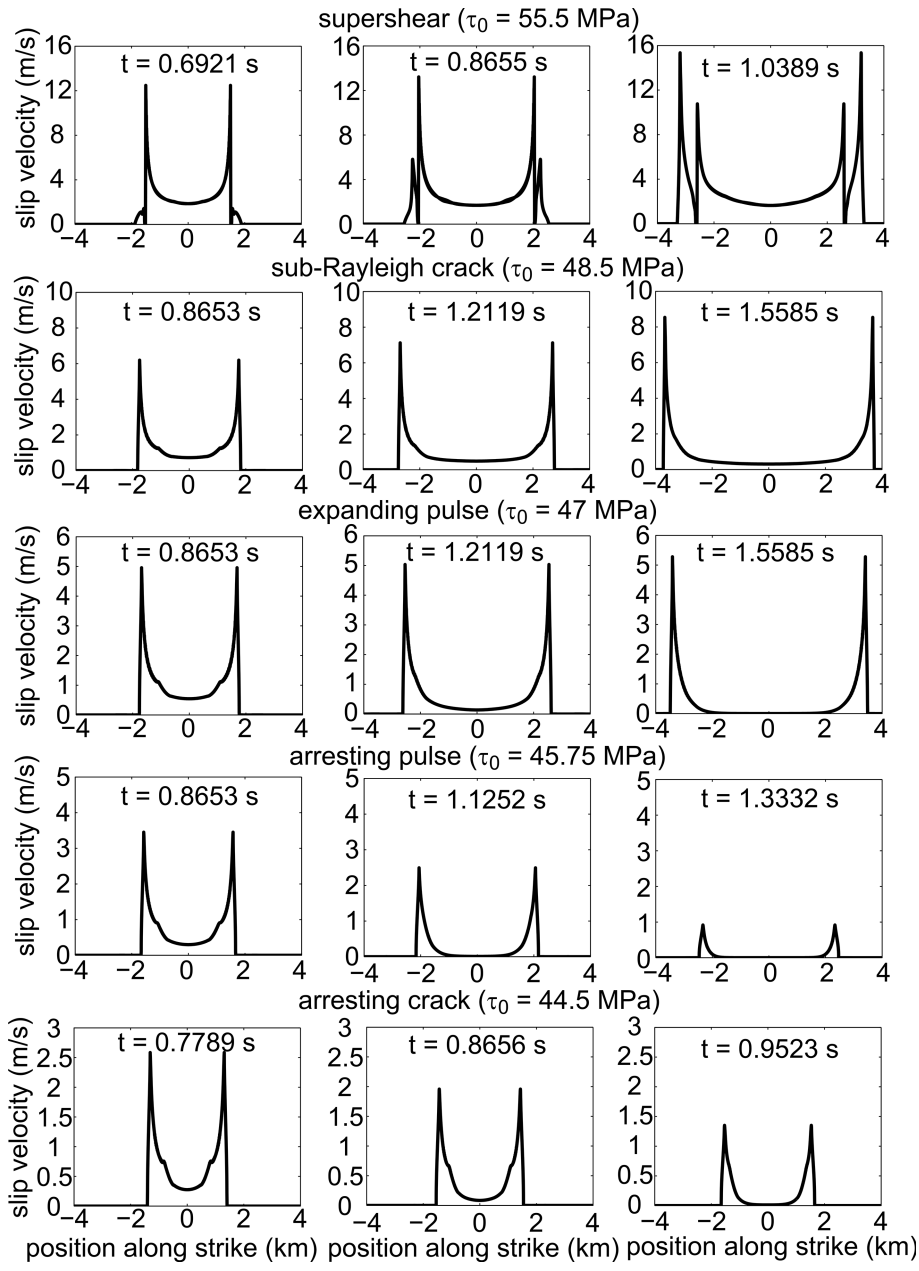

Fig. 4.6. Illustration of the three rupture modes as plots of slip velocity versus distance for three time increments: top panel, supershear crack; middle panel, sub-Rayleigh crack; and bottom panel, pulse-like. (From Daub et al., 2010.)

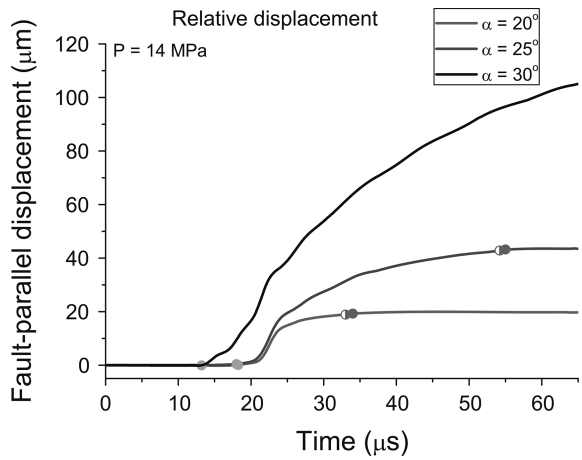

Fig. 4.7. Slip histories at a point from laboratory models of crack-like (top curve), and pulse-like rupture propagation (bottom curves). (From Lu et al., 2010.)

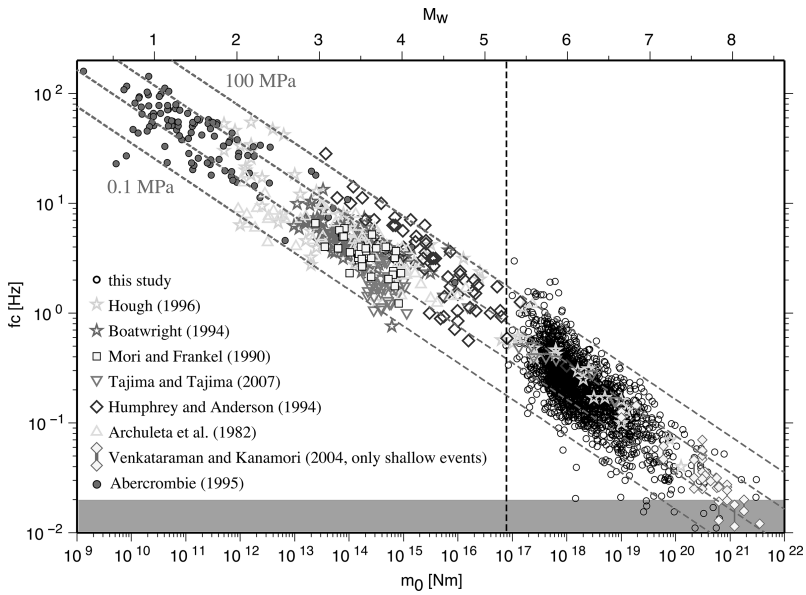

Fig. 4.8. Corner frequency versus moment for shallow earthquakes. The dashed lines show constant stress-drop scaling for 0.1, 1,10, and 100 MPa. (From Allmann and Shearer, 2009.)

source parameters, and simplifying assumptions have to be made regarding the geometry of the source.

However, within a given method of measurement, stress drop is scale–invariant, as shown in Figure 4.8. There, corner frequency is plotted versus moment for shallow earthquakes (Allmann and Shearer, 2009). The dashed lines are for constant stress–drop scaling for values of 0.1, 1,

Table 4.1. Stress parameters for earthquakes by tectonic type

Earthquake type	Stress drop, MPa	Apparent stress, MPa
Subduction Thrust	2.00	0.70
Continental Collision	2.63	0.43
Continental Transform	3.54	1.50
Oceanic Transform	6.03	4.48
Oceanic Intraplate		6.95
Oceanic Ridge	2.82	0.25
Tsunami Earthquakes	0.80	0.10

Data sources: Subduction, Denolle et al. (2016). Tsunami, Ye et al. (2016a). Other stress drops, Allman and Shearer (2009). Other apparent stresses, Choy and Boatwright (1995).

10, and 100 MPa. Thus, although stress drop varies over three orders of magnitude, it is scale–independent.

Apparent stress, like stress drop, is scale–independent (Ide and Beroza, 2001; Ye et al., 2016a). These two parameters vary with tectonic environment, as discussed in Chapter 6 and as shown in Table 4.1.

The scale invariance of stress drop and apparent stress with earthquake size has important ramifications regarding the application to earthquakes of the thermal weakening of friction observed in the high–speed friction experiments described in Section 2.2.5. Such thermal weakening requires a critical slip distance before it is manifested (Figure 4.14(a)), which at seismogenic depths is several tens of centimeters. This implies that thermal weakening should gradually become apparent for earthquakes within the magnitude range of about 4.5–5.5, where a dramatic increase in stress drop should be observed. However, no such change in stress drop is observed over any scale range in Figure 4.8. Kanamori and Heaton (2000) interpreted a change in the scaling of apparent stress over that magnitude range as indicating the onset of thermal weakening, but the more recent results of Ide and Beroza (2001) showed that no such scaling break in apparent stress exists. Thus, there is no verification from seismological data that such pronounced thermal weakening occurs in earthquakes.

If we take a circular rupture as a simple model of earthquakes, the relation (Equation (4.3)) is known from the elastic solution for a circular crack of diameter L, and

$$M_0 = \frac{2}{7}\Delta\sigma L^3 \tag{4.16}$$

If we also assume that the rupture propagates at a steady velocity V_R, then the rupture duration $T_R \approx L/V_R$. This leads to two scaling relations (Kanamori and Anderson, 1975; Kanamori and Brodsky, 2004):

$$M_o \propto T_R^3 \tag{4.17}$$

and

$$M_o \propto L^3 \tag{4.18}$$

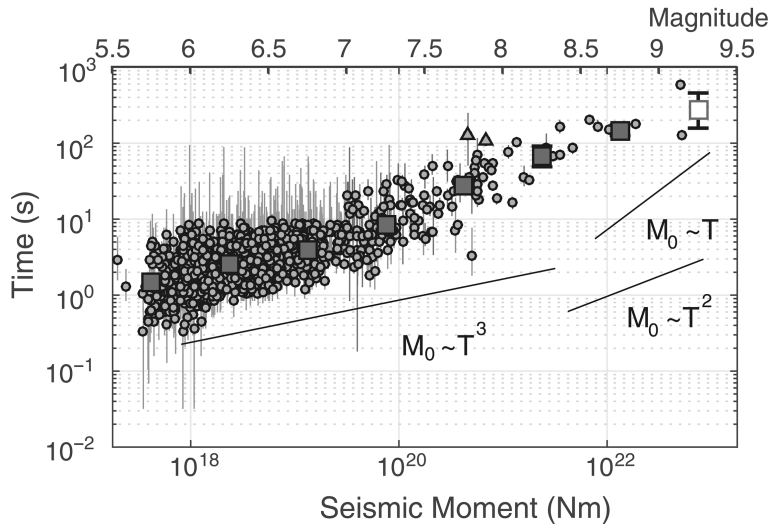

Fig. 4.9. The relation between the seismic moment M_0 and seismic source duration, assumed to be equal to rupture duration T_R for shallow subduction earthquakes. The equivalent M_W is shown at the top of the figure. (From Denolle and Shearer, 2016.)

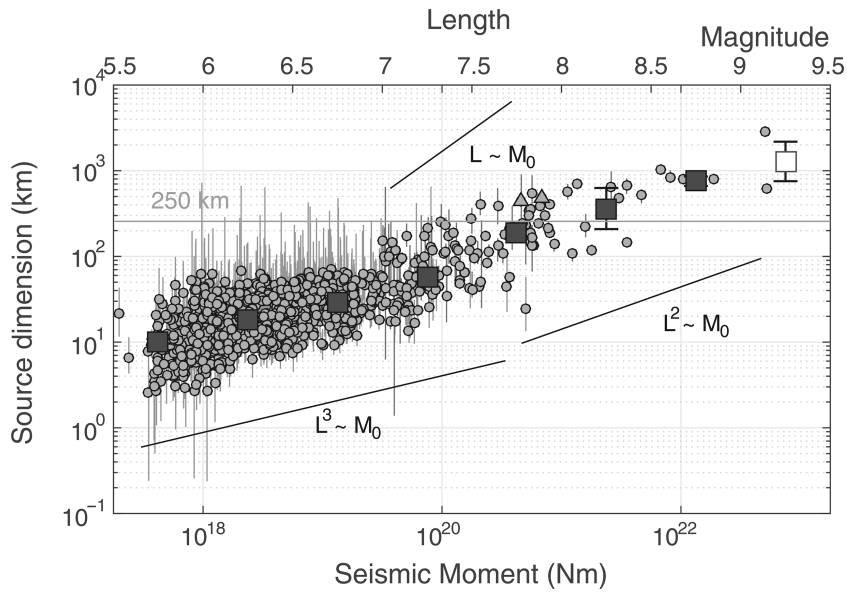

Fig. 4.10. The scaling relation between seismic moment M_0 and rupture diameter L for shallow subduction earthquakes. (From Denolle and Shearer, 2016.)

in which the scaling parameter in both cases is stress drop. In Figures 4.9 and 4.10 moment is plotted versus duration and length for a large dataset of shallow subduction earthquakes (Denolle and Shearer, 2016). We see that the scaling laws expected from the circular crack model apply up to about M_W 7.5, above which they transition to scaling laws with exponent of about 2.

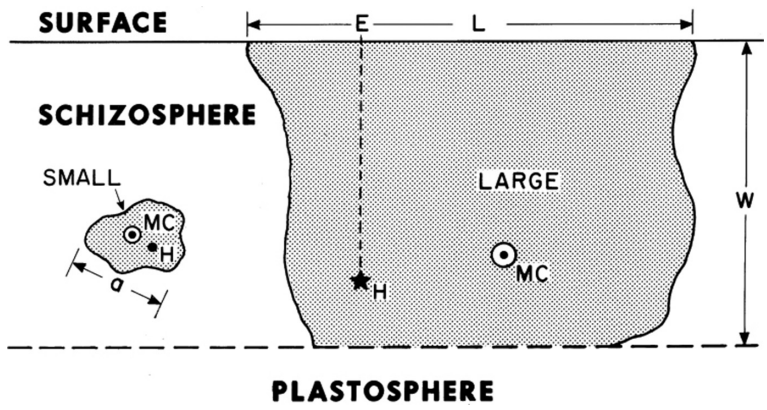

Fig. 4.11. Diagram illustrating the definitions of small (subseismogenic) and large (supraseismogenic) earthquakes, showing hypocenter (H), epicenter (E), moment centroid (MC), and the dimensions of rupture (*a*, *L*, and *W*).

The transition from cubic to squared scaling occurs when the width of the rupture becomes restricted to the width of the seismogenic window, which ranges from 50 to 250 km for subduction zones. Above that point the circular crack approximation no longer applies because the rupture width is held constant while the rupture continues to grow in length. It thus becomes necessary to divide earthquakes into two classes, called simply *large* earthquakes and *small* earthquakes. Small earthquakes are all those events whose rupture dimensions are smaller than the width W^* of the schizosphere, as shown in Figure 4.11. They therefore propagate and terminate entirely within the bounds of the schizosphere, and their behavior may be described as rupture in an unbounded elastic brittle solid. A large earthquake, in contrast, is one in which a rupture dimension equals or exceeds the width of the schizosphere. Once an earthquake becomes large, it is constrained to propagate only horizontally, with its aspect ratio increasing as it grows: its top edge at the free surface, and its bottom at the base of the schizosphere. There are two reasons for making this distinction. The first, motivated by the above observations, is that small and large earthquakes, so defined, are expected to obey different scaling relationships. The second is that we generally need consider only large earthquakes when quantitatively considering the role of earthquakes in tectonics. Note that the magnitude level where earthquakes change from small to large depends on the tectonic environment. For crustal faults like the San Andreas Fault, say, where the schizospheric width is only about 15 km, this occurs at about $M = 6.0$, whereas in subduction zones, where the down-dip width of the schizosphere is much greater, in the range of 50–200 km, it is about $M = 7.5$.

When we restrict ourselves to continental earthquakes, where W^* is 15–20 kms, a similar breakdown in scaling between small and large earthquakes becomes apparent. Figure 4.12 shows moment magnitude for continental strike-slip earthquakes plotted versus their area (Hanks and Bakun, 2002; 2008), based on an updated version of the Wells and Coppersmith (1994) data set. There is a clear break in scaling at about $A = 500$ km^2, or $L \sim 2\,W^*$, for $W^* = 15$ km. For earthquakes below the break, the correlation is $M_W = \log A + 3.98$. This relation, as can be seen from combining Equations (4.2) and (4.16), indicates that small earthquakes can be considered circular ruptures with constant stress drops of ~2.7 MPa. For large earthquakes the relationship is $M_W = 4/3 \log A + 3.07$. This relationship is consistent with observation that

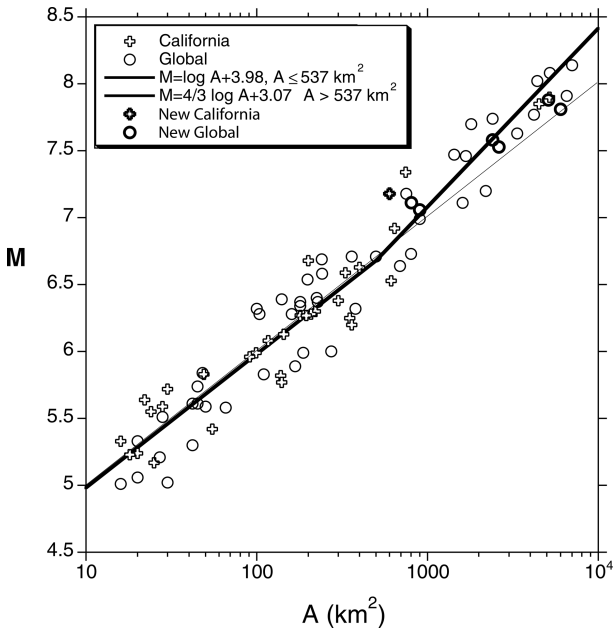

Fig. 4.12. A plot of moment magnitude versus area for continental strike-slip earthquakes. The data is from the compilation of Wells and Coppersmith (1994) with updates of more recent large earthquakes. Thick lines are the bilinear relations of Hanks and Bakun (2002), discussed in the text. Thin line is the linear relation of Wells and Coppersmith (1994). (From Hanks and Bakun, 2008.)

$\Delta\overline{u} \propto L$ and $M_o = \mu\alpha L^2 W^*$ (Scholz, 1982; Shimazaki, 1986) and the fit to the data gives $\alpha = 2.3 \times 10^{-5}$ (Hanks and Bakun, 2002; 2008). Figure 4.13 shows mean slip versus rupture length for large continental earthquakes. This plot shows the $M_o \propto L^2$ regime just described, and also reveals the existence of a third scaling regime for earthquake with extremely long aspect ratios, $L>10\ W^*$. In that regime, mean slip reaches a constant value as expected from the elastic solution of infinitely long strike–slip or dip–slip fault (Knopoff, 1958), in which $\widetilde{L}$ in Equation (4.3) is W^*. This implies that for constant stress–drop scaling, slip in such earthquakes is also constant and that $M_o \propto (W^*)^2 L$. These three scaling regimes are listed in Table 4.2.

The interpretation of the $M_o \propto L^2$ scaling regime for subduction earthquakes, seen at $M_W > 7.5$ in Figure 4.10, is not so straightforward, because W^* is not constant for subduction zones. Denolle and Shearer 2016 adopted the scaling relation of $W \propto L^{2/3}$ (Leonard, 2010). In that case, if one makes the usual assumption that $\Delta u \propto W$, then $M_o \propto L^{7/3}$, which would be indistinguishable from the L^2 case. That scaling is hard to reconcile with the idea that W is set by the local value of W^* and that L is set by other factors such as the history of prior earthquakes. We know that great subduction zone earthquakes can be both nearly equidimensional, such as Tohoku–oki 2011, and be long and narrow, such as Sumatra 2004.

The $\Delta u \propto L$ scaling of regime 2 is difficult to reconcile with the constant stress–drop scaling of the other regimes. One suggestion is that large earthquakes progressively penetrate deeper below the seismogenic layer as they lengthen (Das, 1982; Shaw and Wesnousky, 2008), so that the actual width increases with length. There is at present no conclusive observational data to confirm that suggestion (Hanks and Bakun, 2014), other than from pseudotachylites within the ductile zone (Section 3.2.2), which so far have been observed to penetrate only a few kilometers below the maximum seismogenic depth.

There are thus three scaling regimes for crustal earthquakes, as shown in Table 4.2. Regime 2 encompasses most large crustal earthquakes; events in regime 3 are quite rare, which is why this regime was not noticed for some time (Scholz, 1994c). The first two scaling regimes also are

Table 4.2. Continental earthquake scaling relations

Size regime	Slip scaling	Moment scaling
1. $L<W^*$	$\Delta u \propto L$	$M_0 \propto L^3$
2. $W^*<L<10\ W^*$	$\Delta u \propto L$	$M_0 \propto L^2\ W^*$
3. $L \geq 10\ W^*$	$\Delta u \propto W^*$	$M_0 \propto LW^{*2}$

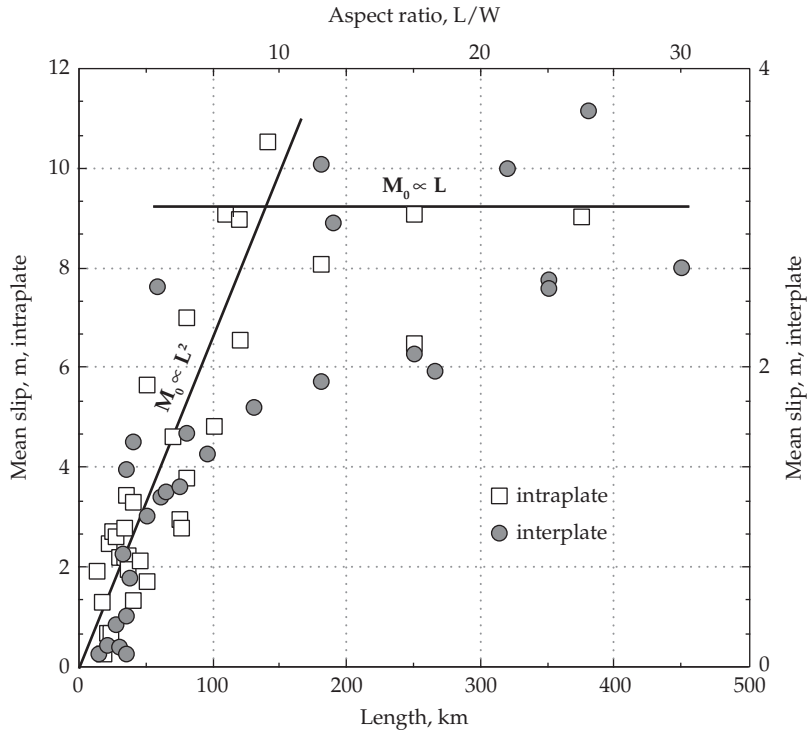

Fig. 4.13. Mean slip versus rupture length for large crustal earthquakes. Two approximate scaling regimes are observed. Slip increases approximately linearly with length for aspect ratios up to about 10, and is independent of length thereafter. There is a broad crossover between these two regimes, which are listed as regimes 2 and 3 in Table 4.1. Notice the difference in scale for intraplate versus interplate earthquakes. (After Scholz, 1994c.)

observed for subduction earthquakes, although the slip scaling relation for regime 2 may differ, as discussed earlier. Regime 3 may also occur for very large aspect ratio subduction earthquakes, as well, but there are too few examples to confirm that.

