1 Introduction

Fracture (or failure) phenomena are observed over a very broad range of size scales, from atomistic-scale to microscopic-scale to macroscopic-scale fractures. A shear failure (or rupture) of laboratory-scale, whether the shear failure of intact rock or frictional slip failure on a precut rock interface, would be of the order of 10^{-3} to 1 m. In contrast, shear rupture phenomena occurring in the Earth's interior, including microearthquakes and huge earthquakes, encompass a much broader range of size scales from 10^{-1} to 10^{6} m. Rupture phenomena over such a broad scale range covering both laboratory-scale and field-scale are encompassed by continuum mechanics. This book deals with shear failures (or ruptures) of a scale range of laboratory-scale and field-scale within the framework of continuum mechanics.

It has been established that the source of shallow focus earthquakes at crustal depths is shear rupture instability along a fault embedded in the Earth's crust, which is composed of rock. At the same time, laboratory experiments have demonstrated that a rock specimen fails (or ruptures) in shear mode under combined compressive stress environments, and that the shear failure (or rupture) of rock is governed by constitutive law. These facts physically mean that earthquake rupture processes are governed by the constitutive law. This enables a deeper understanding of the process of earthquake generation in terms of the underlying physics, if the constitutive law for earthquake ruptures is properly formulated by taking into account the real situation of seismogenic fault zone properties such as fault heterogeneities.

In light of this, we have to correctly recognize the real situation of seismogenic environments and fault zone properties, in order to strictly formulate the constitutive law for earthquake ruptures, and to quantitatively account, in a unified and consistent manner, for the entire process of scale-dependent earthquake rupture in terms of the underlying physics. Since correct recognition of seismogenic fault zone properties is an indispensable prerequisite for strict formulation of the constitutive law for earthquake ruptures, this introductory chapter is mainly devoted to a description of seismogenic fault zone properties.

Since an unstable, dynamic rupture can occur only in the brittle through semi-brittle layer of the Earth's interior, the occurrence of earthquake ruptures is generally restricted to shallow crustal depths, called the *seismogenic layer*, except in plate-boundary subduction zones where deeper earthquakes can occur. The brittle through semi-brittle layer away from plate-boundary subduction zones is usually limited to crustal depths shallower than 20 km, and temperatures and lithostatic pressures in the layer range from the Earth's surface temperature to roughly 500–600 °C, and from atmospheric pressure to about 500 MPa, respectively. The seismogenic layer and individual faults embedded therein are inherently heterogeneous; in other words, heterogeneity is a key property of the seismogenic layer and preexisting faults therein. This fact cannot be denied, even if a part of the preexisting fault zone in a superficial layer of the Earth's crust is narrow and remarkably planar, as some geologists have noted. It is not possible to look at the whole structure of a fault zone from an isolated part of the fault zone viewed in the superficial layer. A view from an isolated part cannot be universalized.

Indeed, seismic observations have demonstrated that a major or great earthquake at shallow crustal depths never occurs alone, but is accompanied by aftershocks, and often preceded by seismic activity (small to moderate earthquakes) enhanced in a relatively wide region surrounding the source fault during the process leading up to the event (see Chapter 7). This is a reflection of the fact that the seismogenic layer and individual faults therein are heterogeneous. If the seismogenic layer and individual faults were homogeneous, then not only foreshocks but also aftershocks could never occur. Specifically, the occurrence of a mainshock rupture on a fault is accompanied by a rapid stress drop and slip displacement on the finite fault, and an adequate amount of the elastic strain energy accumulated in the medium surrounding the fault is dissipated during the rupture. Consequently, the mainshock rupture arrests within a finite region (the fault area), and the rupture arrest results in the re-distribution of local stresses on and around the fault area, leading to aftershock activities in and around the source region. Note that the occurrence of such a sequential process in a finite region is due to heterogeneities of the seismogenic layer and preexisting faults therein.

