4.03 Fracture and Frictional Mechanics: Theory

Y Fialko, University of California San Diego, La Jolla, CA, USA
© 2015 Elsevier B.V. All rights reserved.

4.03.1 Introduction 74
4.03.2 Linear Elastic Fracture Mechanics 74
4.03.2.1 Singular Crack Models 75
4.03.2.2 Planar Breakdown Zone Models 75
4.03.3 The Governing Equations 77
4.03.4 Exact Solutions for Quasistatic Two-Dimensional Planar Cracks 77
4.03.5 Shear Cracks Governed by Rate-and-State Friction Laws 80
4.03.6 Dynamic Effects 81
4.03.7 Fracture Energy 83
4.03.8 Coupling between Elastodynamics and Shear Heating 85
4.03.9 Conclusions 88
Acknowledgments 89
References 89

Nomenclature

- \( e \): Small distance
- \( D \): Average fault slip
- \( f_\theta(\theta) \): Azimuthal dependence of a near-tip stress field
- \( T \): Temperature scale
- \( w \): Nondimensional width of the fault slip zone
- \( \sigma^0_{ij} \): Stress tensor prior to crack propagation
- \( \sigma^1_{ij} \): Stress tensor after crack propagation
- \( A \): Rupture area
- \( c \): Specific heat
- \( D(x) \): Fault slip
- \( D_c \): Critical weakening displacement
- \( D_m \): Maximum coseismic displacement
- \( F \): Half-length of a developed part of a crack on which friction has dropped to a residual level \( \sigma_d \)
- \( f_i \): Body force
- \( f_s \): Coefficient of static friction
- \( G \): Release rate of mechanical energy (per unit crack length)
- \( G_c \): Effective fracture energy (per unit crack length)
- \( \Im \): Imaginary part of a complex argument
- \( \Re \): Real part of a complex argument
- \( R^* \): Dynamic length of the process zone at the crack tip
- \( r_0 \): Effective radius of the crack tip
- \( S \): Minor axis of an elliptical cavity
- \( S_t \): Boundary between elastically and inelastically deforming materials
- \( T \): Temperature
- \( t \): Time
- \( U_{ec} \): Elastic strain energy
- \( U_{f} \): Frictional losses
- \( U_G \): Fracture energy
- \( U_t \): Potential energy of remotely applied stresses
- \( U_r \): Radiated energy
- \( V_p \): P-wave velocity
- \( V_t \): Rupture velocity
- \( V_s \): S-wave velocity
- \( w \): Half-width of the fault slip zone
- \( x, y, z(x_1, x_2, x_3) \): Spatial coordinates
- \( \alpha \): Nondimensional factor
- \( \gamma \): Nondimensional factor
- \( \delta_{ij} \): Kronecker delta function
- \( \Delta S_1 \): Surface bounding a prospective increment of the process zone at the crack tip in the result of crack growth
- \( \Delta S_2 \): Surface bounding a prospective decrement of the process zone at the crack tip in the result of crack growth
- \( \Delta U_p \): Change in the potential energy
- \( \delta W \): Work done by external forces
- \( \Delta \sigma \): Stress drop
- \( \epsilon_{ij} \): Strain tensor
- \( \theta \): Nondimensional temperature
- \( \kappa \): Thermal diffusivity
4.03.1 Introduction

Seismic and geodetic observations indicate that most earthquakes are a result of unstable localized shear on quasiplanar faults (Dahlen and Tromp, 1998; Gutenberg and Richter, 1949). Because the thickness of earthquake rupture zones that accommodate slip is much smaller than their characteristic in-plane dimensions, it is natural to idealize earthquake ruptures as shear cracks. Development, propagation, and arrest of shear cracks are subject of the earthquake fracture mechanics. Unlike the engineering fracture mechanics that mainly concerns itself with criteria and details of failure at the tip of tensile cracks propagating through ‘intact’ solids (Freund, 1998; Lawn, 1993), the earthquake fracture mechanics must consider both the inelastic yielding at the rupture fronts and the evolution of friction (in general, rate- and slip-dependent) on the rest of the slipping surface (see Chapters 4.04, 4.05 and 4.06). Although a distinction is sometimes made between the crack models and friction models of an earthquake source (e.g., Kanamori and Brodsky, 2004), the two processes are intrinsically coupled and should be considered jointly. Note that the shear crack propagation does not necessarily imply creation of a new fault in intact rocks, but also refers to slip on a preexisting (e.g., previously ruptured) interface. Mathematically, the crack growth in unconfined intact media and on preexisting faults is very similar, provided that the slip surface is planar. While shear cracks in unconfined intact media tend to propagate out of their initial planes, such tendency is suppressed at high confining pressures (e.g., Broberg, 1987; Lockner et al., 1992; Melin, 1986), and in the succeeding text, we limit our attention to planar ruptures.

The redistribution of stress and strain due to a spatially heterogeneous fault slip can be described using either kinematic (displacement-controlled boundary conditions) or dynamic (stress-controlled boundary conditions) approach. Kinematic (e.g., dislocation) models are useful if the fault slip is known or can be inferred with sufficient accuracy, for instance, from seismological, geodetic, or geologic observations (Bilby and Eshelby, 1968; Okada, 1985; Savage, 1998; Steketeet, 1958; Vvedenskaya, 1959). Dynamic (fracture mechanics) models potentially have a greater predictive power, as they solve for, rather than stipulate, the details of fault slip for given initial stress conditions (Andrews, 1976a; Ben-Zion and Rice, 1997; Burridge and Halliday, 1971; Day, 1982; Freund, 1979; Madariaga, 1976). The time-dependent boundary conditions on the fault are usually deduced by using constitutive laws that relate kinetic friction to the magnitude and velocity of fault slip, preseismic stress, temperature, and other state variables pertinent to the evolution of stress on a slipping interface. The kinematic and dynamic approaches give rise to identical results for the same slip distribution. Dislocation models are well understood and are widely employed in inversions of seismic and geodetic data for the rupture geometry and spatiotemporal details of slip (e.g., Delouis et al., 2002; Fialko et al., 2005; Hartzell and Heaton, 1983). Fracture mechanics models are intrinsically more complex and less constrained but nonetheless increasingly used to interpret high-quality near-field seismic data (Aochi and Madariaga, 2003; Oglesby and Archuleta, 2001; Peyrat et al., 2001). This chapter will focus on the fracture mechanics approach, and kinematic models will not be discussed. Consequently, the term ‘dynamic’ will be used to describe time-dependent aspects of rupture for which inertial effects are important, that is, in a meaning that is opposite to ‘static’ (rather than ‘kinematic’) descriptions. There exist a number of excellent texts covering the fundamentals of fracture mechanics, with applications to the earthquake source seismology, including Cherepanov (1979), Rice (1980), Rudnicki (1980), Li (1987), Segall (1991), Scholz (2002), and Ben-Zion (2003), among others. This chapter will give an overview of basic aspects of fracture mechanics and some recent theoretical developments.