An important feature of Figure 4.13 is the difference in the scaling parameters between interplate and intraplate earthquakes, the latter having about three times the slip at a given length, as indicated by the different scales on either side of the diagram. In regime 2, the proportionality factor α between Δu and L is 6.5×10^{-5} and 1.5×10^{-5} for intraplate and interplate events, respectively. In regime 3, one can evaluate the stress drops for these two catagories using a two-dimensional crack model: it is ~4 MPa for interplate and ~12 MPa for

intraplate events. This difference in stress drop between intra– and interplate earthquake is thought to be a geologic slip–rate effect (Scholz et al., 1986, Kanamori and Allen, 1986), resulting from the difference in healing during interseismic periods (e.g. Marone, 1998b; Marone et al., 1995). It is seen also in the global data set for shallow earthquakes (Allmann and Shearer, 2009), in which the average stress drops for interplate earthquakes is 3.31 MPa versus 5.95 MPa for intraplate earthquakes.

The scaling in regimes 1 and 2 has the same form as that observed for faults (Figure 3.13). The constant of proportionality for faults is several orders of magnitude larger than that for earthquakes, however, because the associated stress drop is the difference between the intrinsic fracture strength and the frictional strength, whereas for earthquakes it represents the difference between the static and dynamic friction, a far smaller quantity. The displacements on faults develop over many cycles of earthquakes, the respective scaling laws of which are compatible (Cowie and Scholz, 1992b). Regime 3 has not been observed for faults: the maximum fault lengths in Figure 3.13 are about 100 km, and the down–dip width of faults, including their ductile shear zones, is several times the seismogenic width, so the largest aspect ratio for the faults in Figure 3.13 is only about 3, well below the crossover point (n.b., plate boundary faults, such as the San Andreas Fault, do not obey such scaling laws as they have no defined ends).

Scaling of fracture energy

Abercrombie and Rice (2005) estimated E_G (G) from seismological parameters under the assumption of no overshoot and found that $G = 5.25 \times 10^6 \, \Delta u^{1.28}$ Jm^{-2}. Nielsen et al. (2016) estimated G by integrating under the frictional weakening curves from high–speed friction such as in Figure 4.14(a). They compared these values, in red in Figure 4.14(b), with the seismologically determined values of G, from Abercrombie and Rice and others, shown there in other colors. The general finding is that G scales with slip with an exponent that is probably not significantly different from 1. This agrees with the result, based on fault scaling laws, that G scales linearly with L (Section 3.2.2). The experimental values are smaller than the seismological ones, with the discrepancy increasing with slip. Nielsen et al. attributed this difference to the additional dissipation in off–fault damage (Section 3.3.1), which is not included in the experimental values.

Viesca and Garagash (2015) provided a different explanation for the falloff of G in the experimental data for displacements greater than ~1 m. They modeled this as due to thermal pressurization under undrained conditions, resulting in a strength reduction with slip similar to shown in Figure 4.14(a). The earthquake data, on the other hand, show a much gentler falloff in the growth of G at large displacements. This was modeled by Viesca and Garagash as thermal weakening under drained conditions. In that case, the strength reduction would be much more gradual with earthquake slip, remaining substantially greater than the experimental values out to larger displacements. A more general explanation is thermal convection by transient granular vortices in the fault core that prevents strong thermal weakening during earthquakes (Einav, 2018).

4.3.2 Earthquake size distributions

The other fundamental earthquake scaling relationship is expressed in their size–frequency distribution. In any region it is found that during a given period the number $N(M_0)$ of earthquakes occurring of moment $\geq M_0$ obeys a relation

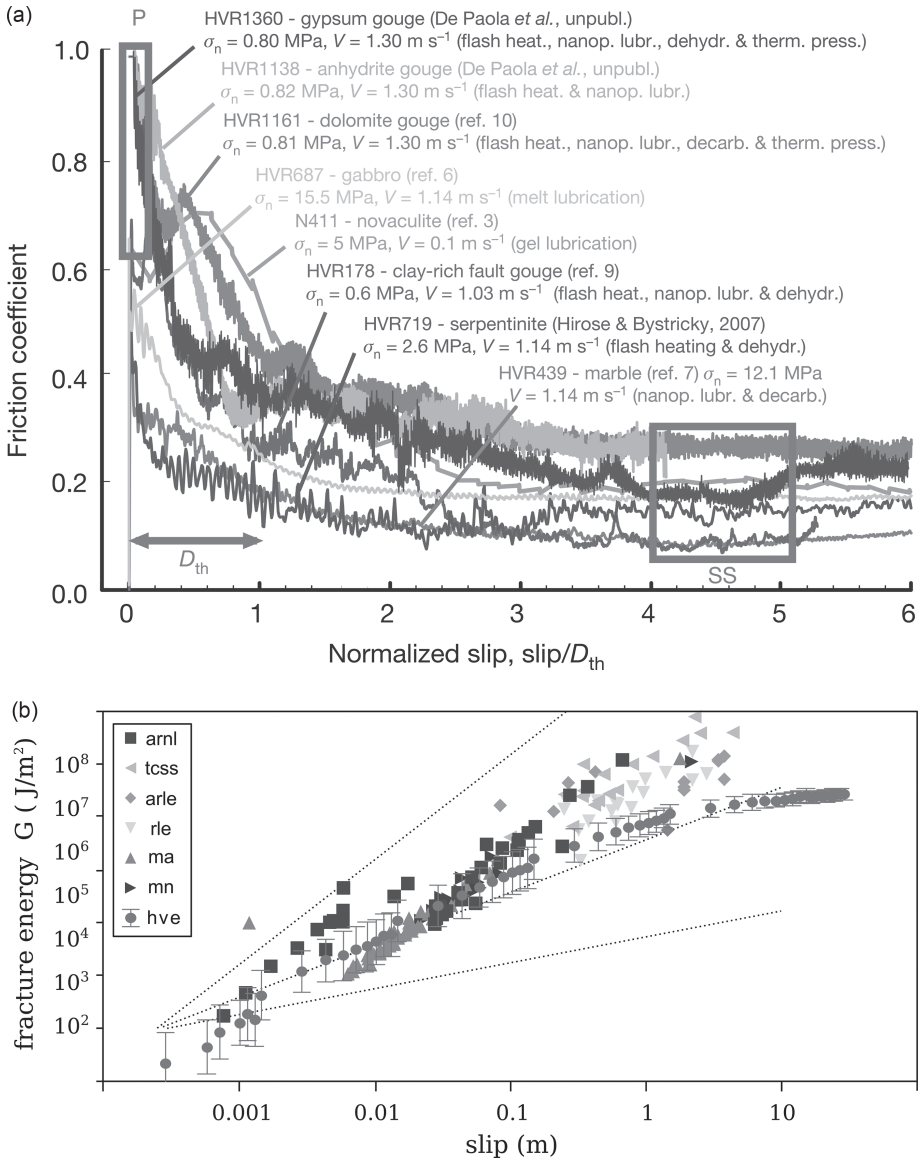

Fig. 4.14 (a) Friction coefficient as a function of normalized slip for high-speed friction experiments with a variety of rock types. The normalized slip displacement D_{th} is the in range 10-20 cm at seismogenic depths. (From DiToro et al., 2011.) (b) Fracture energy from high-speed friction experiments (red) and from seismological measurements (other colors). The dashed lines, for reference, have slopes of 0.5, 1, and 2. (From Nielsen et al., 2016.) (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

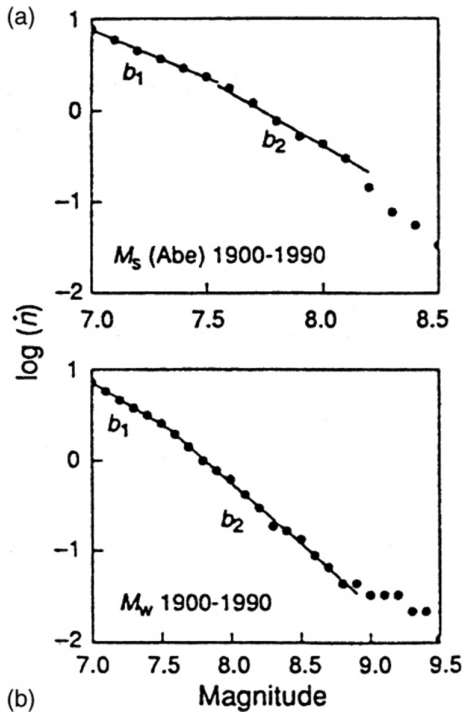

Fig. 4.15. The frequency–size distribution for shallow earthquakes from the global catalog of Pacheco and Sykes (1992): (a) frequency versus M_S; (b) frequency versus M_W. Frequency given as the annual number of earthquakes of a given magnitude or larger. A change in slope from ~1.0 to 1.5 is observed at $M = 7.5$, which corresponds to the change from small to large earthquakes for subduction zone events, which dominate this catalog. For strike-slip earthquakes, the transition occurs at $M = 6$. (From Pacheco et al., 1992.)

$$N(M_o) = aM_0^{-B} \tag{4.19}$$

where a is a variable in time and space. When written in terms of magnitude this is historically known as the Gutenberg–Richter or Ishimoto–Aida relation. The exponent in the familiar Gutenberg–Richter formulation, b, equals $3/2B$, which can be obtained from Equation (4.2). This type of power–law size distribution is typical of fractal sets. Many processes can give rise to power–law size distributions (Newman, 2006; Sornette, 2003). Without considering process, it can be seen as simply a consequence the self–similarity of earthquakes: the exponent $B = 2/3$ for the worldwide distribution and has been shown to be implicit from the self–similar scaling of small earthquakes (Hanks, 1979; Andrews, 1980; Aki, 1981). However, as shown in Figure 4.15, at the crossover between small and large earthquakes, as defined earlier, B changes to 1 for the larger events (Pacheco, Scholz, and Sykes, 1992). This reflects the break in self–similarity at that point, from 2D to 1D propagation (Rundle, 1989). At the lower end, the size–frequency distribution has been found to extend down to M –4 (Kwiatek et al., 2010).

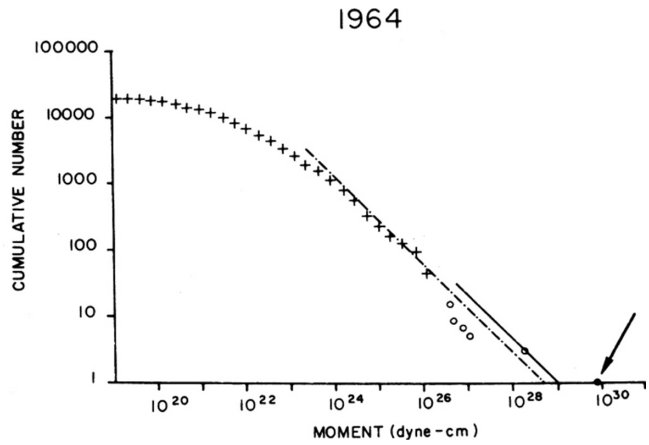

Fig. 4.16. Distribution of small earthquakes within the rupture zone of the 1964 Alaska earthquake, normalized to the recurrence time of that earthquake. The 1964 earthquake is indicated by an arrow. Note that it is about 1.5 orders of magnitude larger than the extrapolation of the small earthquakes would indicate. The roll-off at $M_0 < 3 \times 10^{23}$ dyne cm is caused by the loss of perceptibility of smaller events. (From Davison and Scholz, 1985.)

If the size distribution of earthquakes is determined for a single fault or fault segment, however, it is found that the size of the large earthquake that ruptures the entire fault will be greatly underestimated by the extrapolation of the size distribution of small earthquakes for the same fault (Wesnousky et al., 1983; Singh, Rodriguez, and Esteva, 1983; Schwartz and Coppersmith, 1984; Davison and Scholz, 1985; Stirling, Wesnousky, and Shimazaki, 1996). An example of this is shown in Figure 4.16. This fractal "tear" occurs because of the different scaling relationships for large and small earthquakes, resulting in their belonging to different fractal sets (Scholz, 1994a). This type of distribution, often called the *characteristic earthquake distribution*, has been reproduced with a cellular model of a heterogeneous fault, styled after the Parkfield, California, region (Ben–Zion and Rice, 1993). This distinction is very important in seismic hazard evaluation (Section 7.2). Although this distribution was first discussed in the context of the *characteristic earthquake model* (Section 5.4.1), they are conceptually different phenomena.

In laboratory experiments the *b*-value in the size–frequency distribution is found to decrease with increasing stress (Amitrano, 2003; Goebel et al., 2013; Scholz, 1968c). Observational studies have shown that the same stress dependence occurs for earthquakes (Gulia and Wiemer, 2010; Mori and Abercrombie, 1997; Schorlemmer et al., 2005; Spada et al., 2013). From the depth dependence of the *b*-value in different tectonic regimes shown in Figure 4.17, the relationship is found to be $b = 1.23 \pm 0.06 - (0.0012 \pm 0.0003)(\sigma_1 - \sigma_3)$, where the stress difference $(\sigma_1 - \sigma_3)$ is in megapascals (Scholz, 2015). Lab experiments also show that the b-value is a function of surface roughness (Goebel et al. 2017).

The observations of earthquake populations are summarized in Figure 4.18, in which (a) shows the distribution of earthquakes on a single fault for a single cycle of a single large earthquake that ruptures the entire fault length L_f, and (b) shows the distribution for a region containing a large number of faults. These distributions derive from the size distribution of

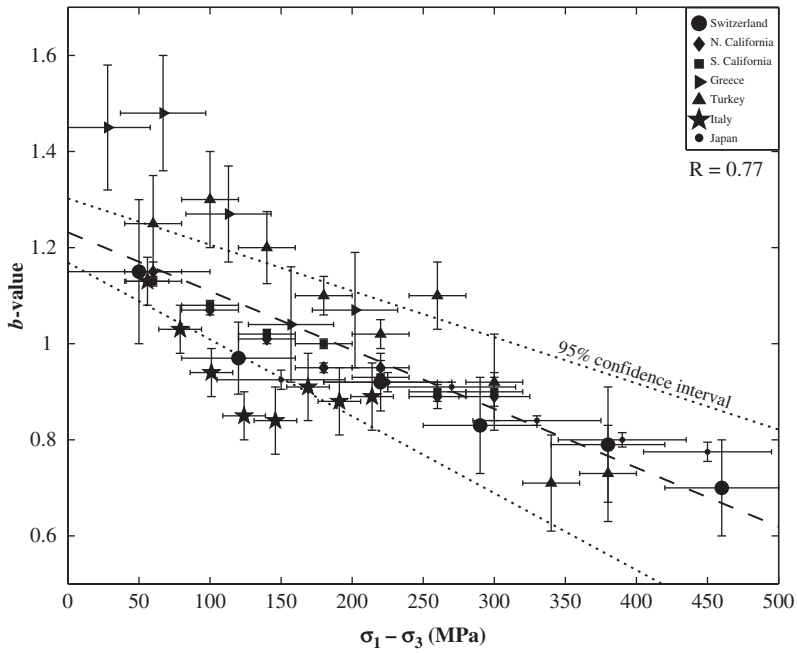

Fig. 4.17. *b*-value for size distributions of earthquakes from a variety of tectonic regimes. Stress-depth gradients are obtained from standard depth distribution of stress from borehole measurements (Zoback and Townend, 2001). (From Scholz, 2015.)

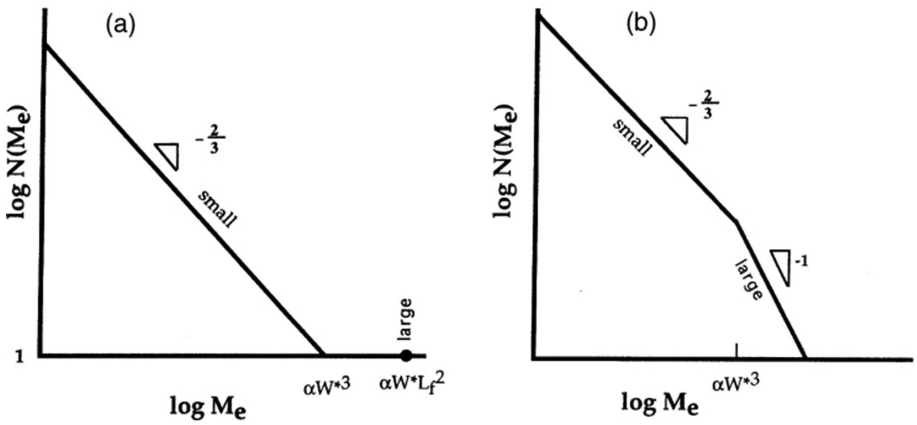

Fig. 4.18. Diagram showing the cumulative size distribution of earthquakes, for (a) a single fault or plate boundary segment and (b) a large region containing many faults or plate boundary segments. (From Scholz, 1997b.)

faults, discussed in Section 3.2.4. If we convolve the earthquake distribution for a single fault (Figure 4.18(a)) over the size distribution of faults, Equation (3.9), with $C = 1$ (Figure 3.25(a)), we obtain the earthquake size distribution for the region, shown in Figure 4.18(b) (Scholz, 1997b). The size distribution for small earthquakes, with $B = 2/3$, on the other hand, can be derived from the size distribution of subfaults (Figure 3.25(b)) for which $C = 2$ (Scholz, 1998b). This distribution of subfaults also explains the ω^{-2} high–frequency falloff on earthquake displacement spectra (Frankel, 1991).

4.3.3 Quantifying heterogeneity: the fractal nature of earthquakes

In Sections 3.31 and it was pointed out that faults are fractal features. They have surface topographies that are self–affine fractals and the size distributions of faults and subfaults both obey the power laws characteristic of self–similar fractals. As discussed in the previous section, earthquakes also have fractal size distributions that can be attributable to the underlying fault distributions.

It should therefore not be surprising that the interior structure of earthquakes, in terms of the distribution of slip and stress drop, should also be fractal–like. Direct evidence for this is in the near–field strong ground motions, which exhibit a flat spectrum typical of white Gaussian noise (Baltay et al., 2013; Hanks and McGuire, 1981; Hanks, 1979), as well as the related ω^{-2} high–frequency falloff on earthquake displacement spectra (Frankel, 1991). Both of these are consistent with a fractal distribution of slip within earthquakes (Candela et al., 2011b; Causse et al., 2010). Inversion schemes for slip distributions within earthquakes produce long–wavelength approximations and do not reveal these complexities.

There are various models of earthquakes that incorporate this fractality in several ways. Gusev (2013) pointed out that the directivity effect could be seen only in the low–frequency end of the radiation and not in the high frequencies, which resemble random noise suggesting incoherent radiators. He proposed that the rupture front resembles a propagating line at the macroscale and as a multiply connected disjoint fractal line at the microscale. This kind of rupture front development has been observed in the laboratory (Delaplace et al., 1999; Schmittbuhl and Maloy, 1997). Ide and Aochi (2005), on the other hand, developed a model consisting of fault patches in which the frictional properties and fracture energies follow power–law scaling. Similar heterogeneous fault rupturing models have been used which simulate the broadband strong ground motion radiation discussed earlier (Dunham et al., 2011; Ruiz et al., 2013). Candela et al. (2011a) related the roughness of the fault to the distribution of stress drops. This provides a connection between the measured topography of faults and the dynamic properties of the earthquakes that run on them.

4.4 OBSERVATIONS OF EARTHQUAKES

4.4.1 Case studies of continental earthquakes

In this section a number of earthquakes are described in some detail, with particular attention to the rupture process. These cases are not necessarily meant to be taken as typical, since there is considerable variability in detail among earthquakes. The cases chosen represent the main faulting types, and all occurred in a continental setting. For a description of subduction zone

and oceanic earthquakes, go to Sections 6.3 and 6.4. The selections here were made primarily of earthquakes that have been unusually well studied.

Strike-slip: Denali, Alaska, November 3, 2002, M_W *7.9*

This earthquake ruptured 340 km of the Denali fault system, a right–lateral strike–slip fault that marks the northern boundary of the Yakutat block and the southern Alaska margin of the North American Plate (Figure 4.19) (Eberhart–Phillips et al., 2003; Haeussler et al., 2004). Quaternary offsets and GPS data indicate that the geologic slip rate of the main branch of the Denali fault is 8–13 mm/yr.

The sequence began with the M_W 6.7 Nenana Mountain earthquake of October 23, 2002. This strike–slip earthquake ruptured 45 km of the Denali fault and terminated 10 km west of the main Denali earthquake (Figure 4.19). The initial rupture of the main earthquake was an oblique thrusting subevent on the previously unrecognized Sustina Glacier fault, followed by 218 kms of primarily right–lateral rupture on the Denali fault. The rupture then propagated through a 14–km–long transfer zone to the Totshunda fault, where it continued for an additional 66 km. Aftershocks were limited to depths less than 10 km, suggesting that the main slip was restricted to depths shallower than this. The horizontal slip distribution obtained by three methods is shown in Figure 4.20. The surface slip distribution correlates well with the other estimates,

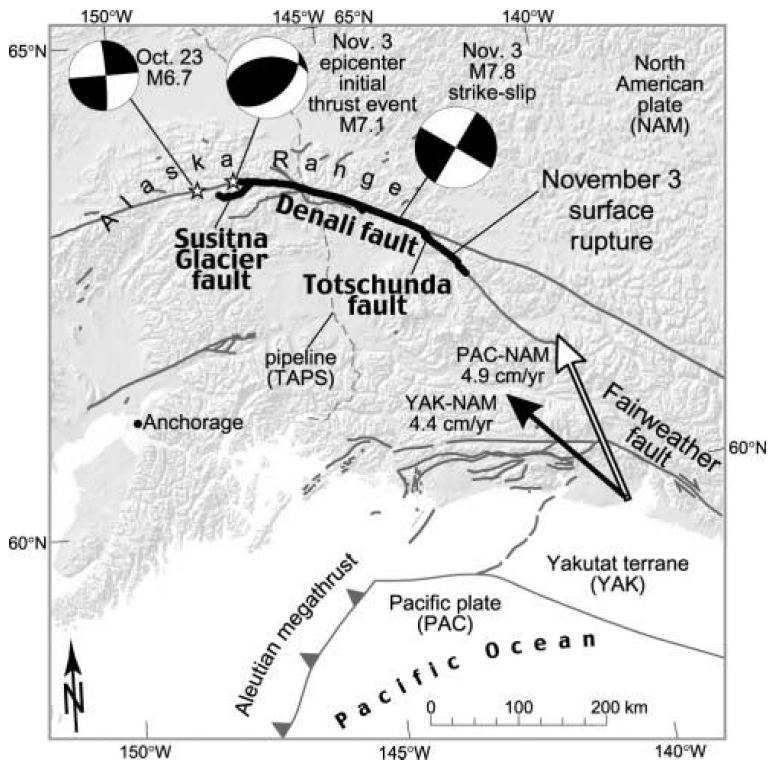

Fig. 4.19. Location map of the 2002 Denali earthquake. Focal mechanism of the foreshock and two subevents of the mainshock are shown. (From Haeussler et al., 2004.)