Seismological observations and analyses (e.g., Kanamori and Stewart, 1976; Aki, 1979, 1984; Beroza and Mikumo, 1996; Bouchon, 1997; Zhang *et al.*, 2003; Yamanaka and Kikuchi, 2004) have revealed that individual faults embedded in the seismogenic layer are heterogeneous, and contain what are called "asperities" (e.g., Lay *et al.*, 1982) or "barriers" (e.g., Aki, 1979) in the field of earthquake seismology. The presence of "asperities" or "barriers" on a fault is a clear manifestation of the fact that a real fault contains strong local areas highly resistant to rupture growth, with the rest of the fault having low (or little) resistance to rupture growth. These seismological observations are consistent with geological observations of structural heterogeneity and geometric irregularity for real faults, as described below.

In general, real faults embedded in the seismogenic crust are nonplanar, being segmented and bifurcated (e.g., Sibson *et al.*, 1986; Wesnousky, 1988). In addition, individual surfaces of fault segments exhibit geometric irregularity with band-limited self-similarity (e.g., Aviles *et al.*, 1987; Okubo and Aki, 1987), and there are gouge layers in between fault surfaces for mature faults (Sibson, 1977). These geometric irregularities and structural heterogeneities for real faults (or fault zones) play a prominent role in causing heterogeneous distributions of not only stresses acting on individual faults but also fault strength and resistance to rupture growth, because these physical quantities are structure-sensitive. The resistance to rupture growth has a specific physical meaning in the framework of fracture mechanics. For the definition of the resistance to rupture growth, see Section 2.2.

Let us consider a specific case where a fault consists of a number of discrete segments which form an echelon array with individual segments nearly parallel to the general trend of the fault (Segall and Pollard, 1980). The stepover zones in such an echelon array possibly have the highest strength, equal to the fracture strength of intact rock, and the resistance to rupture growth may be high enough to impede or arrest the propagation of ruptures at these sites. In addition, nonplanar fault segment surfaces exhibiting geometric irregularity with band-limited self-similarity contain various wavelength components. When such geometrically irregular fault segment surfaces, pressed under a compressive normal stress, are sheared in the brittle regime, the prime cause of frictional resistance is the shearing strength of interlocking asperities (Byerlee, 1967). In this case, the local strength at the sites of interlocking asperities is strong enough to equal the shear fracture strength of initially intact rock in the brittle regime.

Thus, the sites of the zones of segment stepover and/or interlocking asperities are potential candidates for strong areas of high resistance to rupture growth. These sites may act as "barriers" or "asperities." When such strong areas act as "barriers," stress will build up and elastic strain energy will accumulate in the elastic medium surrounding these sites until they break. When a large rupture breaks through these sites and links them together, they will be regarded as "asperities."

Local stress drops at "asperities" on real seismic faults obtain values as high as 50 to 100 MPa (e.g., Papageorgiou and Aki, 1983a and 1983b; Ellsworth and Beroza, 1995; Bouchon, 1997), which is high enough to equal the breakdown stress drop of intact rock tested under seismogenic environmental conditions simulated in the laboratory (see Section 3.4). This strongly suggests that the earthquake rupture process at shallow crustal depths is not a simple process of frictional slip failure on a uniformly precut weak fault, but a more complex process including the fracture of initially intact rock at some local areas on an inhomogeneous fault.

Such local strong areas on a fault, called "asperities" or "barriers," are required for an adequate amount of elastic strain energy to accumulate in the elastic medium surrounding the fault, owing to tectonic loading. Since the elastic strain energy accumulated provides the driving force to bring about an earthquake or to radiate seismic waves, local strong areas highly resistant to rupture growth on a fault play a much more important role in generating a large earthquake or in radiating strong motion seismic waves than does the rest of the fault having low (or little) resistance to rupture growth. Thus, there is no doubt that real faults in the Earth's crust are inherently heterogeneous.

In light of this, the constitutive law for earthquake ruptures must be formulated as a unifying law that governs not only frictional slip failure at precut weak interface (or frictional contact) areas on faults but also the shear fracture of intact rock at some local strong areas on the faults (Chapters 3 and 4).

In addition, rupture phenomena, including earthquakes, are inherently scale-dependent. Indeed, some of the physical quantities inherent in shear rupture exhibit scale-dependence (Chapters 5 and 6). It is therefore essential to formulate the governing (or constitutive) law in such a way that the scaling property inherent in the rupture breakdown is incorporated into the law; otherwise, scale-dependent physical quantities inherent in the rupture over a broad scale range cannot be treated consistently and quantitatively in a unified manner in terms of a single constitutive law.