4.03.2 Linear Elastic Fracture Mechanics

It is well known that structural flaws and discontinuities such as cracks, voids, and inclusions of dissimilar materials give rise to local stress perturbations that may significantly amplify the average applied stress. A classic example is an elliptical cavity in an elastic plate subject to a uniform extension (Inglis, 1913; Lawn, 1993). Provided that the major and minor axes of the cavity are L and S, respectively, the orientation of the remote tensile stress $\sigma_0$ is orthogonal to the major axis, and the cavity walls are stress-free, the component of stress parallel to the remote tension assumes a value of $\sigma_0 (1 + 2L/S) = \sigma_0 (1 + 2\sqrt{L/r})$ at the cavity tip, where $r = S^2/L$ is the radius of the cavity tip. For an extreme case of a sharp slit, $S/L \rightarrow 0$, the stress at the tip scales as $\sigma_0 \sqrt{L}/\sqrt{r}$, i.e., becomes unbounded for however small remote loading $\sigma_0$. Full analytic solutions for the stress distribution around sharp cracks (see Section 4.03.4; also, Freund, 1998; Lawn, 1993; Rice, 1968a) indicate that the stress field indeed has a characteristic square root singularity

$$\sigma_0|_{r \rightarrow 0} \sim \frac{K}{\sqrt{2\pi r}} f_0(\theta)$$

where $K \sim O(\Delta \sigma \sqrt{L})$ is the stress intensity factor that depends on the crack geometry and loading configuration, $\Delta \sigma$ being the

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>Shear modulus</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Poisson ratio</td>
</tr>
<tr>
<td>$\xi$</td>
<td>Dummy variable</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Density</td>
</tr>
<tr>
<td>$\psi$</td>
<td>Complex variable</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>Remotely applied stress</td>
</tr>
<tr>
<td>$\sigma_d$</td>
<td>Residual stress on the developed part of a crack</td>
</tr>
<tr>
<td>$\sigma_{ij}$</td>
<td>Stress tensor</td>
</tr>
<tr>
<td>$\sigma_n$</td>
<td>Normal stress resolved on a fault</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>Static strength, or the yield stress in the crack tip process zone</td>
</tr>
<tr>
<td>$\sigma_f$</td>
<td>Shear stress resolved on a fault</td>
</tr>
<tr>
<td>$\chi$</td>
<td>Nondimensional along-crack coordinate</td>
</tr>
<tr>
<td>$\psi$</td>
<td>Nondimensional half-length of the developed part of a crack</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Analytic function of a complex argument</td>
</tr>
</tbody>
</table>
difference between the far-field stress and stress resolved on the crack walls (hereafter referred to as the stress drop; for an empty crack under remote tension, $\Delta \sigma = \sigma_0$), $r$ is now the distance to the crack tip measured from the crack exterior, and $f_0(\theta)$ is a function characterizing the azimuthal dependence of the near-tip stress field (Lawn, 1993; Rice, 1980). Because the governing equations and mathematical structure of solutions are identical for the tensile (mode I), in-plane shear (mode II), and antiplane shear (mode III) cracks (see Section 4.03.4), we will use examples of both tensile and shear cracks to highlight universal features of and important differences between the shear and tensile rupture modes. Such a comparison is instructive because many concepts of fracture mechanics have been developed for tensile failure that is most common in engineering applications and subsequently borrowed for seismologic applications that mostly deal with shear failure. Significant differences in the ambient conditions warrant a careful evaluation of the range of applicabilities of basic assumptions behind the fracture mechanics models (e.g., Rubin, 1993).

4.03.2.1 Singular Crack Models

The square root singularity in the stress field (eqn [1]) is a common feature of all crack models assuming a constant stress drop and elastic deformation of the ambient solid, regardless of the mode of failure, and rupture velocity (as long as the latter remains subsonic, see Section 4.03.5). While this stress singularity is admissible on thermodynamic grounds (in particular, it is integrable, so that the elastic strain energy is finite in any closed volume containing the crack tip), it is clearly unrealistic as no material is able to support infinite stresses. Theoretical arguments supported by a large number of observations suggest that the assumption of a perfect brittle behavior (i.e., elastic deformation of the unbroken host up to the onset of fracture) is violated in some zone around the crack tip (Atkinson, 1987; Irwin, 1957; Lawn, 1993) (Figure 1).

In this zone (commonly referred to as the process or breakdown zone), inelastic yielding prevents stress increases in excess of a certain peak value $\sigma_0$ that presumably represents the microscopic strength of a material. The size of the process zone $r_0$ that corresponds to equilibrium (i.e., a crack on a verge of propagating) may be interpreted as the effective radius of the crack tip. Outside of the process zone (at distances $r > r_0$), stresses are below the failure envelope, and the material deforms elastically. If the size of the equilibrium process zone is negligible compared to the crack length, as well as any other characteristic dimensions (e.g., those of an encompassing body), a condition termed small-scale yielding (Rice, 1965a) is achieved, such that the stress and strain fields in the intact material are not appreciably different from those predicted in the absence of a process zone. This is the realm of the linear elastic fracture mechanics (LEFM). According to LEFM, the near-tip ($r_0 < r < L$) stress field is completely specified by the scalar multiplier on the singular stress field, the stress intensity factor $K$ (eqn [1]), which can be found by solving an elastic problem for a prescribed crack geometry and loading conditions. The crack propagation occurs when the stress intensity factor exceeds a critical value, $K > K_c$. The critical stress intensity factor, or the fracture toughness $K_c$, is believed to be a material property, independent of the crack length and loading configuration (although fracture properties may vary for different modes of failure, e.g., $K_{cI}$ and $K_{cII}$, similar to differences between macroscopic tensile and shear strengths). To the extent that the microscopic yield strength $\sigma_0$ and the effective equilibrium curvature of the crack tip, $r_0$, may be deemed physical properties, the fracture toughness $K_c$ may be interpreted as a product $\sigma_0 \sqrt{r_0} = \text{const.}$