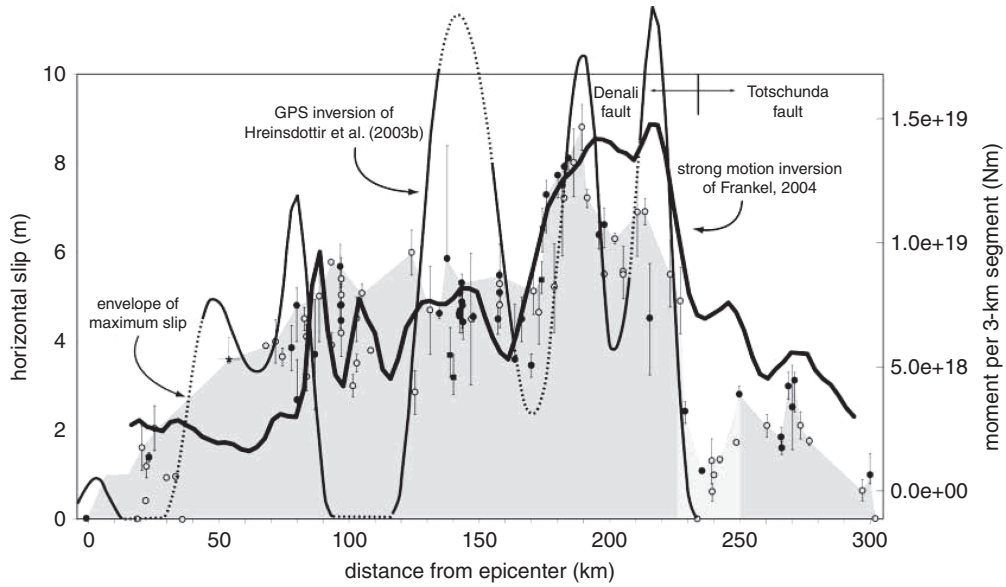

Fig. 4.20. Horizontal slip distribution for the Denali earthquake. Surfaces slip measurements (data points) and the envelope of maximum slip (shaded) from Haeussler et al. (2004). Slip inversions from GPS data from Hreinsdottir et al. (2003) and from strong motion records from Frankel (2004). (Figure from Haeussler et al., 2004.)

which can be expected for such a shallow rupture. Vertical displacements were about 20% of the horizontal displacements and were primarily north–side up. This reflects the transpressional nature of the Denali fault with respect to the motion of the Yakutat block.

The slip maximum at about 200 km east of the epicenter corresponded to a strong pulse of radiation and a locally super–shear rupture velocity of about 5 km/sec (Frankel, 2004) compared to an average rupture velocity from a dynamical model of about 3 km/sec (Dreger et al., 2004). The transfer of slip to the Totshunda fault results from its more favorable orientation within the stress field (Bhat et al., 2004). The transfer zone connecting the faults consists of a series of right–stepping right–lateral segments separated by extensional jogs with short normal fault segments (Haeussler et al., 2004). There was a local slip minimum within this zone (light shading, Figure 4.20). The dynamic model of Dreger et al. (2004) suggested that the rupture front may have jumped over the transfer zone onto the Totshunda fault proper.

The Totshunda fault is a first–order splay fault of the Denali (see Section 3.1.3). It has a strike subparallel to the YAK–NAM slip vector and the Queen Charlotte–Fairweather fault system, a portion of the PAC–NAM plate boundary further south (Figure 4.19). The quaternary slip rate of the Denali fault east of its intersection with the Totshunda fault is 2–3 mm/yr, much less than the 8–13 mm/yr rate to the west, indicating that transfer of slip to the Totshunda fault is a long–term phenomenon.

This earthquake was notable for triggering earthquakes at great distances to the south and southeast in North America, a topic that will be discussed in Section 4.5.3. The concentration of these phenomena along strike in the rupture propagation direction demonstrates the *directivity effect*, namely the amplification of seismic waves due to the relativistic contraction of the stress field in the direction of rupture propagation.

Thrust Faulting: Chi-Chi, Taiwan, September 20, 1999, M_W *7.6*

This was the largest earthquake to strike Taiwan in the twentieth century. It occurred on the Chelungpu fault, a ramp fault in the fold and thrust belt of Taiwan (Figure 4.21(a)/(b)). The Chelungpu fault rises at a 30° dip from a basal decollement that is aseismically slipping at about 35 mm/yr (Hsu et al., 2003). It extends through a thick sequence of sedimentary rocks (Figure 4.21(c)). The earthquake initiated near the base of the Chelungpu fault and exhibited a complex rupture pattern, with several fault strands participating (Kao and Chen, 2000). It initially propagated up–dip and bilaterally, but then continued mainly to the north, where the largest moment release occurred. The rupture process was complex, involving many fault segments and abrupt changes in rupture velocity (Ma et al., 2000), with an average rupture

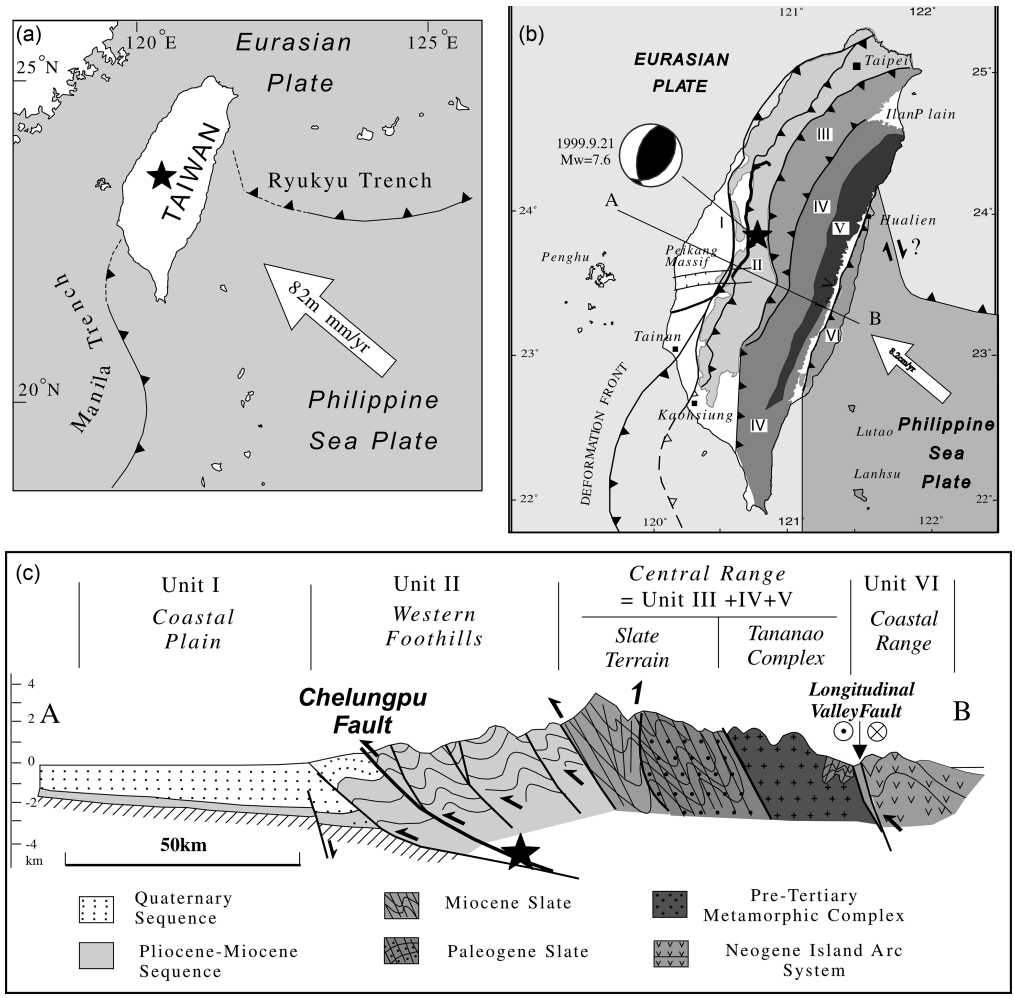

Fig. 4.21. (a) Tectonic framework of Taiwan and location of the Chi-Chi earthquake. (b) Generalized geologic map of Taiwan, with location of the Chi-Chi earthquake. (c) Geologic cross section along line A–B. (From Lee et al., 2002.)

velocity being about 2 km/sec (Ji et al., 2003). The earthquake rupture length was about 80 km and produced surface ruptures over that entire length.

The slip was primarily up–dip thrust with a left lateral component. Inversion of GPS and strong motion data indicate that slip was about 5 m in the hypocentral region, and increased greatly to the north, where at shallow depths the slip reached a maximum of about 20 m (Johnson et al., 2001; Ma et al., 2001; Wu et al., 2001; Zeng and Chen, 2001). Afterslip occurred immediately after the earthquake, primarily in a region just below the hypocenter (Hsu et al., 2002).

The initial part of the rupture radiated a normal amount of high–frequency energy, but when the rupture entered the region of high slip at the northern end, it was greatly depleted in high frequencies. This led Ma et al. (2003) to proposed that this latter behavior resulted from hydrodynamic lubrication (Brodsky and Kanamori, 2001) in which the fault surfaces became separated by a lubricating fault gouge layer such that collisions between asperities did not occur during this part of the rupture.

Large coseismic thrust displacements were found at the surface exposure of the Chelungpu fault. Such sharply defined surface ruptures with large offsets are unusual for thrust earthquakes, which more frequently result in a distributed monoclinal crush zone near the surface with only minor localized surface breaks. In the Chi–Chi earthquake there were many places where the hanging wall block was displaced many meters over the footwall rocks with very little internal deformation (Bilham and Yu, 2000). This suggests that separation may have occurred between the blocks, as observed in foam rubber models (Brune, 1996; Brune et al., 1993) and much more convincingly in a laboratory fault in Homelite (Gabuchian et al., 2017). In such models of shallow angle thrust faults the hanging wall block was observed to separate, rotate, and flip over the footwall block. This is a different model for separation than that discussed above, and may also help explain the lack of high–frequency radiation in the northern region of large slip. Such effects, including large slips and accelerations observed in the near–surface, arise from the asymmetry of faulting – namely, the focusing of deformation in the acute wedge of the hanging wall (Oglesby and Day, 2001).

Normal faulting: L'Aquila, Italy, April 6, 2009, M_W *6.3; Amatice, Italy, August 24, 2016;* M_W *6.0, and Norcia, Italy, October 10, 2016,* M_W *6.5*

These earthquakes occurred along a NW belt of active normal faults that run along the western edge of the Apennines (D'Agostino et al., 2008; Roberts et al., 2004; Roberts and Michetti, 2004). The faults in this area had previously been ruptured in a series of three earthquakes in 1703 in a SE progression over a period of 19 days (Guidoboni and Valensise, 2015). The first, on January 14, was an M~6.7 near Norcia, followed by an M~6.2 on January 16 near Montereale and then an M~6.7 on February 2 at L'Aquila. These ruptured all the active faults between Norcia and L'Aquila (Galli et al., 2005). The present sequence could be said to have begun with the M_W 6.0 Umbria–Marche earthquake sequence of 1997 (Amato et al., 1998), which ruptured the region just to the north of Norcia. These latter earthquakes are shown in Figure 4.22(a).

The city of L'Aquila lies within an intermontane basin created by the normal fault that hosted the 2009 earthquake. Figure 4.22(b) shows the coseismic displacement field represented by fringes of an InSAR radar interferogram. The earthquake activated a group of normal faults dipping ~50° to the SW (Figure 4.23). A swarm of small foreshocks began four months prior to the mainshock, as shown in Figure 4.24 (Chiaraluce, 2012). These occurred in a 4–km–long band on the L'Aquila rupture plane at the depth (9 km) of the eventual mainshock hypocenter. Following an M_W 4.0 foreshock on March 30, activity switched to an antithetic fault plane,

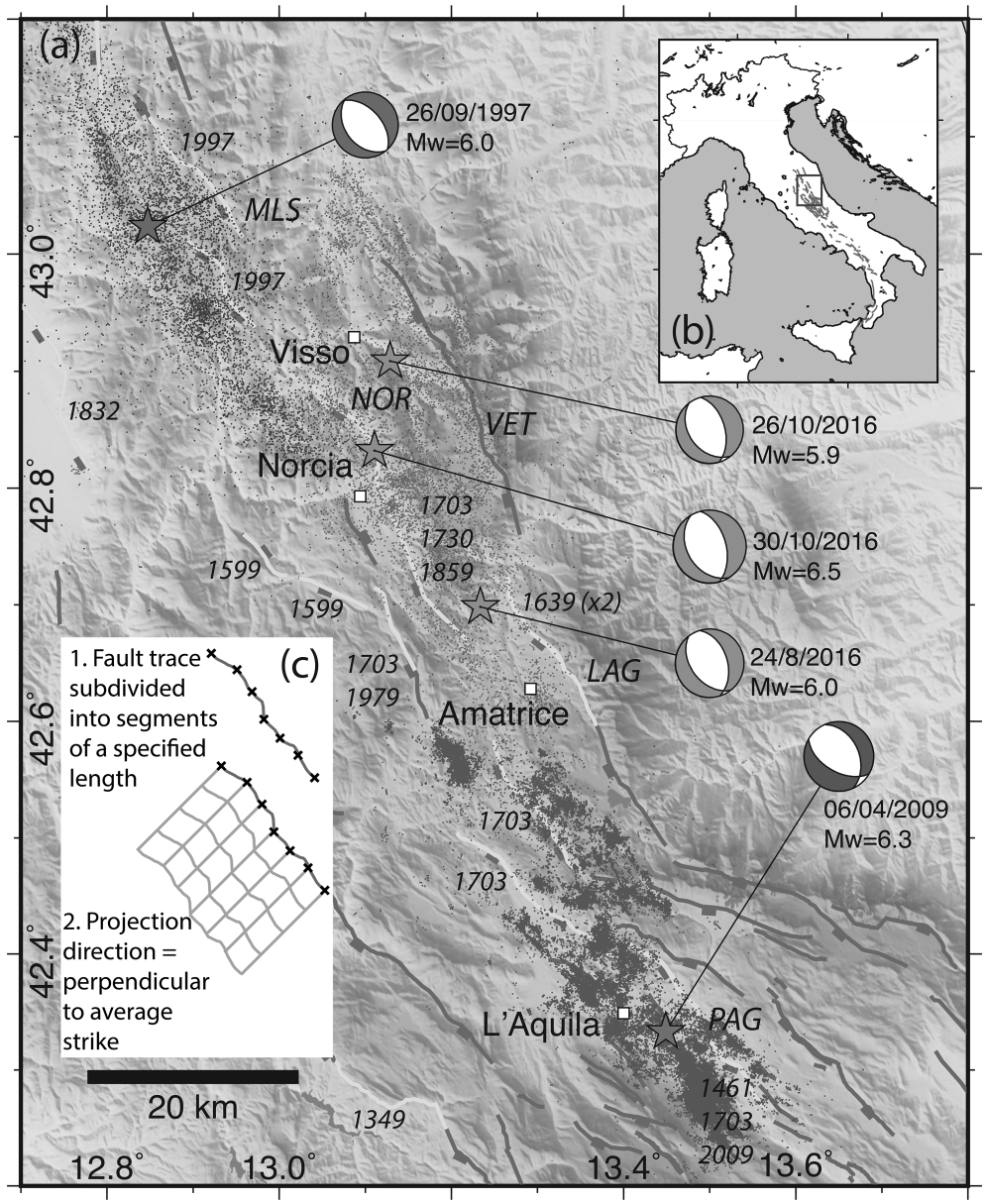

Fig. 4.22. (a) Map of the Northern Apennines seismic belt. Mainshocks, and their aftershocks, of the recent earthquake are color coded: L'Aquila, magenta; Amatrice and Norcia, green; and Umbria–Marche, blue; the active faults are in red. Historic earthquake ruptures are in yellow indicated with dates in black. (From Mildon et al., 2017.) (b) InSAR image of the ground deformation due to the April 2009 L'Aquila earthquake in Central Italy draped over a digital elevation model. The rainbow colors show 2.8 cm contours of ground displacement around the city of L'Aquila, measured using the radar instrument on the Envisat (ESA) satellite (up to ~25 cm motion away from the satellite and ~8 cm toward the satellite). The black line is the causative fault. (From Walters et al., 2009.) (c) Down-dip section of the L'Aquila fault plane showing high precision aftershock locations compared with (i) coseismic slip contours (from Cirella et al., 2009) and (ii) afterslip contours (from D'Agostino et al., 2012). Mapped faults, Mt. Stabiata (MSF) and Paganica (PaF), thick black lines; surface ruptures, thick gray lines (from Boncio et al., 2010). Gray arrows: direction of rupture development. (From Valoroso et al., 2013.) (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

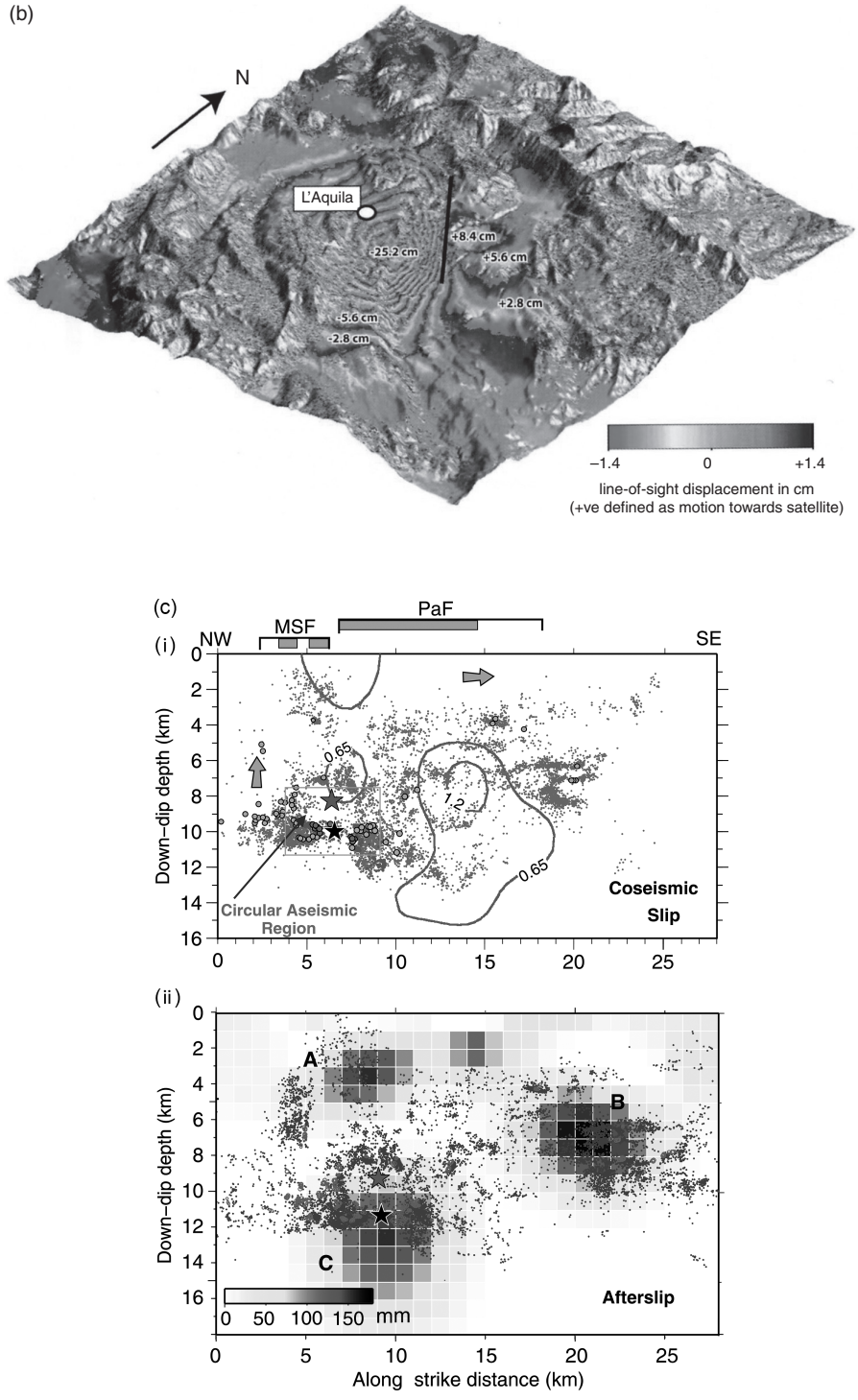

Fig. 4.22. (cont.)

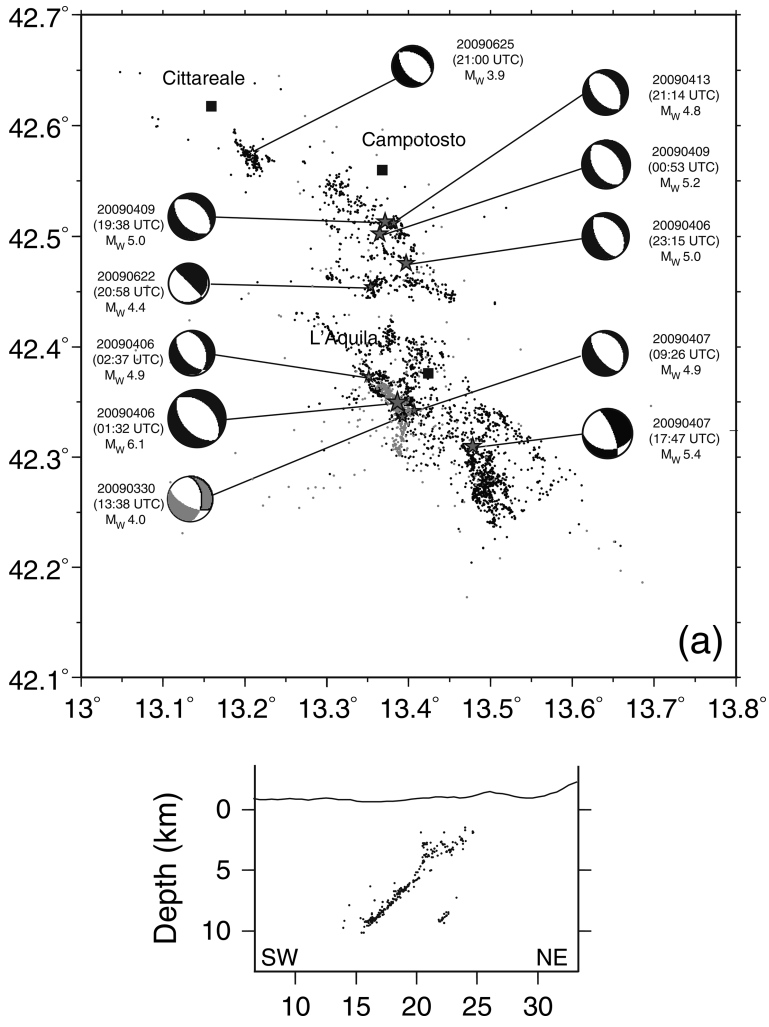

Fig. 4.23. Map and NW-trending vertical section of the L'Aquila seismic sequence. (From Valoroso et al., 2013.)

then switched back to the main rupture plane after an M_W 3.9 foreshock on April 5. The mainshock occurred on the following day. Following the mainshock, activity migrated NW toward the Campotosto fault at a rate of ~3.5 km/day with a series of $M_W > 5$ events.