As described above, the constitutive law governing the behavior of earthquake ruptures provides the basis of earthquake physics, and the governing law plays a fundamental role in accounting quantitatively, in a unified and consistent manner, for the entire process of a scale-dependent earthquake rupture, from its nucleation to its dynamic propagation to its arrest, on a heterogeneous fault. Therefore, it is critically important to strictly formulate the constitutive law for real earthquake ruptures, based on positive facts, from comprehensive viewpoints. Without a rational formulation of the law governing real earthquake ruptures, the physics of earthquakes cannot be a quantitative science in the true sense.

However, the resolution of seismological data observed in the field is not high enough not only to strictly formulate the constitutive law, but also to fully elucidate the physical nature of a scale-dependent earthquake rupture generation process from its nucleation to the subsequent dynamic propagation on a heterogeneous fault. This is because it is not possible to preliminarily deploy a series of high-resolution instruments for measuring local shear stresses (or strains) and local slip displacements along the fault on which a pending earthquake is expected to occur at a crustal depth.

In contrast, in laboratory experiments, the experimental method can be properly devised for the purpose intended, and high temporal and spatial resolution measurements can be made at a series of locations along the preexisting fault on which a shear rupture occurs. Therefore, high-resolution laboratory experiments properly devised for the purpose intended enable us not only to strictly formulate the constitutive law for shear rupture but also to fully elucidate the physical nature of a scale-dependent shear rupture process from its nucleation to the subsequent dynamic rupture on an inhomogeneous fault. In particular, in order to strictly formulate the constitutive law as a unifying law that governs not only frictional slip failure at precut interface areas on a fault but also the shear fracture of intact rock at some local strong areas on the fault, we need detailed data obtained in high-resolution laboratory experiments on the shear failure of intact rock and frictional slip rupture on a precut rock interface. I have devoted myself to conducting such leading-edge research through high-resolution laboratory experiments properly devised for the purpose intended, and contributed to the derivation of underlying physical laws, such as a unifying constitutive law and a constitutive scaling law, and to the elucidation of the physical nature of the scale-dependent shear rupture generation process, to achieve a deeper understanding of the physical process from earthquake nucleation to its dynamic propagation in terms of the underlying physical laws.

As mentioned above, real faults embedded in the Earth's crust are inherently inhomogeneous, and the earthquake rupture process at shallow crustal depths is not a simple process of frictional slip failure on a uniformly precut weak fault, but a more complex process, including the fracture of initially intact rock at some local strong areas on an inhomogeneous fault. In this book, therefore, the constitutive law is formulated as a unifying law which governs not only frictional slip failure at precut interface areas on a fault but also shear fracture at some local strong areas on the fault, and into which the scaling property inherent in shear-rupture breakdown is incorporated, based on positive facts elucidated in high-resolution laboratory experiments properly devised for the purpose intended (Chapters 3 and 4). This enables one to quantitatively account for the physical behavior of scale-dependent shear ruptures, including earthquake rupture nucleation and strong motion during the dynamic rupture propagation, in a unified and consistent manner, in terms of a single constitutive law (Chapters 5 and 6), and consequently to enhance the physics of earthquakes to a more complete, quantitative science in the true sense. This deductive approach based on the results of high-resolution laboratory experiments is the prominent feature of this book, and sets this book apart from others in the fields of rock failure physics and earthquake physics.

2 Fundamentals of rock failure physics

2.1 Mechanical properties and constitutive relations

2.1.1 Elastic deformation

When an external stress is applied to a real material, the material necessarily deforms, and this may eventually lead to fracture; in other words, fracture is commonly preceded by deformation. If there is a unique relation between the applied stress and the resultant deformation (or strain), the material is referred to as *perfectly elastic*. Perfect elasticity implies that there is no energy loss during the loading and unloading processes, which means that the energy stored in the specimen during the loading process is completely released during the unloading process. In other words, the elastic deformation is instantaneous and is completely recoverable when the applied stress is removed. When there is a linear relation between the stress applied to a material and the resultant deformation (or strain), the material is called *linearly elastic*, which is usually valid for small deformation.