The critical stress intensity factor is a local fracture criterion, as it quantifies the magnitude of the near-tip stress field on the verge of failure. However, it can be readily related to global parameters characterizing changes in the strain energy $\partial U_e$ and potential energy of applied stresses $\partial U_i$ in the entire body due to a crack extension $\partial L$, such as the energy release rate $G_c = - (\partial U_e + \partial U_i)/\partial L$,

$$G_c = \frac{2K_c^2}{2\mu}$$

where $\mu$ is the shear modulus of an intact material and $2$ assumes values of $(1 - \nu)$ and unity for mode II and mode III loading, respectively (Irwin, 1957). For ideally brittle materials, the energy release rate may be in turn associated with the specific surface energy spent on breaking the intermolecular bonds (Griffith, 1920; Lawn, 1993). Further analysis of the breakdown process at the crack tip requires explicit consideration of the details of stress concentration in the tip region.

Figure 1  An idealized view of stress variations and yielding at the crack tip. $r_0$ is the characteristic dimension of an inelastic zone in which stresses exceed the yield threshold $\sigma_0$. Solid curve shows the theoretically predicted stress increase with a characteristic $1/\sqrt{r}$ singularity. The theoretical prediction breaks down at distances $r < r_0$ but may be adequate for $r > r_0$.

4.03.2.2 Planar Breakdown Zone Models

A simple yet powerful extension of the LEFM formulation is the displacement-weakening model, which postulates that (i) the breakdown process is confined to the crack plane, (ii) inelastic deformation begins when stresses at the crack tip reach some critical level $\sigma_0$, and (iii) yielding is complete when the crack wall displacement exceeds some critical value $D_s$ (Barenblatt, 1959; Dugdale, 1960; Leonov and Panasyuk, 1959). In case of tensile (mode I) cracks, $\sigma_0$ represents the local tensile strength in the breakdown zone, and $D_s$ is the critical opening displacement beyond which there is no cohesion between the crack walls. In case of shear (mode II and III) cracks (Figure 2), $\sigma_0$ represents either the shear strength (for ruptures propagating through an intact solid) or the peak static friction (for ruptures propagating along a preexisting fault), and $D_s$ is the
slip-weakening distance at which a transition to the kinetic friction is complete (Ida, 1972; Palmer and Rice, 1973). Under these assumptions, the fracture energy may be defined as work required to evolve stresses acting on the crack plane from \( \sigma_s \) to the residual value \( \sigma_0 \) and is of the order of \( (\sigma_s - \sigma_d) / D_c \) depending on the details of the displacement-weakening relation (Figure 3). For an ideal brittle fracture that involves severing of intermolecular bonds ahead of the crack tip, \( \sigma_s \) may approach theoretical strength, \( \mu/10 \sim O(10^7 \sim 10^{10} \text{ Pa}) \), and \( D_c \) may be of the order of the crystal lattice spacing \( (10^{-10} - 10^{-9} \text{ m}) \), yielding \( G_c \sim O(1 \text{ J m}^{-2}) \). This is close to the experimentally measured fracture energies of highly brittle crystals and glasses (Griffith, 1920; Lawn, 1993). At the same time, laboratory measurements of polycrystalline aggregates (such as rocks, ceramics, and metals) reveal much higher fracture energies ranging from \( 10^{-10} \text{ J m}^{-2} \) (for tensile failure) to \( 10^3 \text{ J m}^{-2} \) (for shear failure), presumably reflecting dependence of the effective fracture energy on factors such as stress homogeneity, grain size, and distribution of inhomogeneities 

### Figure 2
Schematic view of a shear rupture expanding at a constant velocity \( v \). Conditions corresponding to the two-dimensional mode II and mode III loading are approximately satisfied at the rupture fronts orthogonal and parallel to the local slip vector, respectively. For a cracklike rupture, slip occurs on the entire boundary bounded by the rupture front. For a pulse-like rupture, slip is confined to an area between the rupture front (solid line) and a trailing healing front (dashed line).

### Figure 3
Possible dependence of the breakdown stress \( \sigma \) on the crack wall displacement \( D \) within the crack tip process zone. Solid line: a schematic representation of the experimentally measured displacement-weakening relations for rocks (Hashida et al., 1993; Li, 1987). Dashed and dotted lines: approximations of the displacement-weakening relation assuming no dependence of the yield stress on \( D \) (dotted line) and linear weakening (dashed line). The area beneath all curves is the same and equals to the fracture energy \( G_c = \alpha \sigma D_c \), where \( \alpha \) is a numerical factor that depends on the particular form of the displacement-weakening relationship. For example, \( \alpha = 1 \) for the constant yield stress model, and \( \alpha = 0.5 \) for the linear weakening model.

where \( \gamma \) is a nondimensional factor of the order of unity that depends on the displacement-weakening relation (e.g., as shown in Figure 3). The small-scale yielding condition requires \( R \ll L \). Parallels may be drawn between the size of the breakdown zone \( R \) in the displacement-weakening model and the critical radius of the crack tip \( r_0 \) in the LEFM model

\[ R = \gamma \frac{R}{\sigma_s - \sigma_d} \frac{D_c}{D} \]
and the average fault slip \( D \):  
\[
D \sim \frac{\Delta \sigma}{\mu} L  \tag{4}
\]
where \( L \) is the characteristic fault dimension (for isometric ruptures) or the least dimension of the slip area (in case of ruptures having irregular shape or high aspect ratios). The assumption of a predominantly elastic behavior of the Earth’s crust appears to be in a good agreement with seismic (Aki and Richards, 1980; Gutenberg and Richter, 1949; Vvedenskaya, 1959) and geodetic (Fialko, 2004b; Reid, 1910) observations of the instantaneous response of crustal rocks to major earthquakes. The brittle–elastic model gives rise to several fundamental scaling relationships used in earthquake seismology. For example, the scalar seismic moment \( M_0 \) is a measure of the earthquake size:

\[
M_0 = DA\mu  \tag{5}
\]
where \( A \) is the rupture area (Aki, 1967; Kostrov, 1974; also, see Chapter 4.02). The seismic moment is related to a coseismic change in the potential energy of elastic deformation \( \Delta U_e \) (e.g., Kanamori, 2004),

\[
\Delta U_e = \frac{\Delta \sigma}{2\mu} M_0  \tag{6}
\]