Details of the rupture are shown in Figure 4.22 (c) (Cheloni et al., 2010; Cirella et al., 2009; D'Agostino et al., 2012; Valoroso et al., 2013). Slip occurred in a patch near the hypocenter and then propagated to the SE, where the area of the largest slip occurred. Small amounts of surface slip were observed on only a few short fault sections, as might be expected from such a small-magnitude earthquake (Alessio et al., 2010; Wilkinson et al., 2015). The aftershocks and after-slip were both concentrated around the edges of the coseismic slip areas. This is fairly typical, as these are both processes by which the stress concentrations induced by the coseismic slip are relaxed.

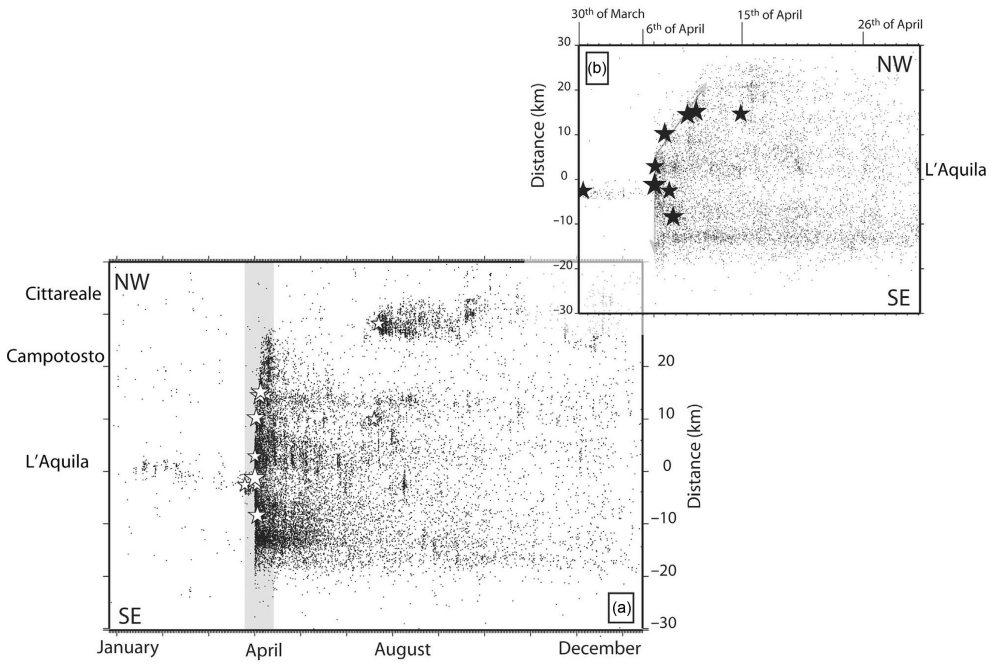

Fig. 4.24. Space–time diagram of the L'Aquila earthquake sequence. B is for the time window shown in gray in A, beginning with the M 4 foreshock of March 30. The stars are events of $M \geq 5$, including the March 30 foreshock and the M 3.9 event in the Cittareale area (June 25) and the M 4.4 (June 22). The gray arrows indicate the direction of seismicity migration. (From Chiaraluce, 2012.)

There was ample evidence for fluid involvement in the L'Aquila earthquake, in particular, for dilatancy and diffusion during the foreshock sequence (Lucente et al., 2010). Discussion of these and other precursory phenomena for this earthquake can be found in Section 7.3.2.

The L'Aquila earthquake was followed by an M_W 6.0 earthquake near Amatrice on August 24, 2016, and one of M_W 6.5 near Norcia on October 30, 2016, as the seismicity continued to migrate northwards (Figure 4.22(a)), filling the gap between the L'Aquila and Umbria–Marche ruptures. The Amatrice earthquake ruptured the southern part of the Mt. Vettore fault and a short strand of the Laga fault to the south (Livio et al., 2016). It was followed by a M_W 5.3 normal faulting earthquake near Norcia. It had a heterogeneous rupture pattern, with a maximum slip of about 1 m in a 5 km diameter patch centered at a depth of 5 km just above the hypocenter, and a second slip patch about 10 km to the north (Tinti et al., 2016). The October sequence involved these same faults further to the north, and were followed by 5 M> 5 earthquakes on January 18, 2017. There was a remarkable observation of coseismic surface rupture of the Mt. Vettore fault by the October 30 earthquake (Wilkinson et al., 2017). By calculating Coulomb stress transfers from earthquakes in this region from 1349 AD combined with stress accumulation, Mildon et al. (2017) showed that the Mt. Vettore fault had been strongly loaded and that, together with a bend in that fault, controlled the location of the 2016 ruptures.

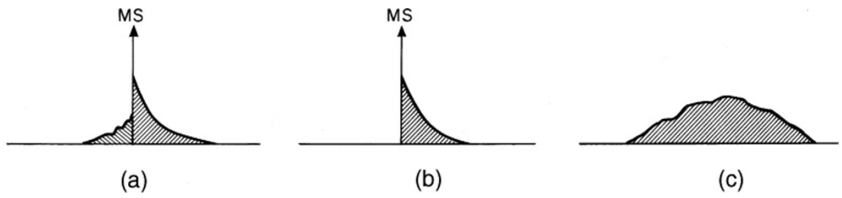

Fig. 4.25. Schematic diagram illustrating the various types of earthquake sequences: (a) mainshock (MS) with foreshocks and aftershocks; (b) mainshock–aftershock sequence; (c) swarm.

4.4.2 Earthquake sequences

Earthquakes seldom occur as isolated events, but are usually part of a sequence with variably well-defined characteristics (Figure 4.25). Foreshock and aftershock sequences are closely associated with a larger event called the mainshock, whereas sequences of earthquakes not associated with a dominant earthquake are called swarms. Occasionally, two or more mainshocks may be closely associated in time and space. These have been called doublets and multiplets. We consider these under the topic of triggered earthquakes, discussed in the next section.

Aftershocks

Of these sequences, aftershocks are the most ubiquitous, being observed to follow almost all shallow tectonic earthquakes of any significant size. They also have the most well-defined characteristics of any of the earthquake sequences. In particular, the decay of aftershock sequences follows the Omori law (named for his observation of it following the 1891 Nobi earthquake):

$$n(t) = \frac{K}{(c+t)^p} \tag{4.20}$$

where $n(t)$ is the number of aftershocks in an interval at time t after the mainshock, K and p are constants, and c is a positive number near zero. The exponent p is usually found to be very close to 1, so this decay law is nearly hyperbolic. It is often stated that the largest aftershock in the sequence is typically at least 1.2 magnitude units smaller than the mainshock, a number that stems from a footnote in Richter (1958, p. 69) in which he called it "Båth's law." There is no documentation in the original source – evidently Markus Båth himself did not publish on this topic. However, in the summary of a series of papers on aftershock sequences of Japanese earthquakes, Utsu (1971) found that the largest aftershock was typically about one magnitude unit less than the mainshock. The sum of seismic moments for the entire sequence usually amounts to only about 5% or so of the moment of the mainshock (Scholz, 1972b). Aftershocks, then, in the classical sense, are therefore a secondary process.

The number of aftershocks increases with the size of the mainshock according to

$$N_{aft} = k10^{b(M_m - M_{th})} \tag{4.21}$$

where M_m is the magnitude of the mainshock, M_{th} is the lower threshold of measuring aftershocks, and k is a usually considered a constant (Utsu, 1970). If we assume the Gutenberg–

Richter parameter b = 1, then, using Equation (4.2), $N_{aft} \propto M_o^{2/3}$. Assuming a circular rupture model, Equation (4.16),

$$N_{aft} \propto \Delta\sigma^{2/3} A \tag{4.22}$$

This scaling was found by Wetzler et al. (2016), who observed that N_{aft} both increased linearly with A and increased with $\Delta\sigma$. Yamanaka and Shimazaki (1990) also observed the linear scaling of N_{aft} with rupture area, and found that it was larger for intraplate than interplate earthquakes. This latter agrees with the greater stress drops for intraplate to interplate earthquakes, as noted earlier (Figure 4.13).

Wetzler et al. (2016) also found the seemingly perverse result that N_{aft} decreases with $\Delta\sigma$ for earthquakes of a given moment. Although an increase of $\Delta\sigma$ at fixed moment will result in a decrease in A, these will just compensate for one another in Equation (4.22) and not yield any expected change in N_{aft}. It is likely that this effect is the consequence of b decreasing with stress, as discussed in Section 4.3.2, which will result in a reduction in N_{aft} according to Equation (4.21).

The time delay c is usually treated as a constant, but it has been found to scale with the GR parameter b and hence inversely with stress (Lippiello et al., 2015; Narteau et al., 2009; Shebalin and Narteau, 2017). It is also a function of M_{th} (Davidsen et al., 2015).

It seems clear that aftershocks are a process of relaxing stress concentrations produced by the dynamic rupture of the mainshock. However, in order to account for the time delay of aftershocks and the characteristic decay law, a time–dependent strength must be introduced. Recalling from Section 1.2.4 that rock strength increases with strain rate, we realize that the dynamic loading during the mainshock may load local regions to stresses much higher than their long–term strength. Such regions will then fail by static fatigue at a later time determined by the imposed stress level. If a static fatigue law of the form of Equation (1.56) is assumed for a large number of such local regions over which the imposed stress levels are assumed to be randomly distributed, it can be shown that the collected rupture behavior of these regions will obey the Omori law (Marcellini, 1997; Scholz, 1968b; Yamashita and Knopoff, 1987). The Omori law is also an expected consequence of the rate–state–friction law (Dieterich, 1994) and subcritical crack growth (Das and Scholz, 1981b; Gomberg, 2001). These are various manifestations of the same underlying physics (e.g. Anderson and Grew, 1977; Brantut et al., 2014; Lockner, 1998; Savage et al., 2005).

There have been efforts to find systematic variations in aftershock productivity. The main finding is that for subduction earthquakes aftershocks productivity is systematically greater in the Western circum–Pacific zones than the Eastern Pacific ones (Singh and Suárez, 1988; Wetzler et al., 2016). In situations where there is significant postseismic expansion of the rupture zone, ηA should be substituted for A in Equation (4.22), where η is the ratio of the long–term to the short–term rupture area. Tajima and Kanamori (1985) found that η is systematically greater for the Western than the Eastern circum–Pacific subduction earthquakes, a factor that explains the difference in aftershock productivity. These differences correlate with a difference in seismic coupling, it being larger in the eastern than in the western Pacific (Scholz and Campos, 1995; 2012). The lower coupled subduction zones in the western Pacific tend to have larger regions of stable to conditionally stable friction such that rupture zones can spread by afterslip into creeping regions (see Figure 5.12(a)), generating more aftershocks as they do (Kato, 2007). A notable difference in aftershock productivity is seen in oceanic transform faults, which are much less productive than continental faults (Figure 6.22). This difference may result from much slower rates of subcritical crack growth for olivine relative to the more siliceaous minerals like quartz. Under all tested conditions Meredith and Atkinson

(1985) found the fracture toughness at a given value of crack velocity for gabbro to be 1.5 greater than of granite. Using the relation of Wiederhorn and Bolz (1970), this implies that the rupture time at a given stress level would be about 30 times longer for the gabbro than the granite. Therefore, the frequency of such failures in the gabbro would be about 1/30 of that in granite: this is within the uncertainty in the difference in the aftershock productivity shown in Figure 6.22.

Although the spatial distribution of aftershocks is usually almost stationary with time, there are several circumstances in which aftershock migration occurs. The one just alluded to is when the rupture zone expands by afterslip. In this case the migration will follow a log t dependence (Kato, 2007; Perfettini and Avouac, 2007). An example is the migration of aftershocks for the 2004 Parkfield earthquake (Peng and Zhao, 2009). Migration may also be driven by fluid diffusion, in which case it follows a $\sqrt{t}$ dependence. An example is the migration of the L'Aquila aftershocks, shown in Figure 4.24 (Di Luccio et al., 2010). Ross et al. (2017) describe a case where both types of migration occur in the same aftershock sequence.

Several aftershock sequences have already been described in the previous section. Aftershocks typically begin immediately following the mainshock and cover the entire rupture area and its immediate surroundings, although they are commonly concentrated in locations where one might expect large stress concentrations to have been produced by the mainshock rupture. These are places with high slip gradients, such as around the rupture periphery as was noted above in the case of the L'Aquila earthquake (Figure 4.22(c)); also see Mendoza and Hartzell, 1988) or in the rupture interior near the periphery of high slip zones (Beroza, 1991; Wetzler et al., 2016).

The traditional view of aftershocks as being limited to the mainshock rupture area and immediate surroundings has been blurred by the discovery of off–fault aftershocks triggered by static stress changes (Das and Scholz, 1981a) and by more distant aftershocks triggered by seismic waves (Hill, 1993). Although from a mechanistic point of view it may seem arbitrary to make a distinction between the traditional near–fault and the more distant aftershocks, the latter are discussed separately in the next section under the heading "Triggered Earthquakes." One reason for presenting it this way is that in the case of more distant triggering it becomes possible to determine the causative stresses and hence discuss the mechanisms involved, whereas that is not possible for the near–fault case because of the fractal–like complexity of the coseismic slip and resulting stresses. That section will lead to a discussion of earthquake clustering in general, and statistical clustering models that have been developed to describe it.

Foreshocks

Foreshocks are smaller earthquakes that precede the mainshock. They usually occur in the immediate vicinity of the mainshock hypocenter and are therefore probably a part of the nucleation process (Das and Scholz, 1981b). (Earthquakes that precede a mainshock but are not near the hypocenter are probably not causally related to the mainshock and, strictly, should not be considered foreshocks.) Unlike aftershocks, the occurrence of foreshocks is quite variable, as indicated by the examples in the previous section. The Denali earthquake was preceded by 10 days by a single large foreshock, whereas the L'Aquila earthquake was preceded by a swarm of foreshocks lasting four months. Further discussion of foreshocks will be deferred till Section 7.2.2.

Swarms

Earthquake swarms are sequences of earthquakes that often start and end gradually and in which no single earthquake dominates in size. Sykes (1970a) made a global survey of swarm occurrence and found that they commonly are associated with volcanic regions, though this is by no means a universal rule. A major swarm at Matsushiro, Japan, was evidently caused by an upwelling of pore fluid, seemingly of plutonic origin (Nur, 1974; Kisslinger, 1975). This seems to have been a natural occurrence of the mechanism responsible for causing earthquake swarms by fluid injection in boreholes (Sections 2.5 and 6.6.1). According to this mechanism, the earthquakes are produced by an increase in pore pressure caused by fluid diffusion. As a result, they occur in a region in which there is an unusually strong stress gradient, so any event in the sequence is prevented from growing very large; strain relief is controlled by the fluid flow, and no dominant large event can occur. For the same reason, the *b*-value in the earthquake size distribution is often observed to be unusually large in swarms (Scholz, 1968c; Sykes, 1970a).

Swarms may also be driven by slow slip events (see Section 4.6) as evidenced by geodetic data (Segall et al., 2006) or by earthquake migration rates that correspond to the propagation rates typical of slow slip events (Lohman and McGuire, 2007; Roland and McGuire, 2009). Both mechanisms are evident in a survey of swarms in southern California (Vidale and Shearer, 2006).

Mogi (1963) has divided these sequences into three types: mainshock–aftershock, foreshock–mainshock–aftershock, and swarm, which he has interpreted as indicating increasing heterogeneity of the source region. He found that these sequence types were variously dominant in different parts of Japan and interpreted this in terms of regional variations of heterogeneity. While this idea has merit, this author feels that it is too generalized. The Adirondacks of New York is a cratonic region in which Precambrian basement outcrops at the surface (Grenville age, 900 My) in a very uniform massif. Several intense earthquake swarms typified by small-magnitude events occurred at Blue Mountain Lake in the central Adirondacks in 1972–1973, then subsided. In 1975, an M4.0 earthquake occurred at Raquette Lake, just 10 km away. In spite of the quick installation of portable seismometers, no aftershocks could be found for this event: it was a singleton. In 1985, an M5.1 earthquake occurred at Goodnow, just 20 km away, and this earthquake had a normal aftershock sequence (Seeber and Armbruster, 1986). Thus, over a very short period of time in a small region of the same structural province, the gamut of earthquake sequences was observed. All the earthquakes had similar thrust mechanisms, the only systematic environmental difference was an increase in depth, from 2 to 3 km for the swarm, 4 km for the Raquette Lake earthquake, and 6 km for the Goodnow sequence. Whether or not this depth difference is relevant to the difference in sequence type is moot.

4.5 EARTHQUAKE INTERACTIONS

The type of earthquake clustering indicated by foreshock and aftershock sequences show that earthquakes are not isolated independent events. It appears, rather, that, as do faults on a longer timescale (Section 3.2.3), earthquakes interact through their stress fields. This is not limited to the immediate vicinity of the mainshock rupture, as in the case of what are traditionally described as aftershocks and foreshocks. Earthquakes can be triggered (or inhibited) at greater distances. Earthquake triggering can be classified as near-field triggering, within one or two fault lengths of the causative earthquake, and far-field triggering, at regional to teleseismic

distances. Near–field triggering can be caused by either static changes in the stress field or by dynamic triggering from seismic waves, whereas more distant effects can be caused only by dynamic triggering. This is because static stresses fall off rapidly with distance from the source, whereas dynamic stresses decrease more slowly, particularly in the case of surface waves. In this section a variety of such interactions, generally described as earthquake triggering, are discussed. A supplemental review may be found in Freed (2005).

4.5.1 Static triggering: Coulomb stress loading

Assuming a simple Coulomb friction model for earthquakes, the potential for slip will be enhanced or retarded by a change in the Coulomb failure stress, ΔCFS, as defined by

$$\Delta CFS = \Delta\tau_s - \mu(\Delta\sigma_n - \Delta p) \tag{4.23a}$$

$$\approx \Delta\tau_s - \mu'(\Delta\sigma_n) \tag{4.23b}$$

in which $\Delta\tau_s$ is the change in shear stress resolved in the slip direction on the potential fault, $\Delta\sigma_n$ and Δp are the changes in normal stress and pore pressure on the fault (positive for compression), and μ is the friction coefficient. Thus, if the $\Delta CFS>0$, slip potential is enhanced and if $\Delta CFS<0$, it is inhibited.

In the case of coseismic stress changes, $\Delta\sigma_n$ and Δp are not independent. Any sudden change in normal stress in a saturated porous medium will induce a corresponding change in pore pressure (see Section 6.6.3). The most general equation relating them is $\Delta p = B\Delta\sigma_{kk}/3$, where B is Skempton's coefficient ($0 \le B \le 1$) and $\Delta\sigma_{kk}$ is the sum of the diagonal elements in the stress tensor (Rice and Cleary, 1976). If the shear modulus of the fault can be considered to be considerably smaller than that of the surrounding rock, then $\Delta p = B\Delta\sigma_n$ (Cocco and Rice, 2002). In that case, the simplified form, Equation (4.23b), can be used, where $\mu' = \mu(1 - B)$ is called the "apparent friction." This simplified form is often used in the literature on this topic, but it merely disguises the fact that we do not know how to handle the poroelastic effect in this case. Not only do we not generally know the appropriate value of B to employ, and so cannot relate μ' to μ, but Equation (4.23b) also hides the inherent time dependence of μ'. The Rice and Cleary relation describes the undrained (instantaneous) response. With time, pore fluid diffusion will allow Δp to return to zero, hence μ' will rise to μ. This is a potentially important effect that will be discussed again in Section 4.5.2.

The calculation of ΔCFS produced by an earthquake depends on knowledge of the geometry and slip distribution of the earthquake, the assumed magnitude and orientation of the regional stress, and the assumed value of μ'. The calculation is specific to the orientation and slip vector of the "target" fault on which the supposed triggered earthquake occurs. The ratio of the regional stress amplitude to the earthquake stress drop is important only close to the fault, where uncertainties in ΔCFS are in any case always dominated by uncertainties in the slip distribution (King and Cocco, 2000). The effect of both becomes less important further from the fault. Regional stress orientation is more important, but can usually be constrained within fairly narrow limits by other information. The effect of the assumed value of μ' is modest in most cases (King and Cocco, 2000).

Figure 4.26 shows the 1979 M 5.2 Homestead Valley, California, earthquake, one of the first cases studied using this method (Das and Scholz, 1981a; Stein and Lisowski, 1983). The mainshock rupture (white line) was constrained with seismic and geodetic data. The four lobes of positive and negative ΔCFS are apparent. Although most aftershocks occurred on or very near

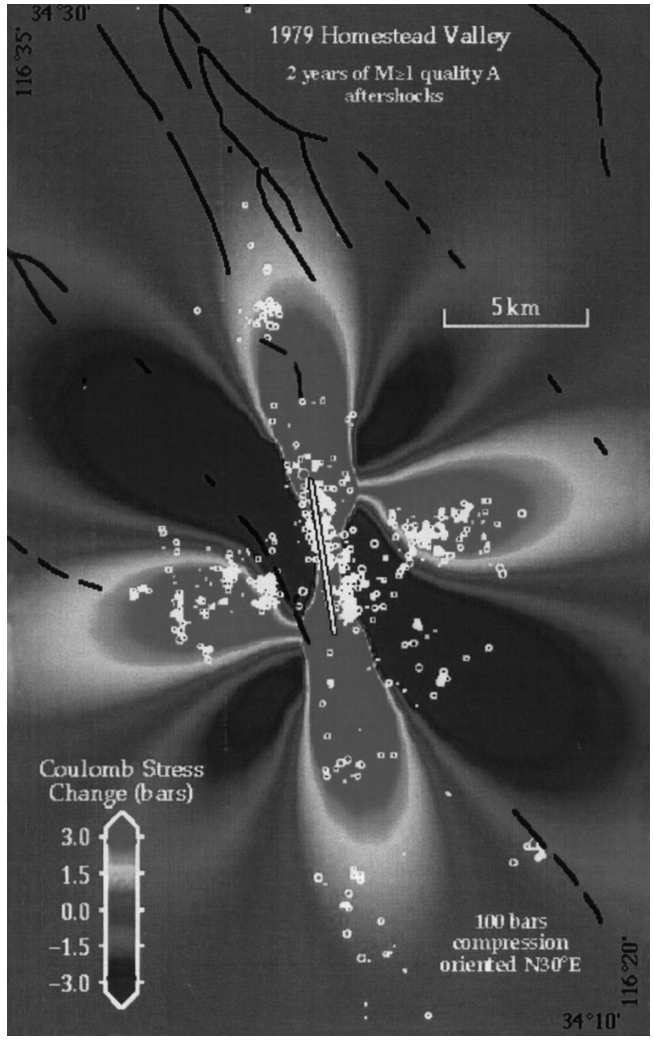

Fig. 4.26. Coulomb stress change induced by the 1979 Homestead Valley earthquake. While line is the mainshock rupture and white symbols are aftershocks. (From King and Cocco, 2000.) (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

the mainshock rupture, there were a considerable number of off–fault aftershocks. The great majority of these occurred in positive ΔCFS regions, with quiescence in negative ΔCFS regions, exhibiting the stress drop shadow effect. Changes of ΔCFS of as little as 1 bar separate regions of enhanced and decreased seismicity.