Geomechanical phenomena in geological and tectonic settings in the Earth are described quantitatively in terms of basic equations, which include mechanical constitutive laws governing deformation or rupture. The relation between an externally applied force and the mechanical response of a material to the applied force is called a (*mechanical*) *constitutive relation*.

A typical, well-known example of such constitutive relations is Hooke's law; this law holds for a material having the property of the aforementioned linear elasticity. If it is assumed that the matter of a linearly elastic body is homogeneous and isotropic, Hooke's law is simply written as

$$\sigma_{ij} = \lambda\delta_{ij}\varepsilon_{kk} + 2\mu\varepsilon_{ij}, \tag{2.1}$$

where σ_{ij} is the stress tensor, ε_{ij} is the strain tensor, δ_{ij} is the unit tensor called Kronecker's delta ($\delta_{ij} = 0$ for $i \neq j$, and $\delta_{ij} = 1$ for $i = j$), and λ and μ are Lame's constants. Lame's constants prescribe the elastic property of a material.

There are two independent elastic constants for a homogeneous and isotropic body with the property of linear elasticity. In Eq. (2.1), Lame's constants have been used as the two independent constants. However, there are other elastic constants defined as those best suited to specific uses. The other elastic constants are Young's modulus E, Poisson's ratio ν, the modulus of rigidity G, and the bulk modulus of elasticity K. These elastic constants are directly related to Lame's constants as follows:

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}, \tag{2.2}$$

$$\nu = \frac{\lambda}{2(\lambda + \mu)}, \tag{2.3}$$

$$G = \mu, \tag{2.4}$$

$$K = \lambda + \frac{2}{3}\mu. \tag{2.5}$$

The elastic constants represent the elastic property of a material, and the material property depends on ambient conditions. Accordingly, the elastic constants also depend on ambient conditions.

2.1.2 Ductile deformation

As the stress (or load) applied to a real material increases, the material may fail (or fracture) within the elastic limit; otherwise, it continues to deform permanently beyond the elastic limit, and eventually it will fail (or fracture) due to the breakdown of inter-atomic bonds. The deformation beyond the elastic limit is not recoverable when the applied stress is removed, and the residual deformation at zero stress is referred to as *permanent deformation*.

The term *ductile deformation* is used when permanent deformation takes place without losing its ability to resist the applied load. The transition from elastic to ductile deformation occurs at the yield stress point. A small amount of ductile deformation is widely observed preceding the shear failure of rock tested under confining pressures, even in the brittle regime. This is because rocks are inherently inhomogeneous. Rocks are made up of various types of mineral grains, and contain a variety of mechanical flaws and stress concentration sources, such as microcracks, pores, and grain boundaries. In compressive loading of one such rock, elastic deformation is usually limited to the first 40–50% of the peak strength. Above the 40–50% level of the peak strength, the number of microcracks increases progressively as the rock is loaded. At higher loads, time-dependent crack growth and the interaction and coalescence of neighboring cracks become progressively more important, forming a thin, planer zone of higher crack density, which eventually results in the macroscopic fracture surfaces in the post-peak region, where the shear strength degrades with ongoing slip (slip-weakening). Thus, when rock in the brittle regime fails by shear mode under confining pressure, non-elastic deformation necessarily precedes the macroscopic shear fracture. The mechanism of this non-elastic (or ductile) deformation in the brittle regime is called *cataclasis*, which involves micro-fracturing and friction acting on microcrack surfaces and fragment particle interfaces developed in rock during loading.

In addition, individual crystalline solids of mineral grains in rocks contain various types of crystallographic defects such as point defects, dislocations, and twin boundaries in their crystal lattices. These crystallographic defects cause plastic deformation of the minerals