For isometric ruptures, \( A \sim L^2 \), and by combining eqns [4] and [5], one obtains a well-known scaling between the scalar seismic moment, the rupture size, and the stress drop (e.g., Dahlen and Tromp, 1998; Kanamori and Anderson, 1975, Chapter 4.02):

\[
M_0 \sim \Delta \sigma L^3  \tag{7}
\]

### 4.03.3 The Governing Equations

Consider a three-dimensional medium which points are uniquely characterized by Cartesian coordinates \( x_i \) \((i = 1, 2, 3)\) in some reference state prior to the fault-induced deformation. Fault slip gives rise to a displacement field \( u_i \). The displacements \( u_i \) are in general continuous (differentiable), except across the slipped part of the fault. The strain tensor is related to the displacement gradients as follows:

\[
\epsilon_{ij} = \frac{1}{2} (u_{i,j} + u_{j,i})  \tag{8}
\]

where the comma operator as usual denotes differentiation with respect to spatial coordinates, \( a_i = \partial a / \partial x_i \). Equation [8] assumes that strains are small \((\epsilon_{ij} \ll 1)\), so that terms that are quadratic in displacement gradients can be safely neglected (e.g., Landau and Lifshitz, 1986; Malvern, 1969). The assumption of an infinitesimal strain implies no difference between the material (Lagrangian) and spatial (Eulerian) reference frames. Typical strains associated with earthquake ruptures are of the order of \( 10^{-4} \)–\( 10^{-5} \) (Kanamori and Anderson, 1975; Scholz, 2002) so that the infinitesimal strain approximation is likely justified. For sufficiently small strain changes (with respect to some reference state), laboratory data and theoretical arguments suggest a linear dependence of strain perturbation on the causal stress change. For an isotropic homogeneous elastic material, the corresponding relationship between stresses and strains is given by Hooke’s law (e.g., Landau and Lifshitz, 1986; Timoshenko and Goodier, 1970):

\[
\epsilon_{ij} = \frac{1}{2(1 + \nu)} \left( \sigma_{ij} - \nu \delta_{ij} \sigma_{kk} \right)  \tag{9}
\]

where \( \nu \) and \( \mu \) are the shear modulus and Poisson’s ratio, respectively, \( \delta_{ij} \) is the Kronecker delta function, and repeating indexes imply summation.

Conservation of a linear momentum in continuous media gives rise to the Navier–Cauchy equations of equilibrium (Malvern, 1969):

\[
f_i + \frac{\partial \sigma_{ij}}{\partial x_j} = \frac{\partial^2 u_i}{\partial t^2}  \tag{10}
\]

where \( f_i \) is the body force (e.g., due to gravity) and \( t \) is time. Due to the linearity of equilibrium equations [10], it is possible to represent the full stress tensor as a superposition of some background (e.g., lithostatic and regional tectonic) stress and a perturbation due to fault slip, such that the latter satisfies a homogeneous case of eqns [10]. Unless noted otherwise, in the succeeding text, we assume that the stress tensor \( \sigma_{ij} \) denotes only stress perturbations due to fault displacements. For convenience, the indicial nomenclature for spatial coordinates \((x_1, x_2, x_3)\) will be used interchangeably with the traditional component notation \((x, y, z)\) throughout the rest of the text.

In order to close the problem formulation, some constitutive law relating slip \( D \), slip velocity \( \partial D / \partial t \), resolved shear and normal stresses, \( \sigma_{ij} \) and \( \sigma_{ij} \), pore fluid pressure \( p \), temperature \( T \), etc. needs to be prescribed on surfaces that violate assumptions of continuity, for example, faults and cracks. Examples are the Mohr–Coulomb (Byerlee, 1978) and rate-and-state friction (see Chapter 4.04), flash melting (Molinaro et al., 1999; Rice, 2006), thermal pressurization (Andrews, 2002; Segall and Rice, 1995; Sibson, 1973), viscous rheology (Fialko and Khazan, 2005), etc.

### 4.03.4 Exact Solutions for Quasistatic Two-Dimensional Planar Cracks

In case of two-dimensional deformation (e.g., plane or anti-plane strain or plane stress), elastic solutions for stresses and displacements can be generally expressed in terms of two analytic functions of a complex variable \( \zeta = x + iy \) (Kieldysh, 1990; Muskhelishvili, 1933). For simplicity, here, we consider loading that is symmetric about the center of the crack \( x = 0 \). In this case, both stresses \( \sigma_{ij} \) and displacements \( u_{ij} \) can be expressed through a single analytic function \( \phi(\zeta) \) (Westergaard, 1939). First, we demonstrate that the mathematical structure of solutions for stresses and displacements in the crack plane \((y = 0)\) is identical for tensile (mode I), in-plane shear (mode II), and antiplane shear (mode III) cracks. In particular, for tensile...
(mode I) cracks, stresses and displacements can be found from \( \phi(z) \) as (e.g., Khazan and Fialko, 2001; Muskheilishvili, 1953)

\[
\begin{align*}
\sigma_{xy} &= 2(\Re \phi' + \gamma \Im \phi^0) \quad [11] \\
\sigma_{yy} &= -2\gamma \Re \phi^0 \quad [12] \\
u_y &= \frac{1}{\mu} (2(1-v)\Im \phi - \gamma \Re \phi') \quad [13]
\end{align*}
\]

where \( \Re \) and \( \Im \) denote the real and imaginary parts of a complex argument. For plane strain shear (mode II) cracks, conditions of symmetry imply that \( \sigma_{xy} = 0 \) on the crack plane \((y = 0)\), and the corresponding equilibrium equations are

\[
\begin{align*}
\sigma_{xy} &= -2(\Im \phi' + \gamma \Re \phi^0) \quad [14] \\
u_y &= \frac{1}{\mu} (2(1-v)\Re \phi + \gamma \Im \phi') \quad [15]
\end{align*}
\]