Coulomb stress changes caused by the four $M > 5$ earthquakes just prior to the 1992 Landers earthquake are shown in Figure 4.27(a). These earthquakes, the 1975 M 5.2 Galway Lake, 1979 M 5.2 Homestead Valley, 1986 M 6.0 North Palm Springs, and 1992 M 6.1 Joshua Tree, progressively increased Coulomb stresses by about 1 bar at the site of the Landers

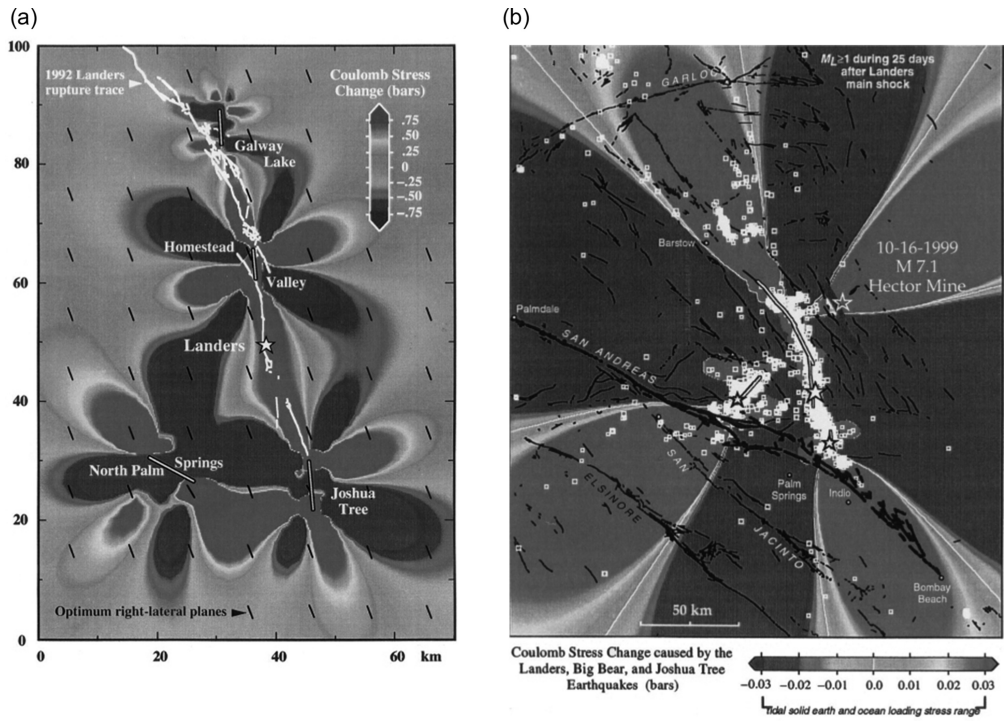

Fig. 4.27. (a) Coulomb stress changes from the major earthquakes that occurred within a few years preceding the 1992 Landers earthquake. Coulomb stress has been enhanced along most of the Landers rupture zone. (From King and Cocco, 2000.) (b) Coulomb stresses produced by the Landers earthquake and the earthquakes that just preceded it. (From King and Cocco, 2000.) (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

epicenter. They also enhanced Coulomb stresses by 0.7–1.0 bars along 70% of the rupture plane of the future Landers earthquake.

The combined effect of the Joshua Tree and Landers events is shown in Figure 4.27(b). There it can be seen that the Big Bear earthquake, which occurred 12 hours after the Landers earthquake on a left-lateral conjugate fault, as well as the cluster northwest of the end of the Landers rupture, both occurred in positive ΔCFS lobes. The overall statistics for static stress change enhancement of the off-fault aftershocks in this sequence is given by Seeber and Armbruster (2000). The Hector Mine earthquake of 1999 culminated a series of smaller events in the same location that were initiated immediately after the Landers earthquakes. At this location $\Delta\tau$ was negative (left-lateral) and $\Delta\sigma_n$ was also negative (clamping), according to the model of King and Cocco (2000). The resulting ΔCFS could be positive at that point only if $\mu \geq 0.8$ (Parsons and Dreger, 2000). However, Harris and Simpson (2002) found that for most models the Hector Mine hypocenter would be in a stress shadow from Landers, so this result makes it equivocal that the Hector Mine earthquake was triggered by Landers. We shall discuss in the next section a solution to this quandary.

Coseismic stress changes are also known to accelerate or decelerate the slip rate on creeping faults. The M 6.8 1968 Borrego Mountain earthquake triggered aseismic surface slip on the Superstition Hills, Imperial, and San Andreas Faults, some distance away (Allen et al., 1972). The Landers, Joshua Tree, and Big Bear earthquakes all triggered creep events on faults to the SW, within the positive ΔCFS lobe of Figure 4.27(b) (Bodin et al., 1994). The stress shadow of the Loma Prieta earthquake on the San Andreas Fault in northern California caused creep to stop on the nearby Hayward Fault, which resumed when tectonic loading erased the shadow (Lienkaemper, Galehouse, and Simpson, 1997).

On a longer timescale, combining ΔCFS from earthquakes with a tectonic loading model for active plate boundaries like the San Andreas and North Anatolian faults shows that most moderate to large earthquakes occur progressively in regions of positive ΔCFS (Deng and Sykes, 1997; Nalbant, Hubert, and King, 1998). Similarly, stress shadows produced by large earthquakes produce seismic quiescences which last until they are overcome by tectonic loading (Jaumé and Sykes, 1996; Harris and Simpson, 1996; 1998; see also Sections 5.4.2 and 7.2.2).

It appears that static stress changes as low as 0.1 bar can trigger observable seismicity (Reasenberg and Simpson, 1992; King, Stein, and Lin, 1994). This is just a small fraction of earthquake stress drops. These very low triggering levels require that the triggered events be very close to their rupture point. The commonness of this phenomenon suggests that there are some fault segments everywhere that are very near their critical point, a conclusion that is also reached from observations of induced seismicity (Section 6.6). Is there a threshold for the triggering stress? We defer this question until Section 4.5.3.

4.5.2 Mechanisms for the time delay

The above examples show clearly that earthquakes are often triggered by Coulomb stress transfer. Simple Coulomb friction, however, offers no explanation for the time delays that are observed, which range from a few tens of seconds to decades. The time delays, Δt, are in thus in the range

$$t_t \leq \Delta t \ll T_R \tag{4.24}$$

where t_t is the travel time of seismic waves from the causative earthquake and T_R is the earthquake recurrence time. Coulomb stress transfer will result in a clock advance (or retardation) Δt_C of the seismic cycle with respect to the expected failure time. With a Coulomb friction model the clock advance will be independent of the time in the earthquake cycle when the stress transfer occurs, as illustrated in Figure 4.28. Thus, if we assume a population of faults, moving at the same geologic slip rate but at different stages of their seismic cycle, all subjected to the same stress transfer, the clock advance for each would be the same, and hence the seismicity rate would not change.

Visco- and poro-elasic effects.

The viscoelastic response of the crust to an earthquake, discussed more thoroughly in Section 5.2.3, introduces a limited mechanism by which nearby earthquakes can be triggered after a time delay. In the case of the Landers and Hector Mine earthquakes discussed earlier (Section 4.5.1), Freed and Lin (2001) showed that the postseismic viscoelastic relaxation of the lower crust and upper mantle in response to the Landers earthquake in the 7 years between

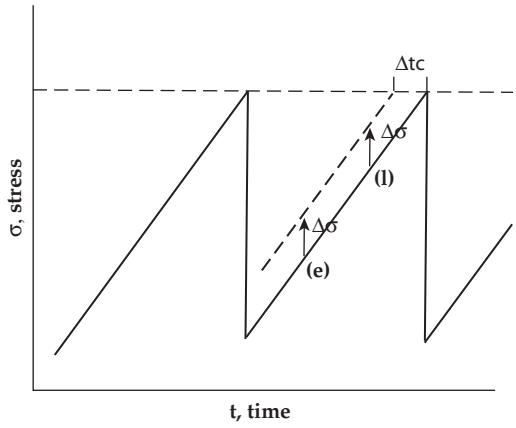

Fig. 4.28. Diagram showing an earthquake loading cycles for a fault obeying Coulomb friction. A stress transfer of an amount $\Delta\sigma$ at any time during the loading cycle will produce the same clock advance Δt.

those earthquakes would have increased the Coulomb stress function by 1–2 bars at the hypocentral location of the Hector Mine earthquake. (This occurs because the postseismic deformation effectively deepens the mainshock rupture depth.) This appears to have been the critical factor in triggering that earthquake, as well as an explanation for the time delay between the two earthquakes.

The poroelastic effect may also play an important role in some cases. The M 6.2 Elmore Ranch earthquake (Figure 4.29(a)), a left–lateral strike–slip earthquake, was followed 12 hours later by the M 6.6 right–lateral Superstition Hills earthquake (Figure 4.29(b)). The latter propagated from a point just abutting the end of the Elmore Ranch rupture. The ΔCFS at that point was almost entirely due to a reduction of the normal stress by about 30 bars (M. Cocco, pers. comm., 2000). However, because of the poroelastic effect this would have been accompanied by an instantaneous reduction of pore pressure almost as large. The immediate ΔCFS would therefore be quite small (Equation (4.23)), but would increase with time as the pore pressure recovered (see Section 5.2.3, Figure 5.14 for a clear example of this). In this case, then, the recovery of pore pressure in the 12 hours following the Elmore Ranch earthquake must have been responsible for bringing the Superstition fault to failure at that time.

Accelerating rate models

The RS friction model and the subcritical crack growth models both feature slip rates that accelerate with time in the earthquake cycle. In these cases, because of the cumulative nature of a static stress transfer, the clock advance decreases with the time in the seismic cycle when the stress transfer occurs (Gomberg, 2001). If we envision a population of faults at different stages of their seismic cycles subjected to a stress transfer, in this case the effect is of a temporal bunching together of the failure times of those faults and, as a result, the seismicity rate will increase. Thus, these processes can explain an extended period of increased seismicity following an earthquake. With these models the rate of seismicity will decrease according to the Omori Law both for the traditional on–fault aftershocks and the off–fault seismicity that has been triggered by static stress changes (Gomberg, 2001).

Triggering of large earthquakes

The finding that triggering occurs even though $\Delta CFS \ll \Delta\sigma_s$, together with Equation (4.24), presents a problem in the case of the triggering of large earthquakes. These observations

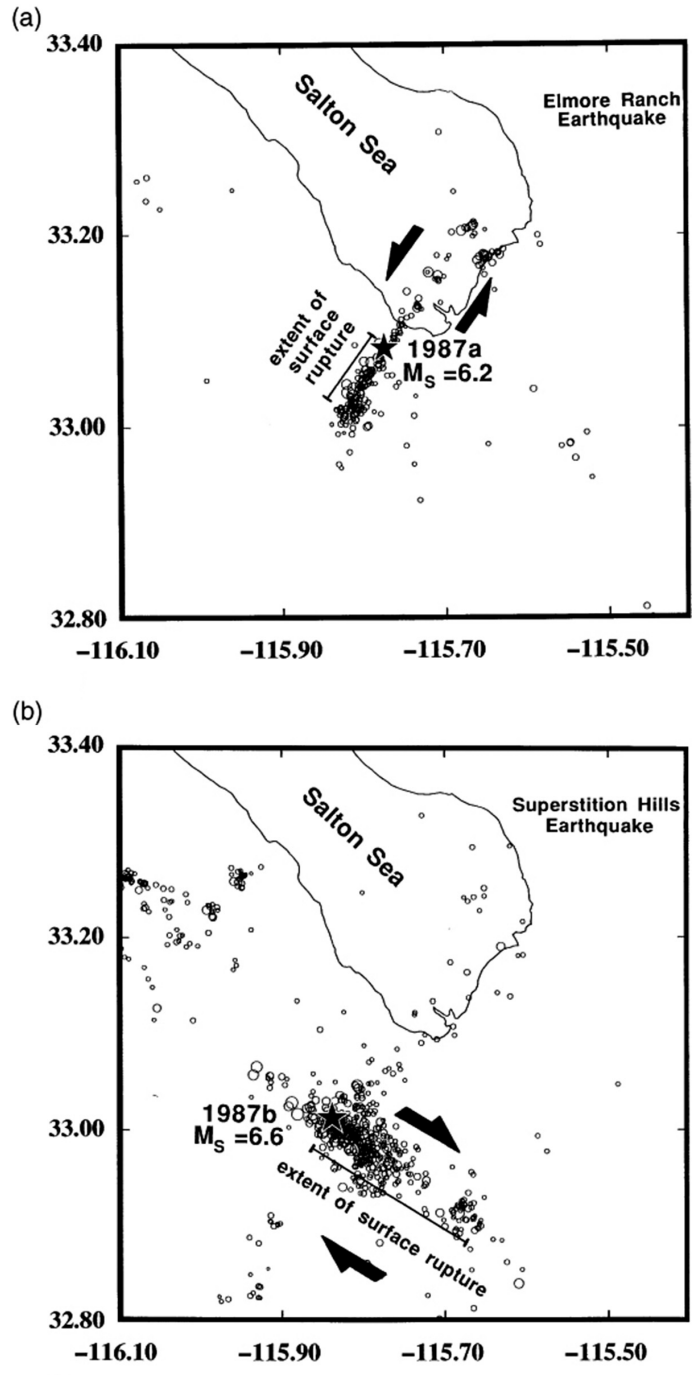

Fig. 4.29. The compound earthquake sequence of November, 1987 near Superstition Hills, California. The left-lateral Elmore Ranch fault ruptured first in an $M_s = 6.2$ event (a), followed 12 h later by rupture of the right-lateral Superstition Hills fault in an $M_s = 6.6$ event (b). The second earthquake initiated from a point near the abutment of the two faults. Stars are the mainshock epicenters. (From Hudnut et al., 1989.)

require that the triggered earthquakes must be extremely close to their natural failure points. As with the similar case of induced seismicity (Section 6.6), we can rationalize this by assuming that there must everywhere be a large population of small faults, some subpopulation of which may be close enough to failure that they can be triggered with very small values of ΔCFS. However, in the case of the triggering of large earthquakes, as in the case shown in Figure 4.29, that rationale cannot be used, because there are only a few faults large enough to host such earthquakes. The Superstition Hills fault must have been at the very end of its seismic cycle in order to be triggered by the Elmore Ranch earthquake. If this was a rare occurrence this might be dismissed as coincidence, but in fact it is not rare – many such similar cases are described in Scholz (2010a). The reason that it is widely believed that the Hector Mine earthquake was triggered by the Landers earthquake is that both of those earthquakes occurred on slowly moving faults with recurrence times of about 5,000 years, so it is hard to believe that the occurrence of two large earthquakes on these nearby faults within a 7–year interval was by chance. Those earthquakes occurred within a set of subparallel strike–slip faults that together constitute the Eastern California shear zone, shown in Figure 4.30(a). The geologic slip rates and times of previous large earthquakes on these faults, obtained by paleoseismological means, are shown in Figure 4.30(b). The Camp Rock and Pisgah fault systems, the site of the Landers and Hector mine earthquakes, both slip at about the same rate and ruptured previously at 4–5,000 and again at about 9–10,000 years ago. Thus, their seismic cycles have been in synchrony over the past three cycles. The slower–moving Helendale and Lenwood faults also appear to rupture in near synchrony (the Lenwood also ruptured in sync with the other synchronous pair at ~5 kbp). Scholz (2010a) argued that nearby faults that have similar slip rates can become synchronized in their seismic cycle over many cycles when they receive positive coupling through their exchanges of stress transfers. The faults behave like *coupled oscillators* that may synchronize if their natural periods are close enough to be within an *entrainment threshold*. Note that the faster–moving Calico fault is not synchronized with the faults on either side, even though it must receive large stress transfers from those faults. Similarly, although the earthquakes in the eastern California shear zone during the 1990s transferred significant stresses to the nearby sector of the much faster–moving San Andreas Fault (Freed and Lin, 2002; Jaumé and Sykes, 1992), they did not induce any increased seismicity there. The synchronization of coupled oscillators, both animate and inanimate, is a common phenomenon (e.g. Strogatz, 2003; Strogatz, 2001). In a historically famous case, Christiaan Huygens observed that two identical pendulum clocks, mounted side–by–side in his laboratory, would, after a brief time, become synchronized. This was experimentally reproduced and explained with a simple theory of stress transfer between the pendulums by Bennett et al. (2002). For a general review of the *Kuramoto model*, the canonical model for coupled oscillators, see Acebron et al. (2005). For theory applied to the types of cases discussed here, the synchronization of coupled oscillators that obey RS friction, see Sugiura et al. (2014).

Such synchronization may, of course, also occur between adjacent segments of a single fault. Figure 4.31 shows a sequence of large earthquake that progressively ruptured the North Anatolian fault from east to west in 1939–1999. Most of these may be considered to have been triggered by static stress loading from the previous earthquakes (Stein et al., 1997). Note, however, that the 1943 earthquake initiated at its distal end from the prior earthquake, where there was no change in the Coulomb stress function, and then propagated from west to east. Thus, it must have occurred naturally in sync with the other earthquakes in this sequence. This underscores the point that even those earthquakes that have been considered triggered by static stress changes must have occurred on fault segments already in synchrony in their seismic cycles. Reinforcing this view, the paleoseismic record for the North Anatolian fault

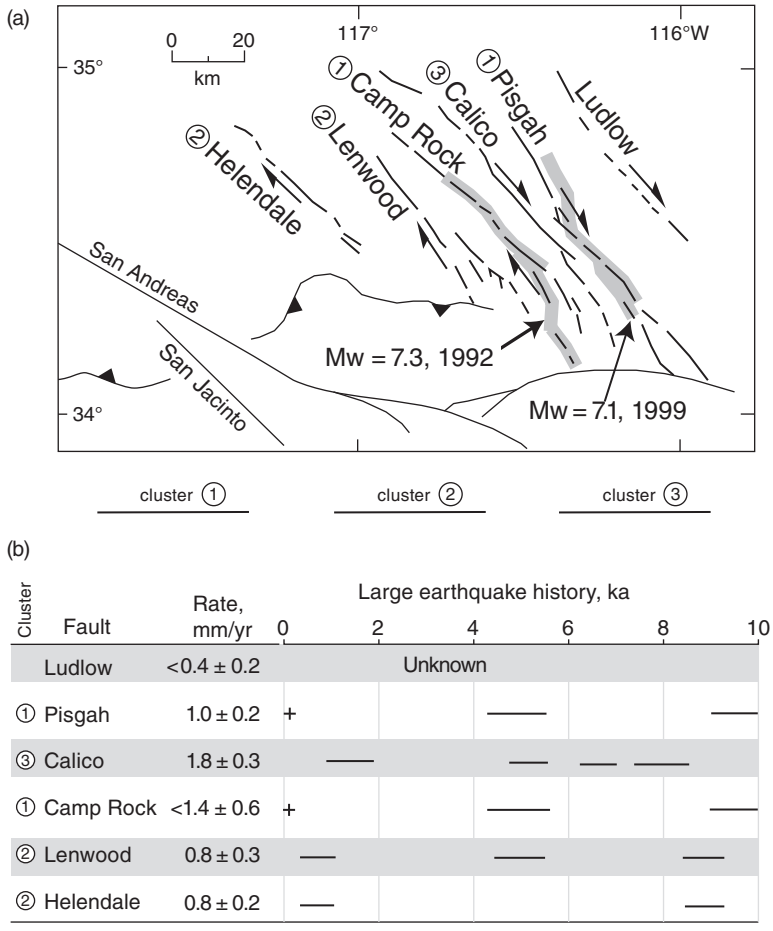

Fig. 4.30. (a) Faults of the eastern California shear zone, Mojave region, southern California. (b) Fault slip rates and paleoseismological histories. The faults are grouped into clusters based on their slip rates and earthquake history. There are two pairs of synchronized faults and two independent faults.

indicates that such sequences have occurred every few hundred years, with the fault being relatively quiescent during the intervening intervals (Hartleb et al., 2006).

On a smaller scale, coupling between nearby fault patches results in the sort of *phase locking* that causes the seismicity of faults to be dominated by occasional system-size large earthquakes rather than by continuous small earthquakes (Brown et al., 1991; Herz and Hopfield, 1995). This is the underlying reason why individual faults and fault segments obey the characteristic size distribution rather than the Gutenberg–Richter law.

Statistical clustering theories

These theories consider earthquakes to be a stochastic point process. Each earthquake is chosen randomly from a truncated Gutenberg–Richter distribution and assumed to trigger

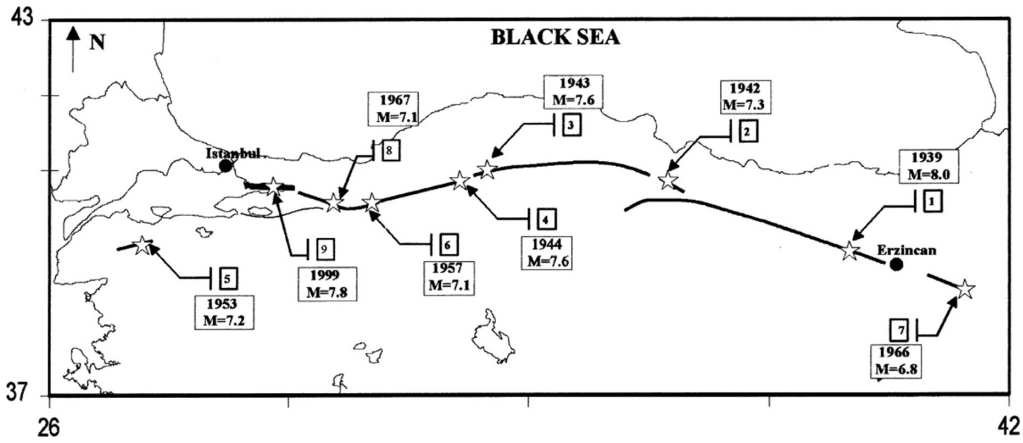

Fig. 4.31. Rupture zones along the North Anatolian fault associated with a sequence of large earthquakes beginning in 1939. Although the western progression of these earthquakes is generally remarked on, the 1943 earthquake initiated at its western end and propagated to the east. (Figure courtesy of Ali Osman Oncel.)

subsequent earthquakes (aftershocks) following a productivity function that depends on its magnitude (Equation (4.21)). Models in which each earthquake has a single parent are called branching models (Felzer et al., 2004; Kagan, 1991). Those that include multiple cascades are called epidemic (ETAS) models (Helmstetter and Sornette, 2002; Ogata, 1998; 1999). These models are species of what are called statistical physics models. For broad reviews of the latter topic as related to earthquake physics, see Kawamura et al. (2012) and Vallianatos et al. (2016).