Table 2.1 Constitutive law parameters for high-temperature plastic creep flow

Rock name	A (MPa^{-n}s^{-1})	n	Q (kJ mol^{-1})	References
Heavitree quartzite (vacuum-heated, dry samples)	4×10^{-6}	4.0	300	Kronenberg and Tullis (1984)
Heavitree quartzite (α-quartz, 0.4% water added samples)	2.2×10^{-6}	2.7	120	Kronenberg and Tullis (1984)
Heavitree quartzite (α-quartz, dry samples)	3.2×10^{-5}	1.9	123	Hansen and Carter (1982)
Simpson quartzite (α-quartz, 0.4% water added samples)	2.0×10^{-2}	1.8	167	Hansen and Carter (1982)
Mt. Burnett dunite (dry samples)	1.3×10^{3}	3.3	465	Carter and Ave'Lallemant (1970)
Mt. Burnett dunite (wet samples)	0.1	2.1	226	Carter and Ave'Lallemant (1970)
Anita Bay dunite (dry samples)	3.2×10^{4}	3.6	535	Chopra and Paterson (1981)
Anita Bay dunite (wet samples)	1×10^{4}	3.4	444	Chopra and Paterson (1981)

contained in rocks at high temperatures. Dissolution creep (pressure solution) also occurs under stress (or pressure) in the presence of water. Accordingly, the mechanisms of ductile deformation (or flow) of rock are the cataclasis, crystal plasticity involving dislocation creep and diffusion creep, and dissolution creep.

Rocks exhibiting brittle behavior at room temperature deform plastically at high temperatures under confining pressures. Another typical example of mechanical constitutive relations may be the constitutive law for plastic flow, in particular, steady-state creep flow at high temperatures. In the case of plastic flow, the physical quantity that is uniquely related to stress is not strain but strain rate. It has been established by laboratory experiments that the steady-state creep flow of rock at a high temperature under a confining pressure obeys the following law (e.g., Carter, 1976; Kirby, 1983):

$$\dot{\varepsilon} = A f(\sigma_{\text{diff}}) \exp\left(-\frac{Q}{RT}\right), \tag{2.6}$$

where $\dot{\varepsilon}$ is the strain rate, $f(\sigma_{\text{diff}})$ is a function of the differential flow stress σ_{diff} defined as the difference between flow stress σ and confining pressure P_c, T is absolute temperature, A is a constant, Q is the activation energy, and R is the gas constant. The specific functional form of $f(\sigma_{\text{diff}})$ is commonly represented by a power law; that is,

$$f(\sigma_{\text{diff}}) = \sigma_{\text{diff}}^{n}, \tag{2.7}$$

where n is an exponent depending on the material and ambient conditions.

Equation (2.6) is a constitutive equation that prescribes the relation between flow stress σ_{diff} and strain rate $\dot{\varepsilon}$ at temperature T. The creep strain rate depends exponentially on the temperature, and hence the strain rate is very sensitive to temperature. For reference, the plastic flow law parameters A, n, and Q determined experimentally for representative rocks are listed in Table 2.1.

Cambridge University Press & Assessment
978-1-107-03006-0 — The Physics of Rock Failure and Earthquakes
Mitiyasu Ohnaka
Excerpt
More Information

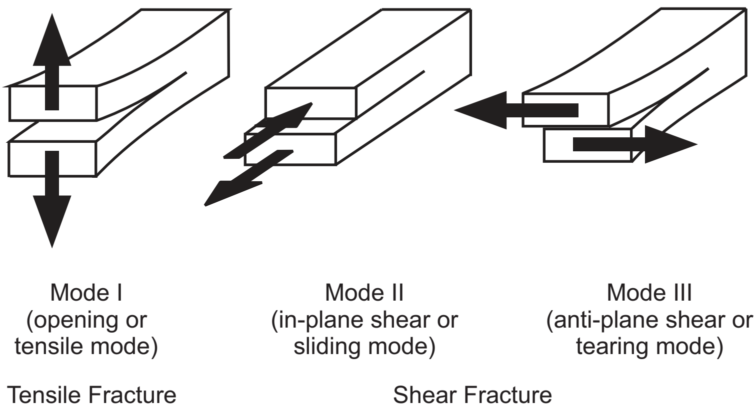

Fig. 2.1 The three fundamental modes of fracture.