Upon making a substitution \( \phi = -i\phi_1 \), one can see that the unknown shear stress \( \sigma_{xy} \) and displacement \( u_y \) on the crack plane may be expressed through a new function \( \phi_1 \) in the same manner as the normal stress \( \sigma_{yy} \) and displacement \( u_x \) are expressed through \( \phi \) (eqns [11] and [13]),

\[
\begin{align*}
\sigma_{xy} &= 2(\Re \phi_1' - \gamma \Im \phi_1^0) \quad [16] \\
u_y &= \frac{1}{\mu} (2(1-v)\Re \phi_1 + \gamma \Im \phi_1') \quad [17]
\end{align*}
\]

For antiplane shear (mode III) cracks, expressions for the relevant stress and displacement components are

\[
\begin{align*}
\sigma_{xz} &= -\Im \phi' \quad [18] \\
u_x &= \frac{1}{\mu} \Re \phi \quad [19]
\end{align*}
\]

Upon making a substitution \( \phi = -i\phi_2 \), one can see that the dependence of the unknown quantities \( \sigma_{xz} \) and \( (1-v)u_y \) on the analytic function \( \phi_2 \) is analogous to the dependence of \( \sigma_{xy} \) and \( u_y \) on \( \phi_1 \) obtained for the in-plane shear crack for \( y = 0 \) (eqns [16]–[17]). This analogy mandates that the mathematical structure of solutions for the tensile, in-plane, and antiplane components of stress for the corresponding crack modes is identical, provided that the boundary conditions on the crack plane (the along-crack distribution of the driving stress and the displacement-weakening relationship) are analogous. Solutions for the crack wall displacements are also identical for different modes, although expressions for displacements for mode III cracks will differ from those for mode I and mode II cracks by a factor of \((1-v)\). Hence, we focus on a particular case of an in-plane shear (mode II) crack. The boundary conditions for the potential function \( \phi_1 \) are as follows:

\[
\begin{align*}
\Re \phi_1' &= \sigma(x)/2 \quad \text{for} \ |x| < L \quad [20] \\
\Im \phi_1' &= 0 \quad \text{for} \ |x| > L \quad [21] \\
\phi_1' &= -\sigma_0/2 \quad \text{for} \ |z| \to \infty \quad [22]
\end{align*}
\]

where \( \sigma(x) \) is the distribution of shear stress on the crack surface and \( \sigma_0 \) is the applied shear stress at infinity (hereafter referred to as prestress). The boundary condition [21] postulates no slip beyond the rupture front (see eqn [17]). An explicit solution for the function \( \phi_1 \) that is analytic in the upper half plane and satisfies the boundary conditions [20]–[22] is given by the Keldysh–Sedov formula (Gakhov, 1966; Khazan and Fialko, 1995, 2001; Muskheilishvili, 1953):

\[
\begin{align*}
\phi_1'(
\frac{\zeta}{\zeta + L}) &= \frac{1}{2\pi i} \int_{L - i \infty}^{L + i \infty} \frac{d\zeta}{(\zeta - \zeta_1)^{1/2}} \frac{\sigma(\zeta)}{\xi - \zeta} + \frac{\sigma_0}{2} \frac{\zeta + L}{\zeta - L} + \frac{C}{(\zeta - L)^{1/2}} \quad [23]
\end{align*}
\]

where \( C \) is an arbitrary constant. From eqn [16], an asymptotic behavior of \( \phi_1' \) at infinity is

\[
\begin{align*}
\lim_{|\zeta| \to \infty} \phi_1' &= -\frac{1}{2\pi i |\zeta|} \int_{L - i \infty}^{L + i \infty} \frac{d\zeta}{(\zeta - \zeta_1)^{1/2}} \frac{\sigma(\zeta)}{\xi - \zeta} + \frac{\sigma_0}{2} + \frac{L \sigma_0 + C}{|\zeta|} \quad [24]
\end{align*}
\]

One can readily determine the unknown constant \( C \) from eqn [16] by satisfying the boundary condition [22]. The final expression for the derivative of analytic function \( \phi_1 \) is

\[
\begin{align*}
\phi_1' &= \frac{1}{2\pi i |\zeta|} \int_{L - i \infty}^{L + i \infty} \frac{d\zeta}{(\zeta - \zeta_1)^{1/2}} \frac{\sigma(\zeta)}{\xi - \zeta} + \frac{\sigma_0}{2} + \frac{L \sigma_0 + C}{|\zeta|} \quad [25]
\end{align*}
\]

For any physically admissible failure model, the maximum stress within the slipped region is bounded by the yield strength, \( \sigma_{xy} < \sigma_y \) for \(|x| < L \). Furthermore, it is reasonable to expect that \( \sigma_{xy} \to \sigma_y \) as \(|x| \to L \). The shear stress \( \sigma_{xy}(z) \) in the ‘locked’ region ahead of the rupture front is then given by the real part of \( \phi_1 \) (see eqn [16]). Sufficiently close to the crack tip (i.e., for \( z = L + \epsilon \), such that \( 3\epsilon = 0, 0 < \epsilon \ll L \), we obtain

\[
\begin{align*}
\sigma_{xy}(L + \epsilon) &= \frac{1}{2\pi i |\zeta|} \int_{L - i \infty}^{L + i \infty} \frac{d\zeta}{(\zeta - \zeta_1)^{1/2}} \frac{\sigma(\zeta) - \sigma_0}{L^2 - z^2} + \sigma_0 \quad [26]
\end{align*}
\]