There is no intrinsic time scale in these clustering models so no distinction is made between instantaneous and delayed triggering. A clock rate is introduced based on the observed rate of seismicity, either than of aftershocks or of background seismicity. In these models the size of the triggered event bears no relation to that of the triggering event, hence there is no distinction between foreshocks, mainshocks, and aftershocks (Felzer et al., 2004). From this viewpoint, the practice of centering the time–frame on the largest event in a sequence and calling it the mainshock is arbitrary.

With some choices of parameters, these models can simulate the Omori Law for aftershock decay, and also Båth's law (Helmstetter and Sornette, 2003a; Shearer, 2012). They also predict that foreshocks should obey an inverse Omori Law (Helmstetter and Sornette, 2003b), as discussed in Section 7.2.2.

Such models are useful in explaining the overall clustering characteristics found in earthquake catalogues, but not so much in exploring the underlying physics. Delayed and instantaneous triggering suggest different processes at work, for example. One also needs to make a distinction between classical aftershocks that occur on or very near the rupture surface of the mainshock and off–fault aftershocks that may occur by static or dynamic triggering at greater distances. The classical aftershocks arise because the slip distribution within earthquakes is strongly heterogeneous, leading to local stress increases on or near the fault. These stresses are relaxed by those aftershocks, hence their spacial distribution reflects the spatial distribution of coseismic slip, as noted earlier (e.g. Figure 4.21(c)). Off–fault aftershocks, on the other hand, are triggered by the relatively smooth far–field static and dynamic stress fields. They represent

places where distant faults happen to be very close to their critical points. Thus, the classical aftershocks are a secondary process, and this very nature is reflected in Båth's law – i.e., the mainshock relaxed that section of fault, and the aftershocks represent a clean–up operation, smoothing out the rough edges left by the mainshock. On the other hand, the distant triggered events may be of any size. If they were included in our definition of aftershocks this would often lead to violations of Båth's law. An example is if the Hector Mine earthquake were to be considered an aftershock of Landers. These points were made by van der Elst and Shaw (2015), who showed that larger aftershocks are more distant than smaller ones.

The clustering models play several important roles. They are a parsimonious description of the statistical observations that must be satisfied by any physics–based model. They also provide a null hypothesis to separate unexpected from expected clustering. An example, regarding foreshocks, is shown in Figure 7.9. They also are used to estimate enhanced seismic hazard due to aftershocks (Gerstenberger et al., 2005; Reasenberg and Jones, 1989) and in that way are an important aspect of operational earthquake forecasting (Section 7.4.3).

4.5.3 Dynamic triggering

The triggering of remote earthquakes by the passage of seismic waves was discovered by Hill et al. (1993), who observed an abrupt increase in microearthquake activity at the Long Valley caldera upon the arrival of surfaces waves from the Landers earthquake several hundred km to the south. As mentioned earlier, the 2002 Denali earthquake triggered earthquakes across western North America at epicentral distances of up to 3,660 km, as shown in Figure 4.32 (Eberhart–Phillips et al., 2003; Gomberg et al., 2004; Prejean et al., 2004). These triggered earthquakes began after the P wave arrival and during the arrival of the largest amplitude surface waves. The locations of the triggered seismicity down range in the direction of rupture propagation of the Denali earthquake clearly demonstrates the focusing of energy due to the directivity effect. The energy flux in the S waves at the range of distances where triggering occurred was 4–5 times larger than recorded at stations in Europe and Asia at similar distances in the back azimuth direction (Eberhard–Phillips et al., 2003). For a review of dynamic triggering see (Brodsky and van der Elst, 2014).

It is clear that triggering at such great distances must be caused by dynamic stresses in the seismic waves rather than by static stress changes, which would be insignificant at such distances. But there is no reason why we should not expect dynamic triggering to occur at short epicentral distances as well. To distinguish such dynamic triggering from static triggering at short distances, Kilb et al. (2000) and Gomberg et al. (2001) took advantage of the directivity effect to demonstrate near–in dynamic triggering for the Landers and Hector Mine earthquakes. Off–fault aftershocks were dominant to the north of Landers and to the south of Hector Mine, corresponding to the rupture directions of each earthquake.

There was a global increase in seismicity lasting 6 days following the 2012 M_W 8.6 Wharton Basin earthquake (Pollitz et al., 2012) (see Section 6.4.3). That strike–slip earthquake was the largest intraplate earthquake ever recorded. The triggered earthquakes, M≥5.5, occurred on the four lobes of maximum Love radiation from the mainshock. Other observations of triggering at teleseismic distances are reviewed by Parsons et al. (2014). The number of earthquakes of magnitude ≥ 6 globally that follow within one day a mainshock of magnitude M is shown in Figure 4.33. The dashed line is the fit to the data of Equation (4.21). In a simple conceptual model, a dynamic stress of magnitude s will trigger all faults in which $s \geq f$, the difference between the pretrigger stress and the failure strength (Brodsky and van der Elst, 2014). The agreement of the data in Figure 3.33 with Equation (4.21) means that the distribution of faults is

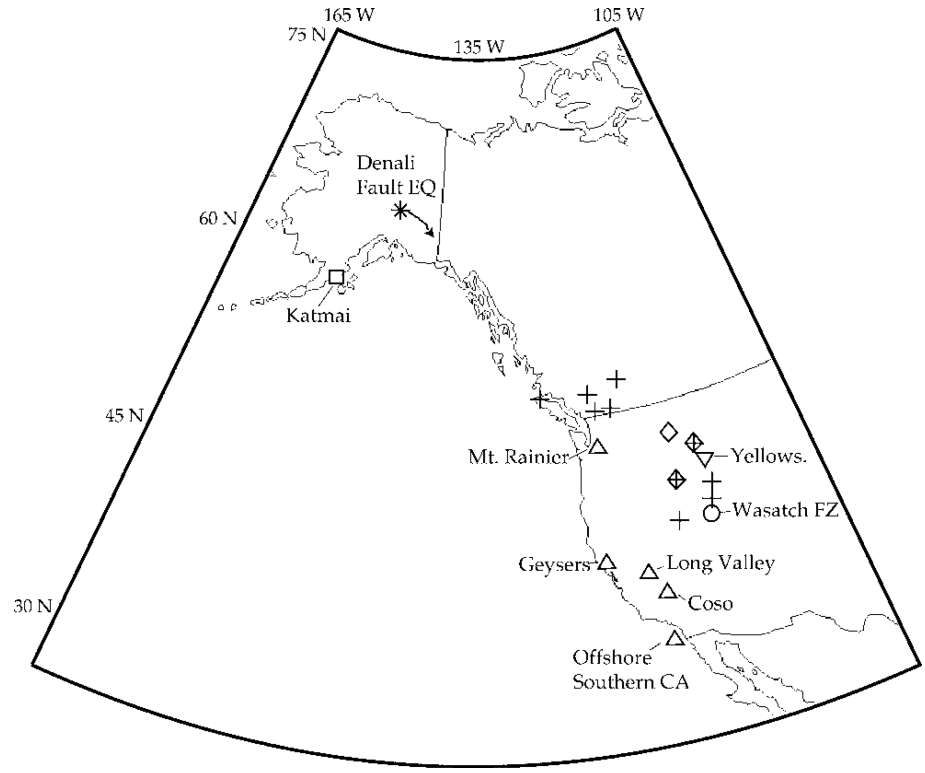

Fig. 4.32. Map of North America showing location of the Denali earthquake (star), its rupture propagation direction (arrow), and locations of remotely triggered seismicity. (From Prejean et al., 2004.)

uniform over all values of f. This would occur if the faults in the system have consistent strengths and are evenly distributed over their loading cycles. This supports the underlying assumption in the seismicity model of Dieterich (1994).

In a statistical study of the frequency of triggered earthquakes as a function of distance it was found that the rate of triggered events decreases with distance at a rate consistent with it being proportional to the peak dynamic strain (Felzer and Brodsky, 2006). Those authors proposed that this might mean that all aftershocks are dynamically triggered, but their result was contradicted by Richards–Dinger et al. (2010). In a later study it was found that extrapolating this relation back to the near field accounted for only 15–60% of the near–field aftershocks – presumably the remainder were produced by static triggering (van der Elst and Brodsky, 2010). The above conclusions were based on small earthquakes treated as point sources. Powers and Jordan (2010) found that seismicity rate decays from faults with distance to a power ~–1.5. These results agree with the predictions of a 2D model of static triggering from a fractal fault (Dieterich and Smith, 2009).

Near instantaneous dynamically triggered earthquakes within a few fault lengths of a large earthquakes have been found to be fairly common (Fan and Shearer, 2016). Such events are lost in the coda of the mainshock waveform – Fan and Shearer identified them with the back projection technique. These triggered events need not be much smaller than the forerunning

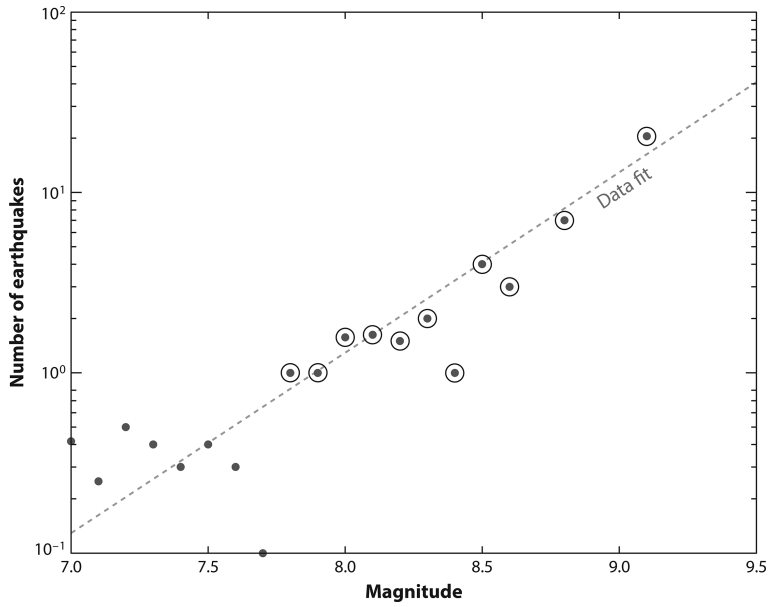

Fig. 4.33. The number of earthquakes of magnitude ≥ 6 globally in 1 day following a mainshock of magnitude M. Mainshocks are defined here as earthquakes with no larger event within 4,500 km and 2 years prior. The dashed line is the fit to Equation (4.21). (From Brodsky and van der Elst, 2014.)

earthquake. In 2009, an M_W 8.1 normal faulting earthquake on the outer rise of the Tonga subduction zone triggered shortly thereafter two M_W 7.8 underthrusting earthquakes on the other side of the trench (Beavan et al., 2010; Lay et al., 2010) – events that were first detected solely from nearby GPS measurements. The ensuing local tsunami was mainly due to the thrust earthquakes.

The mechanism for a time delay in triggering by dynamic stresses is poorly understood. How can transient stresses have a lasting effect? One possibility (Gomberg et al., 2004) is that the dynamic stresses somehow weaken faults. This perhaps could occur by dynamic stressing altering the permeability of faults and thus inducing fluid flow and pore pressure changes (Elkhoury et al., 2006; Manga et al., 2012; Parsons et al., 2017). The Parsons et al. paper provides some evidence that fluid diffusion is involved in producing the time delay. Another suggestion is that the delayed events are the result of cascades of aftershocks (by static stress changes) of dynamically triggered earthquakes (Brodsky, 2006). There has been some success in experimentally triggering stick–slip events in the laboratory with oscillating loads (Savage and Marone, 2008). Those results can be partially explained by simulations with some versions of the RS friction law within restricted parameter ranges (van der Elst and Savage, 2015). This issue is discussing in more detail by Brodsky and van der Elst (2014).

The fractional seismicity rate change for dynamically triggered earthquakes in California is shown in Figure 4.34 (Brodsky and van der Elst, 2014). These data imply that there is no lower limit for dynamic stress triggering – or, at least, that it is exceedingly low: a dynamic strain of $\sim 3 \times 10^{-9}$ corresponds to a stress of 100 Pa! This contrasts with

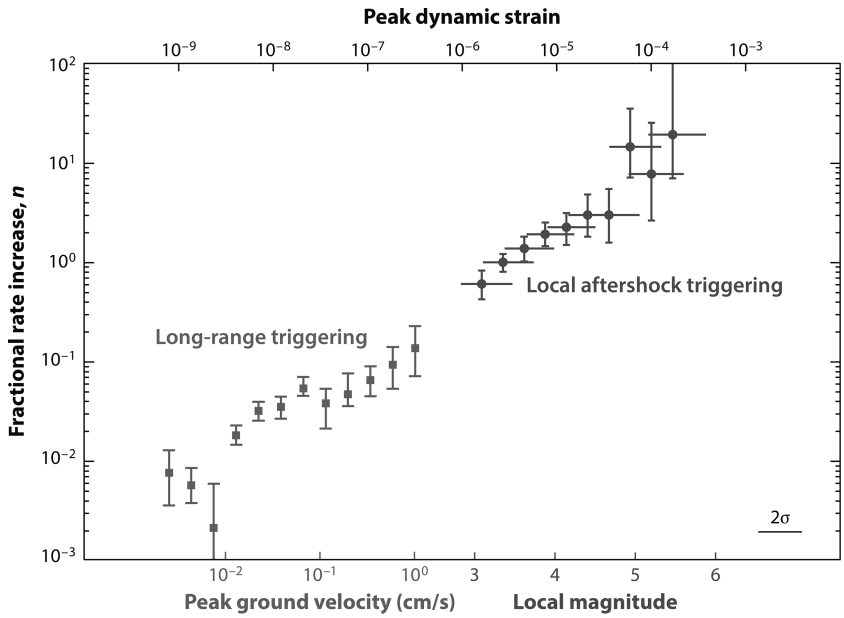

Fig. 4.34. Triggered earthquake rate change versus peak dynamic strain. Long-range triggering is at least 800 km distant. Local is less that 6 km. (Modified from Brodsky and van der Elst, 2014.)

experience from induced seismicity, which always indicates a lower stress threshold to be exceeded for triggering earthquakes in the 0.1–1.0 MPa range, as in the Rangely experiment (see Sections 2.5 and 6.6).

Whatever the mechanism involved, the dynamic stresses at these remote triggering sites are very small. Hence, the faults involved must have been very near their critical points for such triggering to occur. It is observed that sites of hydrothermal activity show the most sensitivity to dynamic triggering. This perhaps indicates that elevated fluid pressures at such sites are responsible for bringing the faults very close to their failure stress. Indeed, sites of recent fluid injection in boreholes are particularly sensitive to triggering from remote earthquakes (van der Elst et al., 2013). These same issues will be revisited with respect to the related phenomena of earthquakes induced by reservoir impoundment or fluid injection in boreholes (Section 6.6).

Triggering from tides

Correlations of earthquakes with tides have long been sought (Emter, 1997). The finding that earthquakes can be dynamically triggered by seismic waves at very small stress levels has reinvigorated that search. Many such attempts to find a correlation have been negative (e.g. Hartzell and Heaton, 1989; Vidale et al., 1998; Wang and Shearer, 2015). The exceptions are in regions of hydrothermal activity, which, as in the case of dynamic triggering, seem to be particularly susceptible (Hurwitz et al., 2014; Stroup et al., 2009). This lack of sensitivity to tidal triggering has been ascribed to the long duration of earthquake nucleation relative to tidal periods (Beeler and Lockner, 2003).

Ocean tides often have stress magnitudes an order of magnitude larger than solid earth tides (~10 kPa versus ~1 kPa), and Cochran et al. (2004) found, as a result, a correlation of tides with shallow submarine underthrusting earthquakes at Coulomb stress levels >20 kPa. The primary triggering correlation was from the reduction of the normal stress at low ocean tides. Similar results were obtained by Wilcock (2009). Ide et al. (2016a) presented evidence that large $M_W > 8.2$ earthquakes of the same class tend to occur during high tidal shear stresses and that in such regions the b-value decreases with increasing tidal stress.

In a continental setting, Bucholc and Steacy (2016) found a weaker correlation for reverse faults, particularly shallow ones, with solid earth tides at a stress levels >1 kPa, but no correlation for strike–slip earthquakes, for which the tidal stresses are smaller.

Strong tidal triggering was observed within the focal regions and just before the great Aceh–Andaman (Sumatra) earthquake of 2004 and the Tohoku–oki earthquake of 2011 (Tanaka, 2010; 2012). The tidal correlation vanished after the great earthquakes, suggesting that triggering becomes important only when the crust is critically stressed.

4.6 SLOW EARTHQUAKE PHENOMENA

With the installation of networks of continuously recording GPS receivers and other instrumentation in subduction zones and other tectonic regions, an entirely new class of seismotectonic phenomena has been discovered. These phenomena, generally classed under the heading "slow earthquakes," consist of a variety of processes by which shear motion can take place either quasi-statically on fault surfaces with little or no seismic radiation or dynamically with seismic radiation at much lower frequencies than that of regular earthquakes. A sampler of these phenomena is shown in Figure 4.35. For a supplemental review, see Beroza and Ide (2011).

In this section we will be concerned with slow slip transients that are spontaneously generated. Afterslip, a postseismic relaxation phenomenon, will be discussed in Section 5.2.3, and steady-state aseismic slip (creep) of faults in Section 3.4.2.

4.6.1 Quasi-periodic slow slip events and tremor in subduction zones

The upper part of Figure 4.36 shows daily changes in the east component of a GPS station at Victoria, BC, relative to stable North America. This station lies on the upper plate of the Cascadia subduction zone, and the long-term trend in the data indicates the eastward compression of the upper plate in response to the convergence of the Juan de Fuca plate at a rate of 40 mm/yr. This trend is punctuated every 13–16 months by a reversal over a period of several weeks (see Figure 4.35(e) for a detail). Each of these relaxations results from a slow slip event (SSE) on the plate interface passing below the GPS station. The lower part of Figure 4.36 shows (nonvolcanic) tremor recorded at a nearby seismograph station. Tremor is a low-amplitude band-limited seismic noise that is phase incoherent when recorded by regional seismic networks. As seen in the figure, it occurs in bursts that correlate with the SSEs. The combination of these two phenomena is called episodic tremor and slip (ETS).

The slip in the SSEs is up-dip thrusting on the plate interface in a channel within the depth interval of 25–40 km, as can be seen in the examples shown in Figure 4.37. The SSEs propagate, often for several hundred km along strike, at a velocity of about 10 km/day. In the example shown in Figure 4.37(a) the SSE propagated bilaterally, first to the south and later to the north. Figures 4.37(b) and (c) show the close spatial association of tremor and slip. As can be seen from

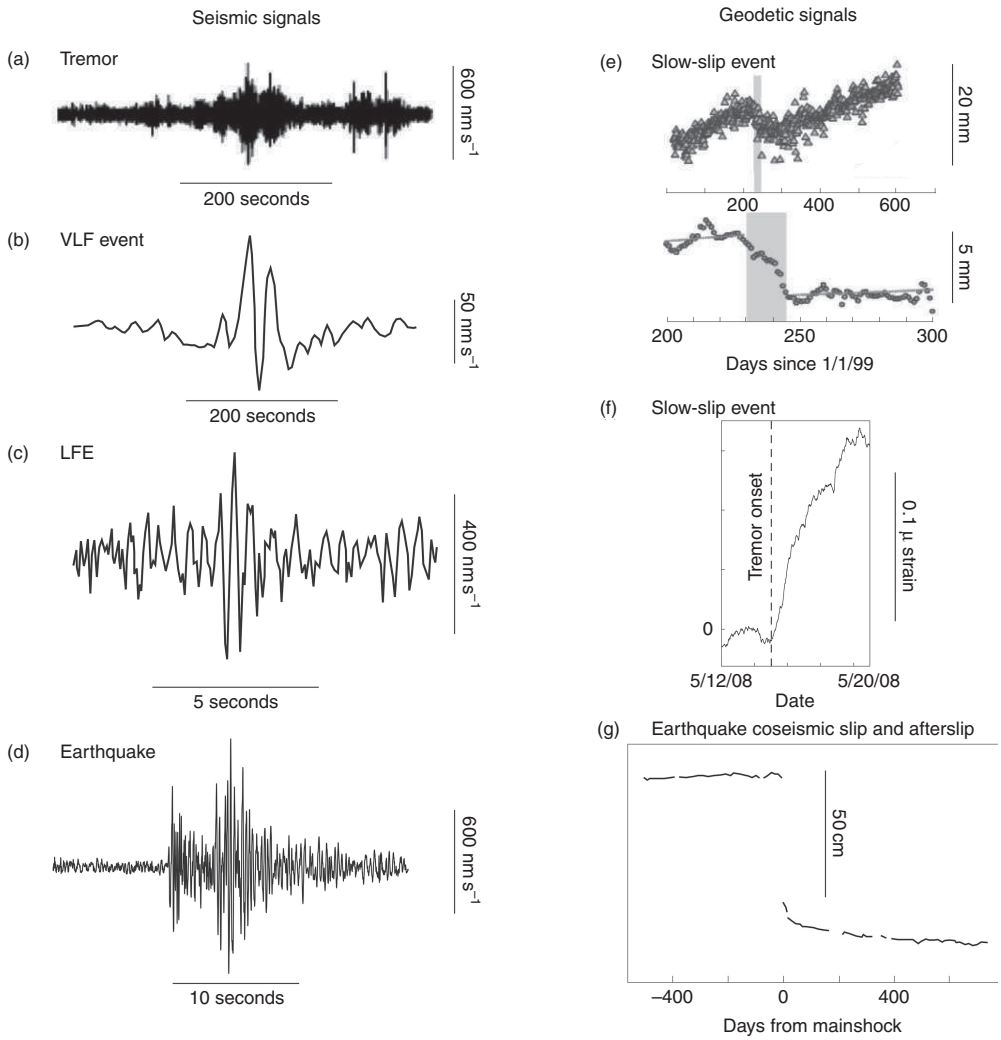

Fig. 4.35. Illustrative examples of slow slip signals: (a) nonvolcanic tremor; (b) a very low-frequency earthquake (VLF) from Japan; (c) a low-frequency earthquake (LFE) from Japan; (d) a M 1.9 earthquake from western Washington; (e) Top: daily E-W GPS displacements measured in Vancouver Island. Bottom: averaged and de-trended GPS data reveal a slow slip event (shaded); (f) a slow slip in differential shear strain measured in western Washington. The strain transient onset coincides with increased tremor activity; (g) GPS N55° displacement ~100 km from the 2001 M 8.4 Peru earthquake. The large offset reflects the coseismic slip and the subsequent decaying deformation is postseismic relaxation, probably from afterslip. (Figure from Peng and Gomberg, 2010.)

these examples, the slip amplitude in SSEs is relatively constant along strike, tapering off near the ends. The equivalent M_w of these events is in the range 6.3–6.8. In northern Cascadia SSEs relax 45–65% of the plate convergence rate within the depth range of the ETS channel (Wech et al., 2009).