2.1.3 Fracture

If increasing stress is applied to a real material, the deformation eventually concentrates in a narrowly localized zone in the material, the fracture finally occurs, and new crack surfaces are created as a result of the breakdown of inter-atomic bonds in the localized zone. Therefore, fracture is characterized by the fact that the strength deteriorates with ongoing relative displacement between the fracturing surfaces during the breakdown process. When the relative displacement is perpendicular to the fracture plane to open the crack, it is called *tensile fracture* (mode I). When the relative displacement is parallel to the fracture plane, it is called *shear fracture*. Shear fracture consists of two fundamental modes: one of which is called *sliding mode* or *in-plane shear mode* (mode II), in which the relative displacement (or slip) is perpendicular to the crack edge, and the other is called *tearing mode* or *anti-plane shear mode* (mode III), in which the relative displacement (or slip) is parallel to the crack edge. These three different fracture modes are illustrated in Figure 2.1.

A fracture that occurs within the elastic limit is called a *brittle fracture*, though a fracture including a small amount of permanent deformation may also be categorized as a brittle fracture type for practical purposes. In contrast, a fracture that occurs after an extensive amount of permanent deformation is referred to as a *ductile fracture*. Once fracture occurs in a real material, the stability or instability of the fracture process is governed by how progressively the strength during the fracture process degrades with ongoing relative displacement. Thus, the constitutive relation for fracture is determined by the transient response of the traction on the fracturing surfaces to the relative displacement. The constitutive relation for rock fracture and for frictional slip failure is an important, fundamental theme of this book, and hence it is described in detail in Chapter 3. In this subsection, we evaluate the breaking strength of atomic bonds that hold atoms together, and the constitutive relation for the bond breaking process.

To evaluate the cohesive strength (or atomic bond strength) σ_c of a material with ideal crystal structure that does not contain any flaws or defects (Orowan, 1948; Gilman, 1959), let r_0 be the equilibrium spacing between atomic planes under no applied stress (see Figure 2.2).

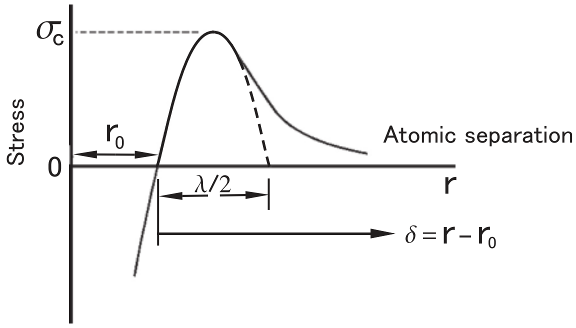

Fig. 2.2 The tensile stress σ required for separating atomic planes, as a function of distance r. The equilibrium spacing between atomic planes under no applied stress is denoted by r_0. Fracture occurs when $\sigma = \sigma_c$.

The tensile stress σ required for separating the atomic planes is a function of the distance r, and increases with an increase in r until the stress σ attains the cohesive strength σ_c, at which atomic bonds are broken, and after which the strength degrades with ongoing displacement between the atoms (Figure 2.2). In this case, the stress σ is a function of the displacement δ within the cohesive zone near the crack tip; that is,

$$\sigma = f(\delta), \tag{2.8}$$

where δ is the displacement from equilibrium, defined by $\delta = r - r_0$. If the specific functional form of $f(\delta)$ is known, relation (2.8) is a self-consistent constitutive relation governing the bond breaking process.

Let us assume that the stress versus displacement curve can be approximated by a sine curve with wavelength λ (Figure 2.2). Under this assumption, the stress σ is simply expressed as follows:

$$\sigma = \sigma_c \sin\left(\frac{2\pi\delta}{\lambda}\right). \tag{2.9}$$

At small displacements where $(2\pi\delta/\lambda) \ll 1$ holds, Eq. (2.9) is reduced to

$$\sigma \cong \sigma_c \left(\frac{2\pi\delta}{\lambda}\right). \tag{2.10}$$

If it is assumed that Hooke's law holds at these small displacements, then

$$\sigma = E\left(\frac{\delta}{r_0}\right), \tag{2.11}$$

where E is Young's modulus. From Eqs. (2.10) and (2.11), we have

$$\sigma_c = \frac{\lambda}{2\pi}\frac{E}{r_0}. \tag{2.12}$$

The energy G_c required for breaking atomic bonds is equal to the work done up to the breakdown of the bonds, so that

$$G_c = \int_0^{\lambda/2} \sigma_c \sin\left(\frac{2\pi\delta}{\lambda}\right) d\delta = \frac{\lambda\sigma_c}{\pi}. \tag{2.13}$$