The first term on the right-hand side of eqn [26] is of the order of \( 1/\sqrt{\epsilon} \) and represents the LEFM approximation (see Section 4.03.2). By comparing eqns [26] and [1], one can formally introduce the stress intensity factor,

\[
K = \frac{1}{\pi} \left[ L \right]^{1/2} \int_{-L}^{L} d\zeta \frac{\sigma(\zeta) - \sigma_0}{(L^2 - z^2)^{1/2}} \quad [27]
\]

which exhibits the expected scaling \( K \propto \Delta \sigma \sqrt{L} \). Physically plausible crack models require that stresses are finite everywhere. From eqns [26] and [27], the requirement of the absence of a stress singularity is met if (and only if)

\[
\int_{-L}^{L} d\zeta \frac{\sigma(\zeta) - \sigma_0}{(L^2 - z^2)^{1/2}} = 0 \quad [28]
\]

Any realistic distribution of the driving stress resolved the crack surface must satisfy the integral constraint [28]. This implies that the driving stress \( (\sigma(x) - \sigma_0) \) must change sign along the crack \((|x| < L)\). For example, the stress drop \( (\sigma(x) - \sigma_0) \) in the central part of the crack needs to be balanced by the material strength or the high transient friction \( \sigma_y \) within the process zone.

Displacements of the crack walls \( u(x) \) corresponding to the instantaneous shear stress \( \sigma(x) \) can be found by differentiating equation [17] for \( y = 0 \),
and making use of expression [26] to integrate the resulting differential equation [29] with the initial condition \( u_0(-L) = 0 \). The respective solution is

\[
D(x) = \frac{2(1-v)}{\pi \mu} \int_{x}^{L} \left( \frac{L^2 - \zeta^2}{c} \right)^{1/2} \, d\zeta
\]

where \( D(x) = 2u_0(x) \) is slip between the crack walls. The problem is closed by specifying a constitutive relationship between slip \( D \) and kinetic friction \( \sigma \). In general, such a relationship may include dependence of friction on slip rate, slip history, local temperature, and other state variables (Dieterich, 1979; Blanpied et al., 1995; Ruina, 1983, Chapters 4.04, 4.05 and 4.06). To gain a further analytic insight, here, we consider a simple slip-weakening relationship

\[
\sigma(x) = \sigma_s \text{ for } D(x) < D_c
\]

\[
\sigma(x) = \sigma_d \text{ for } D(x) > D_c
\]

where \( \sigma_s \) is the yield strength or static friction in the process zone, \( \sigma_d \) is the residual kinetic friction on the developed part of the crack, and \( D_c \) is the critical slip-weakening displacement corresponding to a transition from \( \sigma_s \) to \( \sigma_d \) (see Figure 3). The size of the process zone, \( R \), is defined by a requirement that the fault slip is subcritical within the process zone, \( D(\vert F \vert) = D_0 \), where \( F = L - R \) is the length of the developed part of a crack on which the friction has dropped to a residual value \( \sigma_d \). Figure 4 illustrates the geometry of the problem. In case of a piece-wise constant distribution of shear stress along the crack \( \sigma(x) \) given by eqn [32], as shown in Figure 4, expressions [28] and [30] can be readily integrated to provide closed form analytic solutions. In particular, integration of expression [28] allows one to determine a relative size of the process zone with respect to the crack half-length:

\[
R \over L = 2 \sin^2 \left( \frac{\pi \sigma_0 - \sigma_d}{4 \sigma_s - \sigma_d} \right)
\]

Evaluation of integral [30] gives rise to

\[
D(x) = \frac{2(1-v)I}{\pi \mu} (\sigma_0 - \sigma_d) I(\chi, \psi)
\]

where \( \chi = x/L \) and \( \psi = F/L \) are the nondimensional along-crack coordinate and the half-length of the developed part of the crack, respectively, and function \( I \) is given by the following equation:

\[
I(U, V) = (V + U) \log \left( \frac{1 + U V + \sqrt{(1 - U^2)(1 - V^2)}}{V + U} \right)
\]

\[
+ (V - U) \log \left( \frac{1 - U V + \sqrt{(1 - U^2)(1 - V^2)}}{V - U} \right)
\]

At the base of the process zone, \( \chi = \psi \), eqn [33] reduces to

\[
D(F) = \frac{4(1-v)}{\pi \mu} F(\sigma_s - \sigma_d) \log \frac{1}{\psi}
\]

For a crack that is on the verge of propagating, slip at the base of the process zone (\( \chi = F \)) equals the critical weakening displacement \( D_c \). A simple dimensional analysis of eqn [35] then reveals a characteristic length scale \( L_c \):

\[
L_c = \frac{\pi D_c}{4(1-v)\sigma_s - \sigma_d}
\]

Using eqn [36], eqn [35] may be written

\[
\frac{L_c}{F} = \log \frac{L}{F}
\]

A further rearrangement of eqn [37] yields an explicit expression for the length of the process zone \( R \):

\[
R = F \left( \exp \left( \frac{L_c}{F} \right) - 1 \right)
\]

It is instructive to consider two end-member cases, the long crack limit, \( F \gg L_c \), and the microcrack limit, \( F \ll L_c \). In case of sufficiently long cracks (\( F \gg L_c \)), an asymptotic expansion of eqn [38] (i.e., \( \exp(x) \approx 1 + x \) for \( x \ll 1 \)) indicates that the process zone length is independent of the crack length and equals \( L_c \). This is the realm of small-scale yielding, \( R = \text{const} = L_c \ll L_c \), in which cracks propagate by preserving the structure of the near-tip stress field. By comparing eqns [37] and [3], one can see that the general scaling relationship suggested by dimensional arguments holds; for the case of a constant breakdown stress \( \sigma_s \) within the process zone, the nondimensional coefficient \( \gamma \) in estimate [3] equals \( \pi/4(1-v) \). Exact analytic solutions assuming small-scale yielding give rise to \( \gamma = \pi/2(1-v) \) for a linear slip-weakening relationship (see Figure 3) (e.g., Chen and Knopoff, 1986) and \( \gamma = 9\pi/16(1-v) \) for a breakdown stress that linearly decreases with distance away from the crack tip (Palmer and Rice, 1973). In the microcrack limit, the length of the equilibrium process zone is not constant even if the critical weakening displacement \( D_c \) and the strength drop \( (\sigma_s - \sigma_d) \) are intrinsic material properties.
independent of the ambient stress and loading conditions. In particular, \( R \) is predicted to exponentially increase as the length of the stress drop region \( F \) decreases. Equations [38] and [32] suggest that the presence of sufficiently short cracks (such that \( F < L_{c} \)) has no effect on the macroscopic 'strength' of rocks. In particular, the prestress required for the crack extension must approach the static yield limit, \( \sigma_{y} \), and the size of the yield zone increases without bound, \( L \approx R \to \infty \), for \( F \to 0 \) (see eqn [38]).