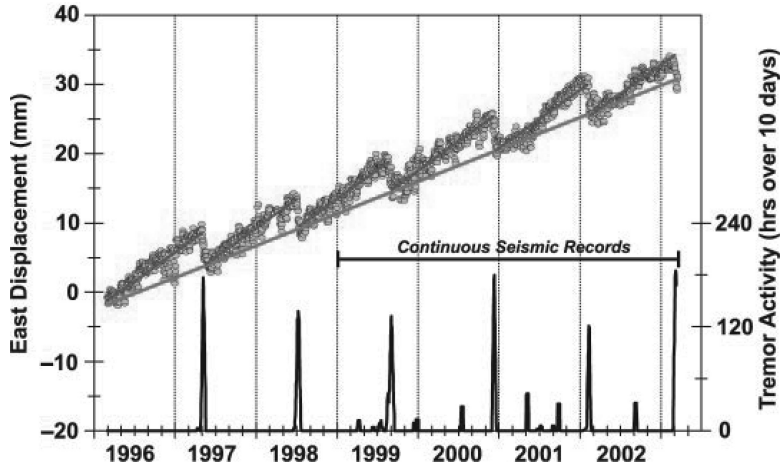

Fig. 4.36. Upper part: eastward motion of a station at Victoria, Vancouver Island, relative to stable North America. The continuous line is the steady-state eastward motion of the site that results from the convergence of the Juan de Fuca plate at 40 mm/yr. The steeper lines give the rate of "fully stuck" motion, punctuated by reversals that correspond to slow slip events at depth. Lower part: tremor activity recorded at a local seismic station. (From Rogers and Dragert, 2003.)

It has often been speculated that a great subduction earthquake might be triggered by an SSE impinging upon it from below. On a minor scale this has already been observed: an SSE event in the Guerrero section of the Mexican subduction zone appears to have triggered an M_W 7.3 interface thrust earthquake in the coupled region just above it (Radiguet et al., 2016). There are also shallow SSEs that seem to be precursory to some great subduction earthquakes. These are discussed in Section 7.2.3.

Tremor is found to be composed of swarms of low–frequency earthquakes (LFEs) and very low–frequency earthquakes (VLFs) (Figures 4.35(b) and (c)) in which the focal mechanisms of the LFEs indicate thrusting on the plate interface (Ito et al., 2007; Shelly et al., 2007). In Figure 4.37(b) and (c) are shown the locations of such tremor events compared with the geodetically determined slip. There is a very close spatial correlation between the two. The tremor also tracks the temporal migration of SSEs. In Figure 4.38 (middle panel) are shown tremor locations in northern Cascadia over the period 1997–2007. Each streak represents an ETS episode: those recorded by GPS receivers in the south (ALBH) or the north of Vancouver Island (NTKA) are linked by arrows. Figure 4.39 shows one such ETS in detail. This one propagated unilaterally at a constant velocity of 9.4 km/day. Within such an episode there are also very rapid migratory streaks in the slab dip direction that propagate at 25–150 km/hr (Shelly et al., 2007) and sweep out the slip area as the SSE propagates along strike (Ghosh et al., 2010). It has been suggested that these may be localized by irregularities on the plate interface (Ide, 2010). Alternatively, they may represent the jerkily propagating leading edge of the SSE (Ando et al., 2010).

One can judge from the width of the ETS in Figure 4.39 at any given time that the SSE propagates with a constant along–strike pulse width of about 70 km. This is approximately the same as the down–dip width of the SSE, (40–25)sinδ ~ 80 km (the dip δ of the interface in northern Cascadia is ~10° [McCrory et al., 2004]).

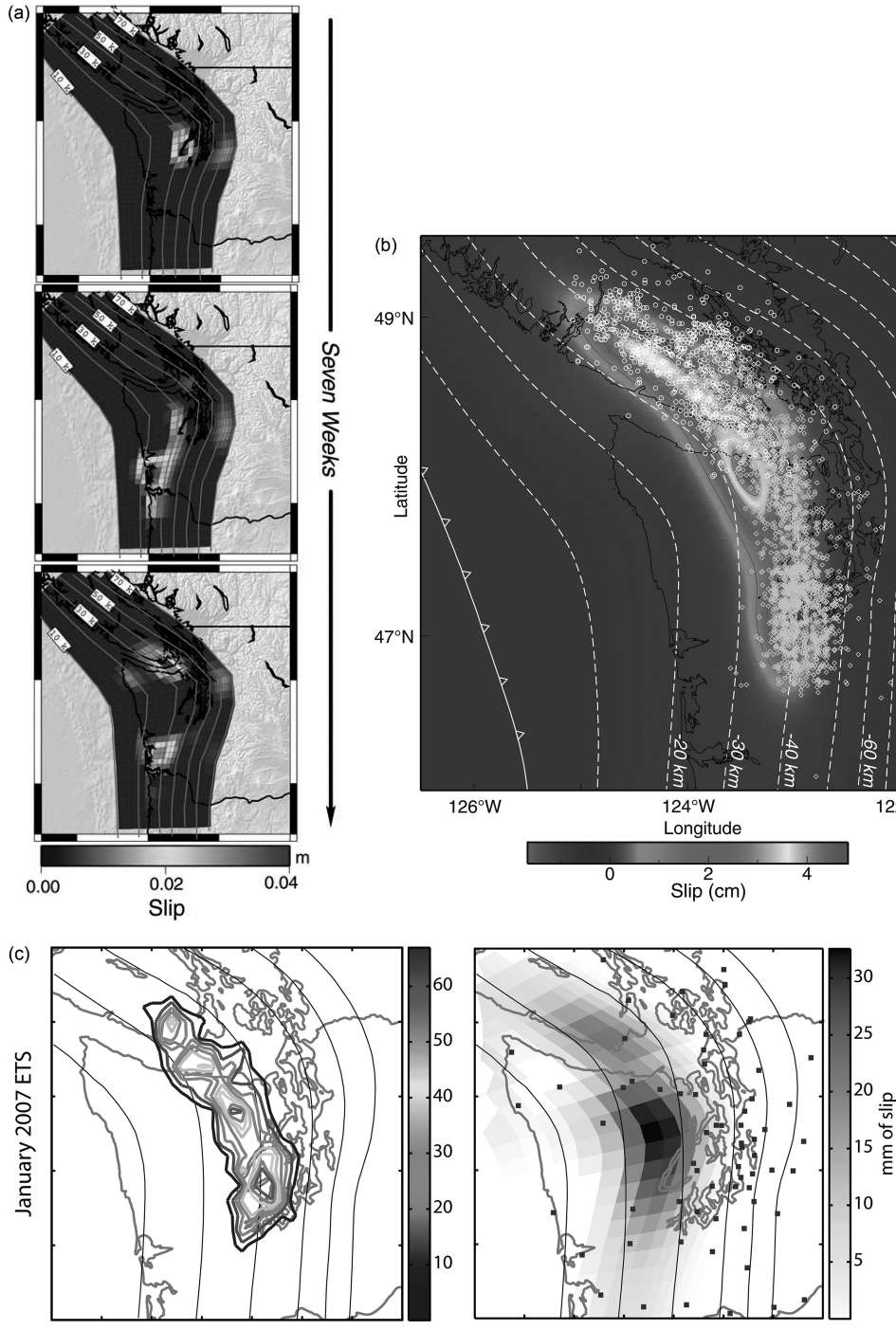

Fig. 4.37. Examples of the slip distributions in slow slip events beneath the northern part of the Cascadia subduction zone. (a) propagation of an early 2003 SSE: each frame shows the slip in an approximately two-week interval. Maximum slip was 3.8 cm (Melbourne et al., 2005); (b) slip distribution in a SSE with tremor locations overlaid (Gomberg, 2010); (c) left, summed tremor epicenters contoured and right, slip distribution of an SSE of January 2007 (Wech et al., 2009.) (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

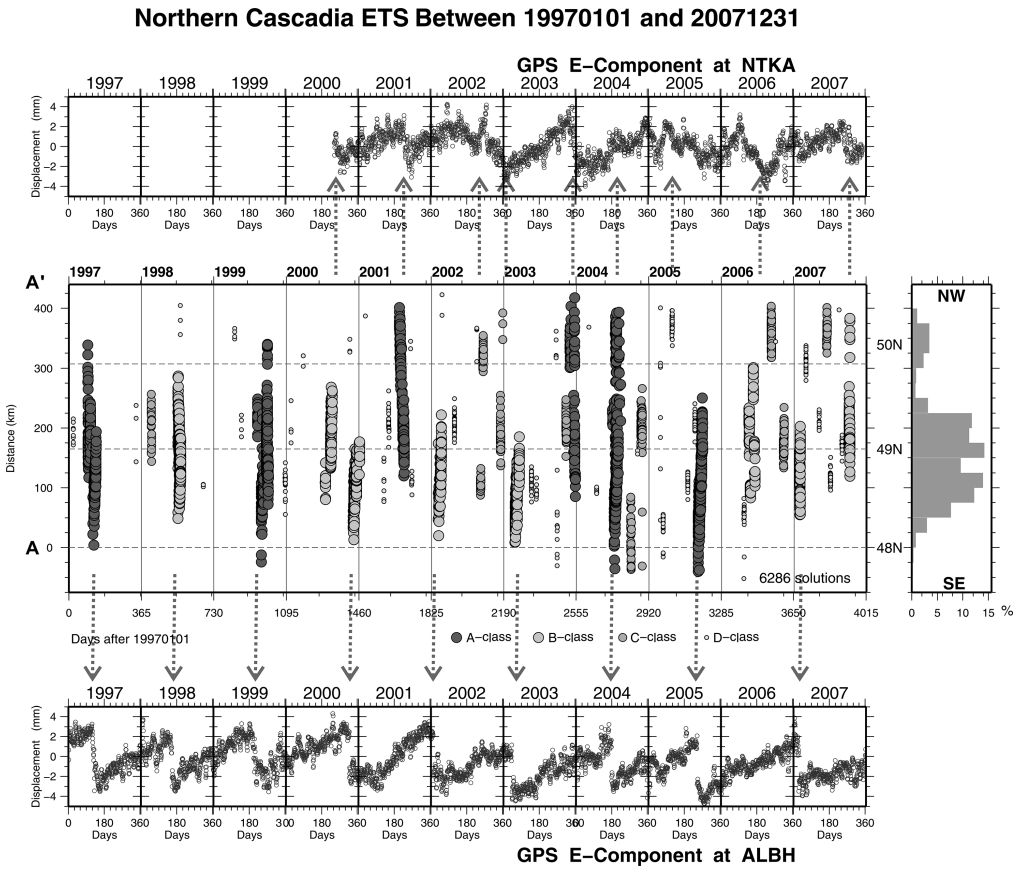

Fig. 4.38. Spatial-temporal distribution of ETS events observed in northern Cascadia between 1997 and 2007. The middle panel shows tremor locations in a north–south profile approximately parallel to the strike of the subducted slab. At top and bottom are daily east component records of GPS stations at ALBH, at the southern end of Vancouver Island and NTKA in the north of the island. (From Kao et al., 2009.)

There is also sporadic tremor that occurs between major ETS events. This occurs below the ETS channel and shows minor episodic clusters that may perhaps represent SSEs below the resolution of GPS (Wech and Creager, 2011). These clusters decrease in amplitude and period until the plate motion becomes silent at about 80 km depth.

The tremor responds to very small stress changes. It is modulated by solid earth tides – increasing when the tidal stresses enhance the Coulomb stress function (Lambert et al., 2009; Rubinstein et al., 2008). Tremor is also observed to be triggered by the passage of surface waves (Gomberg et al., 2008; Rubinstein et al., 2007) or P waves from distant (Ghosh et al., 2009a) or local earthquakes (Han et al., 2014). This has led to the conclusion that the plate interface must be very weak in the ETS channel, perhaps because of near lithostatic pore pressures there, a topic that will be taken up again in Section 6.3.3. Houston (2015), however, has argued that the friction coefficient itself must be very low, <0.1.

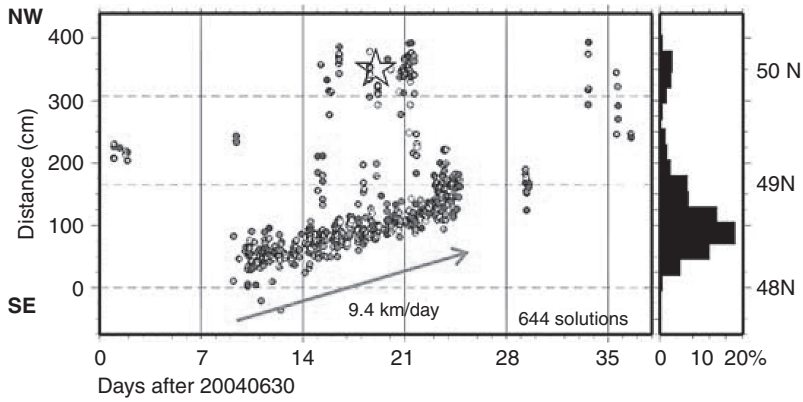

Fig. 4.39. Along-strike migration pattern for an ETS episode in northern Cascadia from July-August 2004. Some tremors were apparently triggered by a M~6.4 earthquake (star) during this episode. (From Kao et al., 2009.)

The above observations, mainly from Cascadia, apply in almost all particulars to the Nankai subduction zone of SW Japan, which is the other area that has been subject to intense study (Obara, 2010; Obara and Kato, 2016). The major difference is that the ETS episodes are more frequent there (every 6 months), and smaller in amplitude. They are below the resolution of GPS and are recognized by along–strike tremor streaks, similar to those shown in Figure 4.38, and are sometimes detected with borehole tiltmeters.

In Nankai the seismically coupled area is well defined, both by the recent great megathrust earthquakes (Tonankai 1944, M 8.0, and Nankaido 1946 M 8.1) that ruptured most of this plate boundary, and because the lower part of the coupled area is beneath land and so can be well resolved with GPS measurements. Figure 4.40(a) is a map of the region showing the location of tremor and 1 m slip contours of the 1944 M8.0 Tonankai and 1946 M8.1 Nankaido earthquakes. In Shikoku, tremor is mostly confined to the depth interval 32–38 km and closely follows the base of the rupture zones of the earthquakes.

The seismic coupling coefficient versus depth of the plate interface along profile A–A' in Figure 4.40(a) is shown in Figure 4.40(b). This result was inverted from horizontal and vertical GPS data for the period 1996–1999, with 95% confidence limits (dashed curves) (Aoki and Scholz, 2003). The wide span between the 95% confidence limits off the coast shows the uncertainty in resolving offshore deformation from land–based GPS. The data points from offshore GPS (circles with error bars) are from Yokota et al. (2016). They indicate that the upper bound is the more accurate one offshore. The locations of coseismic slip in the 1946 earthquake, the ETS channel, and the location of offshore LFEs are shown. The seismic coupling curve shows a plateau in the depth range 25–35 km, corresponding to the ETS channel. This reinforces the conclusion from Figure 4.40(a) that the ETS channel abuts the lower edge of the seismically coupled region. The LFEs and VLFs at shallow depths near the trench also represent shear on the plate interface or splay faults (Ito and Obara, 2006) and may indicate a similar stability transition at the upper edge of the locked zone. The lack of offshore instruments is perhaps preventing the observation of ETS in that location.

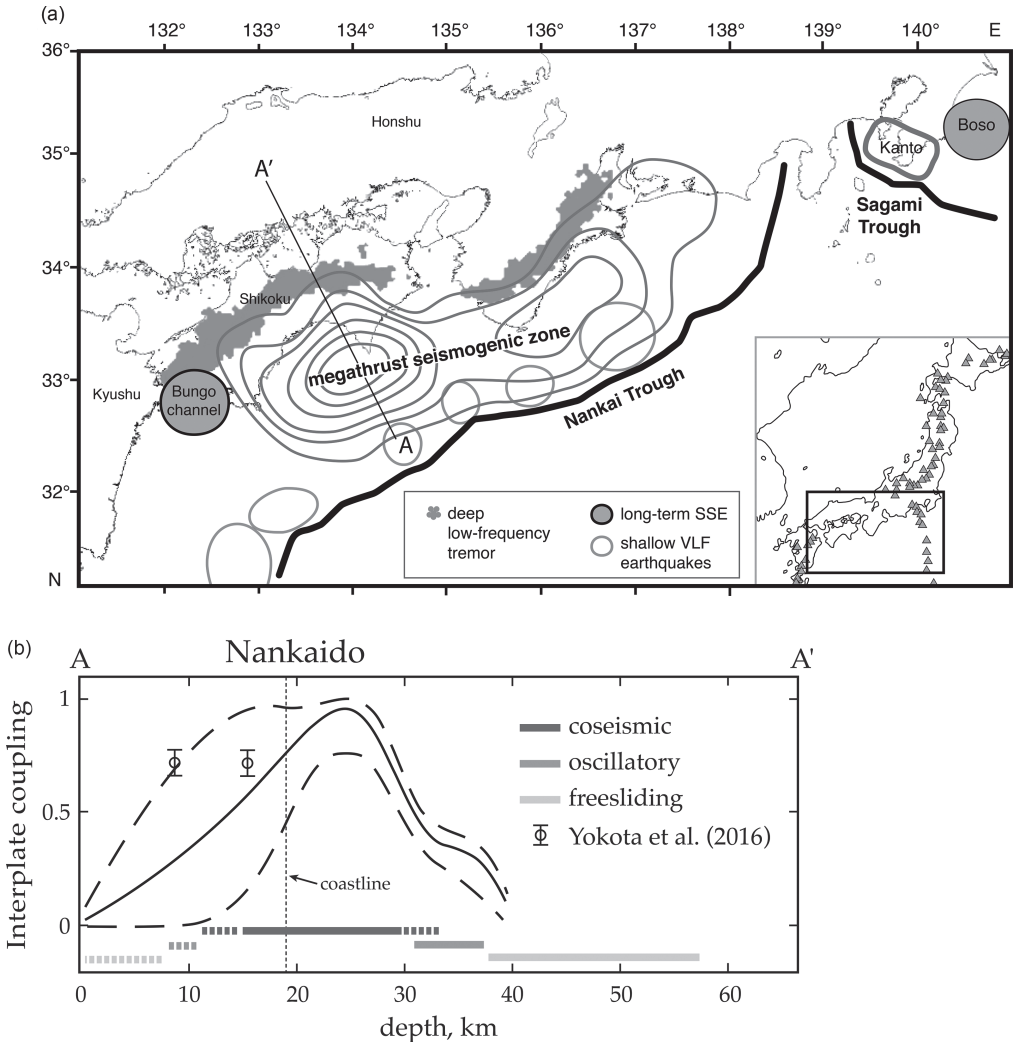

Fig. 4.40. (a) Map of location of tremor and contours of coseismic slip for southwest Japan. The coseismic slip is in 1 m contours (outer contour 1 m) from the 1944 Tonankai and 1946 Nankaido earthquakes from Sagiya and Thatcher (1999). These are based on geodetic data and must include some afterslip because the measurement points were relocated some time after the earthquakes. Tremor locations from Obara (2002). Shallow offshore VLFs are from Ito and Obara (2006). (b) The seismic coupling coefficient along profile A-A' in Figure 4.40a inverted from GPS data for the period 1996–1999 with 95% confidence limits (dashed curves) (Aoki and Scholz, 2003). The location of the coast (Muroto Point) is given by a vertical dashed line and the offshore GPS data points from Yokota et al. (2016) are given by circles with error bars. The locations of the coseismic slip in 1946, the ETS channel, and the zone of offshore LFEs are shown. (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

A very different picture arises in Cascadia. Based on a thermal model and onshore geodetic data, Hyndmann (2013) has concluded that a 70 km gap exists between the lower edge of the coupled region, which is far offshore, and the ETS channel. In this area, however, there has been no modern great earthquake that can be used to confirm the extent of the coupled zone. There is also the problem of poor offshore resolution from onshore GPS that lends uncertainty to the coupling region inferred by Hyndman (2013).

4.6.2 Other slow slip events and tremor

The ETS episodes discussed in the previous section occur in a narrow transition zone within the lower stability boundary that separates the unstable, seismically coupled plate interface and the deeper region that is stably sliding. There are also SSEs that occur where no such transition exists, particularly in subduction zones that are seismically decoupled at all depths.

The central and northern parts of the Hikurangi subduction zone are such decoupled zones. Within those regions there have been observed deep (25–45 km) SSE of duration 2–3 mo. with recurrence intervals of ~2 y in the size range *Mw* 6.3–6.6 (Wallace and Beavan, 2010; Wallace et al., 2012; Wallace and Eberhart–Phillips, 2013). A deeper event of duration 1.5 years and *Mw* 7.1 occurred over the depth range 20–70 km. These deep events are silent – they are not accompanied by tremor outbursts, although tremor does occur just below them. They are also not confined to a narrow channel but expand outwards in all directions. In the same region shallow SSEs also frequently occur, in the depth range 6–12 km. These shallow events are accompanied by swarms of microearthquakes. In the southern Hikurangi trough area, which is seismically coupled, SSEs occur below the coupling depth. A shallow short duration SSE was also observed above the coupled zone, which may be associated with the upper stability transition (Wallace et al., 2012).

Similar long duration and equidimensional SSEs occur in decoupled areas or far from stability transitions. Examples are those in the Bungo Channel separating the Japanese islands of Shikoku and Kyushu to the west of the coupled region encompassing the 1946 rupture zone (Figure 4.40(a)) (Sagiya, 2004). In the Guerrero segment of the Mexican subduction zone *Mw* ~7.5 slow slip events with average slips of ~10 cm occur with 6–7 months duration at intervals of about 5 years (Kostoglodov et al., 2003; Radiguet et al., 2012). These events are silent and occur at depths of 35–50 km, far from the shallow coupled zone that is just offshore. A long–term slow slip event was observed in the Shumagin region of Alaska (Beavan et al., 1984), a "seismic gap" that is now known to be decoupled (Scholz and Campos, 2012). Shallow (10–20 km) slip events of several weeks duration occur on the periphery of the coupled part of the Sagami Trough off the Boso Peninsula of SE Honshu (Figure 4.40(a)) (Ozawa et al., 2003; Sagiya, 2004). These events are not accompanied by tremor, but like the shallow SSEs in Hikurangi, are accompanied by microearthquake swarms. For other examples, see Schwartz and Rokosky (2007).