Quantitative estimates of the critical length scale \( L_{c} \) are not straightforward because it is not clear whether the slip-weakening distance \( D_{s} \) is indeed a scale-independent material constant (e.g., Barton, 1971; Ohnaka, 2003; Rudnicki, 1980). Laboratory measurements of the evolution of friction on smooth slip interfaces indicate that \( D_{s} \) may be of the order of \( 10^{-5} \) m (Dieterich, 1979; Li, 1987; Marone, 1998). For \( \mu \approx 10^{10} \) Pa and \( (\sigma_{y} - \sigma_{a}) \sim 10^{-7} - 10^{8} \) Pa (likely spanning the range of stress drops for both 'strong' and 'weak' faults), from eqn [36], one obtains \( L_{c} \approx 10^{-5} - 10^{-1} \) m, negligible compared to the characteristic dimension of the smallest recorded earthquakes but comparable to the typical sample size used in the laboratory experiments. An upper bound on \( L_{c} \) may be obtained from estimates of the effective fracture energies of earthquake ruptures. For large (moment magnitude > 6) earthquakes, the seismically inferred fracture energies \( G_{s} = (\sigma_{y} - \sigma_{a})D_{s} \) are of the order of \( 10^{8} - 10^{9} \) J m\(^{-2} \) (Abercrombie and Rice, 2005; Beroza and Spudich, 1988; Hussein, 1987; Ida, 1972), rendering the effective \( D_{s} \approx 0.01-1 \) m. Assuming that the seismically inferred values of \( D_{s} \) are applicable to quasistatic cracks, eqn [36] suggests that a transition from 'micro' to 'macro' rupture regimes occurs at length scales of the order of \( 10^{-3} \) m.

The magnitude of a prestress required to initiate the crack propagation can be found by combining eqns [37] and [32]:

\[
\frac{\sigma_{y} - \sigma_{a}}{\sigma_{y} - \sigma_{a}} = 2 \pi \arcsin \left( \frac{L_{c}}{F} \right) \tag{39}
\]

In the microcrack or large-scale yielding limit, \( F \ll L_{c} \), eqn [39] predicts \( \sigma_{y} \approx \sigma_{a} \), that is, the crack propagation requires ambient stress comparable to the peak static strength of crustal rocks, as discussed in the preceding text. In the small-scale yielding limit, \( F \gg L_{c} \), eqn [39] gives rise to a well-known inverse proportionality between the stress drop \( (\sigma_{y} - \sigma_{a}) \) and the square root of the crack length \( F \) (e.g., Cowie and Scholz, 1992; Kostrov, 1970; Rice, 1968a):

\[
\sigma_{y} = \sigma_{a} + 2 \pi (\sigma_{y} - \sigma_{a}) \left( \frac{2L_{c}}{F} \right)^{1/2} \tag{40}
\]

It follows from eqn [40] that for sufficiently large ruptures, the background tectonic stress required for the rupture propagation does not need to appreciably exceed the residual friction on the slipped surface \( \sigma_{a} \). That is, the stress drop associated with the rupture propagation, \( (\sigma_{y} - \sigma_{a}) \), may be much smaller than the strength drop, \( (\sigma_{y} - \sigma_{a}) \), provided that the rupture size significantly exceeds the critical nucleation size \( L_{c} \). This statement forms the basis of the 'statically strong, but dynamically weak' fault theory (Lapusta and Rice, 2004). According to this theory, major crustal faults may operate at relatively low driving stresses (e.g., sufficient to explain the so-called heat flow paradox of the San Andreas fault (Brune et al., 1969; Lachenbruch, 1980)), even if the peak failure stress required for the onset of dynamic weakening is consistent with Byerlee's law and hydrostatic pore pressures (Byerlee, 1978; Marone, 1998; Scholz, 2002), provided that \( \sigma_{a} \ll \sigma_{y} \). If so, earthquake ruptures must nucleate in areas where \( \sigma_{y} \) approaches \( \sigma_{a} \) (due to either locally increased ambient stress or decreased static strength, e.g., due to high pore fluid pressures) and propagate into areas of relatively low ambient stress. Under this scenario, the overall fault operation must be such that the average stress drop \( \Delta \sigma \) remains relatively small (of the order of 0.1–10 MPa) and essentially independent of the rupture size (Abercrombie, 1995; Kanamori and Anderson, 1975; Scholz, 2002). Because the overall seismic moment release is dominated by the largest events, the implication from the Lapusta and Rice (2004) model is that the Earth crust is not able to support high deviatoric stresses in the vicinity of large active faults. Phenomenologically, this is consistent with the 'weak fault' theory maintaining that the average shear stress \( \sigma_{a} \) resolved on mature faults is of the order of the earthquake stress drops (i.e., up to a few tens of megapascals) and is considerably less than predictions based on Byerlee's law (a few hundreds of megapascals) (e.g., Kanamori and Heaton, 2000). The 'statically strong but dynamically weak' fault theory seeks to reconcile laboratory results from rock friction experiments with seismic observations. Neither the peak shear stress \( \sigma_{y} \) nor the residual dynamic friction \( \sigma_{d} \) can be estimated from seismic data. Both parameters are likely scale-dependent; for example, the peak shear stress \( \sigma_{y} \) may vary from gigapascals on the scale of microasperities and gouge particles (\( 10^{-3} \) m) to the Mohr–Coulomb stress \( f_{s}(\sigma_{a} - \rho) \), where \( f_{s} \) is the static coefficient of friction, \( \sigma_{a} \) is the fault-normal stress, and \( \rho \) is the pore fluid pressure, on the scale of centimeters to meters (i.e., consistent with laboratory data), to values that may be lower still on scales of hundreds of meters to kilometers. Similarly, the residual dynamic friction may depend on the amount of slip (and the rupture size) (e.g., Abercrombie and Rice, 2005; Brown and Fialko, 2012; Kanamori and Heaton, 2000; Rice, 2006); implications from such behavior are further discussed in Section 4.03.7.