Nonvolcanic tremor is also observed below the San Andreas Fault in central California (Nadeau and Dolenc, 2005). This occurs in migrating clusters that suggest ETS episodes (Shelly, 2010) but which have not produced detectible geodetic signals. The LFEs associated with this activity occur at a depth of 25 km, which indicates a gap of 10 km from the seismogenic zone. Short–term creep episodes are recorded with surface measurements in the creeping section of the San Andreas Fault, but they may be due to near–surface effects (Schulz et al., 1983; Schulz et al., 1982). However, a shallow SSE of larger duration and size was observed in the northern part of the creeping section (Linde et al., 1996).

Slow slip events beneath an Antarctic ice stream

Motion of the Whillans Ice Stream, West Antarctica, is dominated by twice–daily tidally modulated slow slip events (Pratt et al., 2014; Walter et al., 2011; Wiens et al., 2008). These events, which occur just before the maximum low tide and just after the maximum high tide, produce ~0.5 m slip of a 100 x 100 km area over a duration of 20–30 min. They have equivalent moment magnitudes of ~7 but produce radiation in the seismic band that is equivalent to only M_S 3.6–4.2.

An image of the Whillans Ice Stream is shown in Figure 4.41(a), where the seismic coupling coefficient, the ratio of slip that occurs in the SSEs to total slip, is contoured (Winberry et al., 2014). There are two maxima in coupling: the central "sticky spot" (CSS), and the northern "sticky spot" (NSS). These are areas that slip less rapidly and thus accumulate more stress during interevent periods – they are equivalent to aperities in earthquake parlance. They correspond to elongated elevations in the basement around which subglacial water flow is diverted, and may correspond to regions of higher effective normal stress due to bending over the basal hill and more consolidated subglacial till (Luthra et al., 2016). Also shown are three regions near the grounding line, 1–3, which produce high–frequency (30–100s) radiation pulses during SSEs.

The high tide SSEs initiate from the locked CSS asperity and the low–tide ones from a slowly sliding asperity at point 1 near the grounding line (Figure 4.41(b)). Rupture velocities near the initiation points are ~1000 ms^{-1} for the high–tide events and ~1,500 ms^{-1} for the low–tide events (the S–wave velocity of ice is 1900 ms^{-1}). Both decelerate rapidly away from the asperities to an average rupture velocity of ~150 ms^{-1}. Radiation is produced from late subevents at the NSS asperity and at points 2 and 3, both close to the grounding line. The GPS receiver nearest the initiation point CSS shows a nucleation phase, and the SSEs are accompanied by increased microearthquake activity and tremor (Winberry et al., 2013), the latter of which appears to be due to swarms of repeating microearthquakes (Lipovsky and Dunham, 2016).

The slip in each event increases with the time elapsed from the previous event, as does the average slip velocity (Walter et al., 2015). At first glance this appears to obey the classic slip–predictable recurrence model (Figure 5.17). However, that is a one–dimensional model that applies to the behavior of a single asperity. In this case there are several different types of SSEs: short recurrence–time low–tide events that nucleate at point 1, long recurrence–time high–tide events that nucleate at CSS, intermediate recurrence–time neap–tide events, and very long recurrence–time events which occur when a low–tide event is skipped and occur after the peak in the next high tide. There is not a continuum: each of these event types has a characteristic recurrence time (Winberry et al., 2014). The effect of Ross Sea tides is to apply a pressure at the grounding line that enhances or reduces loading on the ice sheet upstream. Low tides increase the loading rate and high tides decrease it. As a result, each type of event reaches the healing curve at a different time and stress level: the faster–loading low–tide events reach the critical point quicker than the low stressing rate high–tide events, and hence produce smaller slips (Winberry et al., 2009). The intrinsic recurrence time of the SSEs is about 12 hours, observed during neap tides. This is near enough to the diurnal tidal period so that tidal modulation occurs. High tides retard the SSEs to about 15 hours, whereas low tides advance them to about 9 hours. It would be incorrect to suggest that the tides trigger the SSEs, in the sense of precipitating an event that otherwise would not occur within that time–frame. The net effect is an apparent relation between slip and recurrence time.

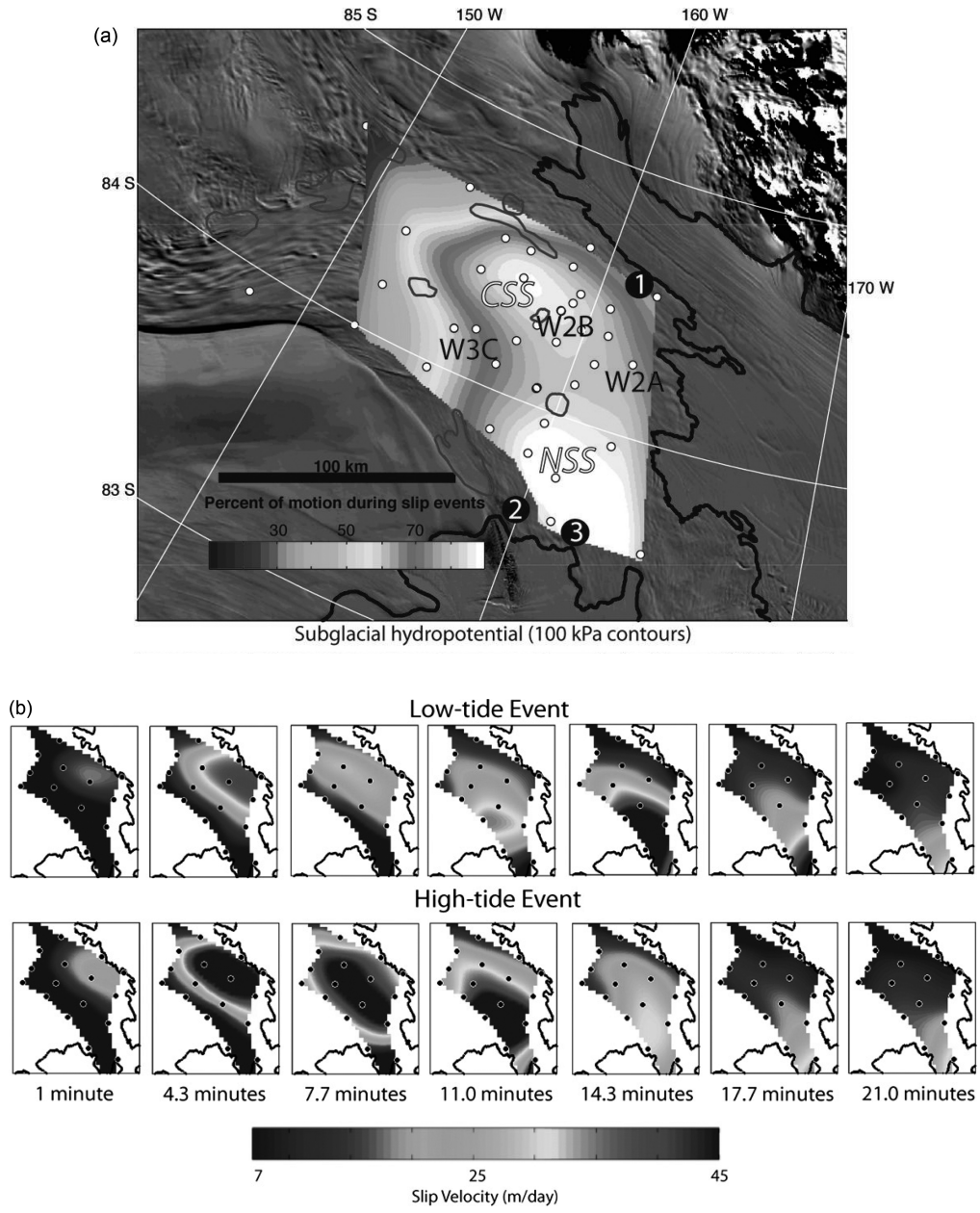

Fig. 4.41. (a) Image of the Whillans Ice Sheet, showing the contoured seismic coupling coefficient (ratio of slip that occurs in SSEs to total slip). CSS and NSS are two areas of high coupling, termed "sticky spots." Points 1–3 are other regions that radiated high-frequency pulses during SSEs. Heavy curve is the grounding line. Overall motion of the ice sheet is from upper left to lower right. (From Winberry et al., 2014.) (b). Slip evolutions of high-tide and low-tide events. (From Winberry et al., 2011.) (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

The more common behavior of ice streams is stable sliding that is modulated by diurnal or fortnightly ocean tides (Anandakrishnan et al., 2003; Gudmundsson, 2006). In the Whillans case, variations in basal friction lead to the retardation of the stable sliding velocity of the ice sheet by a few strong asperities. When these asperities fail dynamically, the stable sliding velocities in the surrounding regions accelerate to release the pent-up slip. The subcritical rupture velocities that occur within those stable sliding areas, ~ 150 cms^{-1}, are proportional to the stress drop in the event (Walter et al., 2011) and are much faster than the 10 km/day of the oscillatory SSEs beneath subduction zones, so this seems to be a different class of process, perhaps more akin to tsunami earthquakes (Section 6.3.2). However, it does not seem that exotic frictional behavior is required. In lab friction experiments with inhomogeneous loading, stick–slip events are observed that likewise propagate into lower-stress regions with subcritical rupture velocities proportional to the local stress drop (Walter et al., 2015).

The topic of earthquake phenomena related to glaciers is a large and fascinating one that has received considerable attention in recent years. For those interested in pursuing it more fully, see the review of Podolskiy and Walter (2016).

4.6.3 Mechanics of slow slip events

The relation between duration and moment for a large variety of slow and regular earthquake phenomena are shown in Figure 4.42. The scaling for regular earthquakes, in which moment scales with the cube of duration, is the same as shown in Figure 4.9 below the crossover. Also

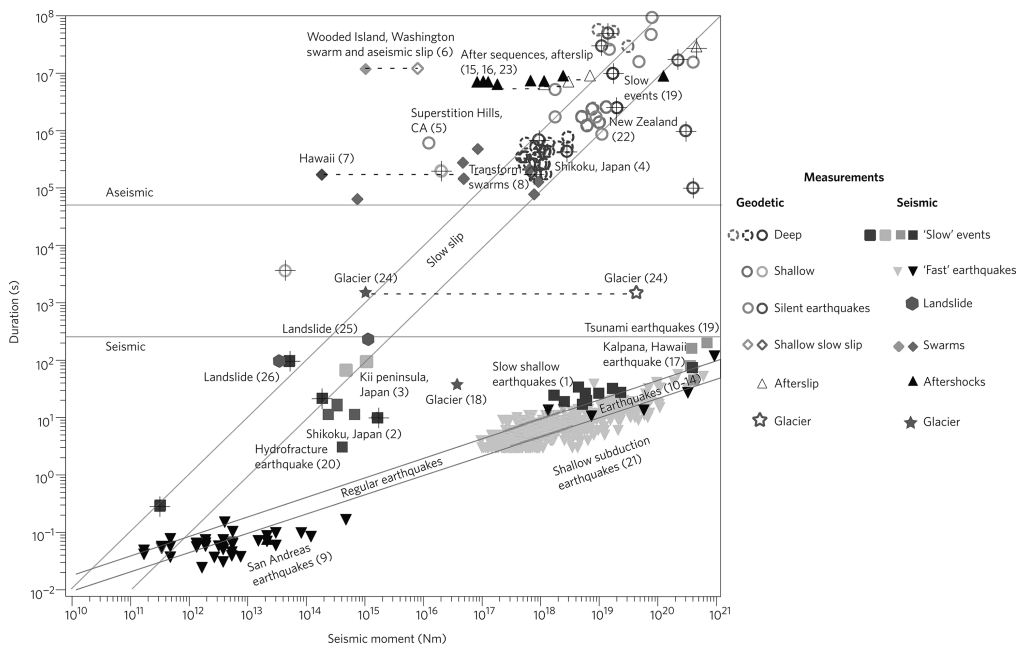

Fig. 4.42. Seismic moment versus duration for a variety of fault slip phenomena. This is a plot from Ide et al. (2007), which has been augmented with data from Peng and Gomberg 2010. Measurements from Ide et al. have plus signs over the symbols. (From Peng and Gomberg, 2010.)

shown is a proposed scaling law for slow slip events in which moment scales linearly with duration (Ide et al., 2007). In Figure 4.42 a number of new data points for diverse phenomena have been added by Peng and Gomberg (2010) to the original plot by Ide et al. (2007). This introduces considerable scatter, which led Peng and Gomberg to suggest that a continuum of slip modes exists with no clear separation between slow and regular earthquakes.

It is certainly hazardous to include diverse phenomena such as LFEs and SSE in the same scaling law, particularly when they are separated by the wide observation gap between the seismic and the aseismic (geodetic) bands in Figure 4.42. However, one can be more successful if one limits oneself to seeking scaling relations among phenomena that are clearly of the same type. For example, Aguiar et al. (2009) obtained a good linear relationship between moment and duration for SSEs in Cascadia: $M_o = 1.4 \pm 0.1x10^{13}t$($M_o$ in N–m and t in sec). Plugging in μ = 50GPa, W = 80 km, and growth rate L/t = 10 to 20 km/day, depending on whether the SSE growth is unilateral or bilateral, this scaling relation implies a scale–independent slip of 3.6–1.8 cm, respectively. This is in reasonable agreement with inversions from GPS data, such as shown in Figure 4.37. This corresponds to a stress drop in the range 30–15 kPa. This scaling is thus the same as that of high aspect ratio earthquakes (Regime 3 in Table 4.2) with the scaling parameter, stress drop, being about two orders of magnitude smaller (c.f. Brodsky and Mori, 2007).

As mentioned earlier vis–à–vis the sensitivity of tremor to tidal triggering, it is generally believed that slow earthquakes occur in regions of near lithostatic pore pressure (Peng and Gomberg, 2010; Saffer and Wallace, 2015). Evidence for this comes from tomographic imaging of such regions that show high ratios of P to S–wave seismic velocities, high Poisson's ratio, or ultralow shear velocity layers (Audet et al., 2010; Audet et al., 2009; Bassett et al., 2014; Kodaira et al., 2004; Matsubara et al., 2009). However, the variety of depth ranges and thermal histories of zones in which SSEs occur indicates that there is no single metamorphic reaction that is the source of the high fluid pressures (Peacock, 2009; Saffer and Wallace, 2015). This does not mean that there are not mechanisms for overpressures specific to particular locales. Hyndman et al. (2015), for example, point out in the case of Cascadia the close association of the ETS channel with the forearc mantle corner, and suggest mechanisms by which that could focus high fluid pressures in the ETS channel. However, at such depths, the use of the linear effective stress law is questionable, because as $A_r/A \rightarrow 1$, Equation 2.16 indicates that $\alpha \rightarrow 0$ in Equation (1.46) (Beeler et al., 2016).

The periodic SSEs below the seismogenic depths of subductions zones are reminiscent of the oscillatory stable sliding that occurs close to the stability transition, as was discussed in Section 2.3.3 (see Figure 2.29(b)). The periodic SSEs have been modeled in 2D by Liu and Rice (2007) and Rubin (2008), who assumed RS friction with a velocity–weakening zone of down–dip width W between an upper seismogenic locked zone and a lower zone of velocity strengthening. At some greater depth a steady slip at the plate rate is imposed. The phase response diagram is shown in Figure 4.43. The control parameter is W/L_c, where L_c is the critical patch length for stable sliding, given in Equation (2.36). This is similar to Figure 2.28, except that the effective normal stress $\overline{\sigma}_n$ is multiplied by W. The conditionally stable regime contains three fields. For a fixed value of W, at low effective stresses steady stable sliding prevails. At higher effective stresses this changes to a region of periodic oscillations, which in turn gives way at yet higher stresses to a region of period doubling and chaotic motion just adjacent to the unstable field. Within the region of periodic oscillations both the period and the amplitude of the oscillations increase with effective stress. Using this scaling, Liu and Rice estimated the effective stress for Cascadia to be 2–3 MPa. They explained this low value by assuming the existence of high pore pressures, justifying this with the same reasons

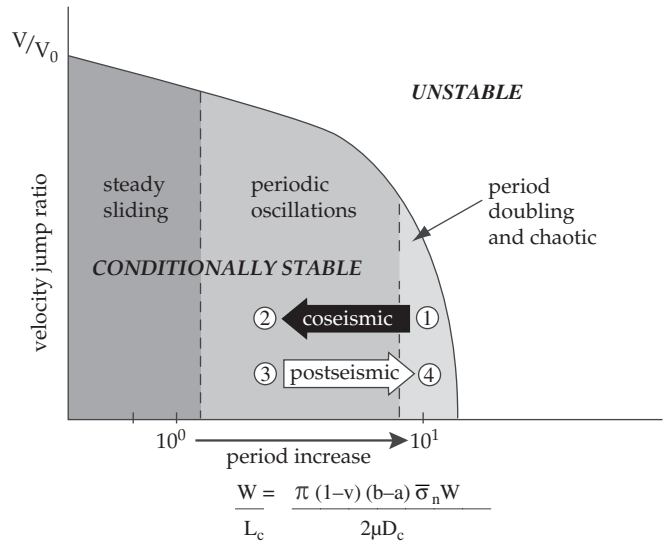

Fig. 4.43. Phase response diagram for the model of Liu and Rice (2007). (Modified from Liu and Rice, 2007.)

mentioned in the previous paragraph. Interestingly, a very similar phase response diagram was obtained in a block–slider model of Gu et al. (1984) with RS friction and two state variables. Many of those features have been corroborated by laboratory friction measurements (Leeman et al., 2016).

This mechanism is not obviously applicable to SSEs that occur far from the locked zone or stability boundary, such as discussed in the previous section. This is because these cases are not restricted to a narrow channel of width W as in the Liu and Rice model. As can be seen from Figure 4.43, with large values of W, smooth stable sliding is expected. One class of models that may apply to such cases contain mechanisms that quench the instability, such as a change from velocity weakening to velocity strengthening with increasing velocity (Shibazaki and Iio, 2003) or dilatancy hardening that clamps the fault at moderate sliding velocities (Rubin, 2008; Segall and Rice, 1995). A type of slow slip event has been observed in subduction zone clay gouge deformed in the laboratory at plate tectonic slip rates (Ikari et al., 2015). Those cases involve a slowing down of the slip rate that results in an increase of shear stress that is subsequently released by more rapid slip. There have not been any reports of such a period of reduced slip preceding SSEs in nature. Other models invoke pore pressure buildup followed by release during slip (Audet et al., 2009; Hyndman et al., 2015) in the manner of the pump–seal mechanism of Sibson et al. (1988) and the inhibition of nucleation due to a rate–dependent slip–weakening mechanism (Ikari et al., 2013). Because the periodic SSEs retain their pulse shape during propagation they resemble solitons, which are solutions to certain nonlinear wave equations such as the sine–Gordon equation (Dodd et al., 1982). There have been several attempts to follow this line of investigation (Bykov, 2014; Gershenzon et al., 2016).

The region marked "period doubling and chaotic motion" in Figure 4.43 has not been widely explored. However, there are observations of LFEs beneath the San Andreas Fault in the vicinity of Parkfield that appear to exhibit this behavior. Two nearby families of bursts of LFEs were

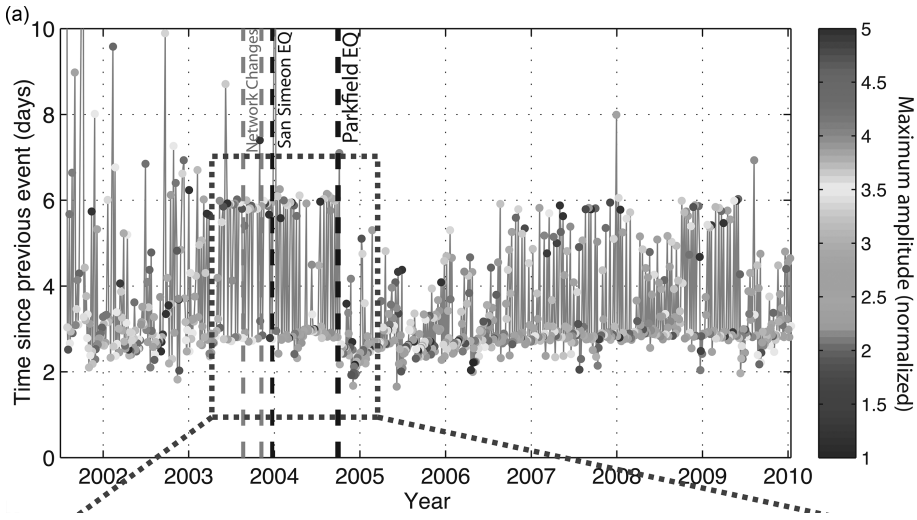

Fig. 4.44. Recurrence interval versus time for LFE family bursts south of Parkfield. Gray lines connect consecutive events. Color scale indicates maximum amplitude of ground velocity for each burst. (From Shelly, 2010.) (A black and white version of this figure appears in some formats. For the color version, please refer to the plate section.)

observed with recurrence periods of 3 days and 6 days, as shown Figure 4.44. This seems to indicate period–doubling behavior. The nearby Parkfield earthquake of 2004 had a profound effect on this behavior. The period–doubling ceased and the behavior switched to simple periodicity with the primary period shorter than previously. This would be the predicted effect if the Parkfield earthquake reduced the effective normal stress in the vicinity of the LFEs, such as from point 1 to 2 in Figure 4.43. It then appears that the effective normal stress subsequently recovered, say by poroelastic recovery, from point 3 to 4. The primary oscillation period gradually increased until it was approximately the value before the earthquake, whereupon the period doubling resumed.

We observe that there are three types of background radiation that accompany SSEs: tremor that accompanies deep SSEs, microearthquakes that accompany shallow SSEs, and silent SSEs. It is possible that the last category radiates some noise that is below the resolution of present instrumentation. These different modes all represent the rubbing noise of the SSE. The moment release rate of each mode is insignificant with respect to that of the SSE itself: they are secondary phenomena. These modes of behavior are generally modeled as heterogeneous faults in which small unstable patches are distributed on an overall stably sliding fault. For faults within the brittle regime this results in microearthquakes, and within the ductile regime LFEs. In the brittle regime, *repeating earthquakes* occur that result from rapid loading of unstable patches by a steady slip accumulation surrounding them (Nadeau et al., 1995; Nadeau and McEvilly, 1999). In the ductile regime, the equivalent process occurs, resulting in repeating LFEs (Frank et al., 2015). There are several variations of models of repeating earthquakes (Beeler et al., 2001; Chen and Lapusta, 2009; Rubinstein et al., 2012). A similar model was developed for LFEs and tremor, in which a viscous damping term was included in the source term to increase the rupture time, resulting in LFEs instead of microearthquakes (Ando et al., 2010).