4.03.5 Shear Cracks Governed by Rate-and-State Friction Laws

More sophisticated models of earthquake faults combine elasticity with empirical friction laws that relate shear stress to normal stress and slip velocity and slip history. In particular, the rate-and-state phenomenology (Dieterich, 1979; Ruina, 1983) provides a system of differential equations that can be used to define shear stress (e.g., \( \sigma \) in eqn [30]) on a fault plane without oversimplifying assumptions, such as eqns [32]. The resulting system of equations precludes analytic solutions, and numerical experiments are required to investigate the evolution of stress and slip rate on a fault. In the framework of rate-and-state friction, material properties that govern fault slip are \( a \) (rate parameter), \( b \) (state parameter), and the characteristic evolution distance analogous to the critical weakening displacement \( D_{c} \) discussed in the preceding text (e.g., Figure 3; also, see Chapter 4.04).

In particular, numerical simulations performed by Dieterich (1992) revealed that slip instabilities on a
rate-and-state fault require a nucleation zone having a characteristic size greater than \( L_b = D_m/(1 - \beta_0)\sigma_0 \), where \( \sigma_0 \) is the fault-normal stress (see Figure 4 in Chapter 4.04). The existence of a minimum nucleation length scale was previously suggested based on a conditional stability of a spring-box slider (e.g., Gu et al., 1984; Rice, 1993; Ruina, 1983). However, the spring-box slider model predicts that the minimum nucleation length should scale as \( b/(b-a)^{1/2} \), rather than \( b^{-1} \), as reported by Dieterich (1992). This discrepancy may stem from the fact that slip on a deformable fault is intrinsically a higher dimensional (2-D or 3-D) problem compared with a (1-D) problem of motion of a spring-box slider.

Rubin and Ampuero (2005) showed that \( L_b \) is in fact the minimum nucleation length and that for a weakly weakening \((a/b \rightarrow 1)\), that is, nearly velocity neutral friction, larger nucleation areas can exist, with the characteristic length of the order of \( b^{2}/(b-a)^{2}L_b \). The latter can be reached under quasi-steady-state conditions (such that the state variable is nearly constant), while \( L_b \) is the minimum nucleation length when the fault is well above steady state (the state variable is rapidly changing) (Ampuero and Rubin, 2008; Rubin and Ampuero, 2005).

Some parallels may be drawn between the analytic solutions presented in Section 4.03.4 and numerical results for rate–state faults (Ampuero and Rubin, 2008; Dieterich, 1992; Rubin and Ampuero, 2005). If one approximates the velocity change at the tip of a slipping patch on a rate–state fault by a step function (e.g., see Figure 8 in Chapter 4.02), then parameters \( \sigma_0 \) and \( \beta_0 \) of rate-and-state friction are analogous to the difference between the yield strength and the background stress \((\sigma_s - \sigma_0)\) and the strength drop \((\sigma_s - \sigma_d)\), respectively (see Figure 4 in this chapter and Figure 1 in Chapter 4.04). The characteristic length of the process zone \( L_c \) (eqn [36]) is then essentially coincident with the minimum nucleation length \( L_b \) established by Dieterich (1992) based on numerical simulations coupling the rate-and-state friction and elasticity. The correspondence between the two length scales may be interpreted as indicating that no elastodynamic instability is possible in the microcrack regime \((F < L_c)\). On the other hand, a condition \( z/b \rightarrow 1 \) implies \( \sigma_0 \approx \sigma_d \) that is, small-scale yielding (see eqn [40]). Indeed, such conditions were satisfied prior to the onset of elastodynamic instability in numerical experiments of Rubin and Ampuero (2005) performed under the assumption of weak velocity weakening. Using eqns [32] and [38], one can see that in the limit \( R/L \ll 1 \), \( b^{2}/(b-a)^{2}L_c \sim L/R \), that is, the ‘velocity neutral’ nucleation length of Rubin and Ampuero (2005) is essentially the length of a quasistatic crack on the verge of propagation under a constant remote stress \( \sigma_0 \). This is the maximum length of a slip patch satisfying conditions of a quasistatic equilibrium. The main differences between the rate–state model and that shown in Figure 4 is that the latter assumes a piece-wise constant stress distribution on the crack surface and zero slip velocity ahead of the crack tip, while in the rate–state model, the shear stress is a continuous function of distance along the fault, and the fault slip rate is never zero. This analogy between the analytic solution for a static crack and a deformable rate–state fault may hold because of large (orders of magnitude) increases in the slip velocity at the tip of an accelerating slip patch compared to the nominally locked region ahead of the tip; variations in slip rate behind the rupture front are much smaller in comparison, so that a step-like change in slip rate may be a reasonable approximation.

4.03.6 Dynamic Effects

Quasistatic solutions considered in Section 4.03.4 are valid only for rupture speeds that are well below the shear wave velocity \( V_s \). As the rupture velocity increases, the inertial term in the equilibrium equations [10] eventually becomes non-negligible, and the near-tip stress field is significantly altered when \( V_r \) becomes a sizeable fraction of \( V_s \). (Andrews, 1976a; Broberg, 1978; Freund, 1979; Rice, 1980). The most pronounced effects of a high rupture speed are the relativistic shrinking of the in-plane process zone \( R \) and simultaneous increase of stress perturbations off the crack plane (Kame and Yamashita, 1999; Poliakov et al., 2002; Rice et al., 2005). The net result is an increased tendency for branching, bifurcation, and nonsteady propagation, all of which significantly complicate the analytic and numerical treatment of the elastodynamic rupture problem. In particular, the intermittent propagation of the rupture front invalidates the equivalence between the LEFM and slip-weakening formulations (e.g., Freund, 1979), implying a greater dependence of the model results on a specific choice of fracture criteria.

Analysis of the full elastodynamic equilibrium equations [10] reveals that solutions exist for rupture velocities below a limiting speed \( V_l \) which equals to the Rayleigh wave velocity \( V_R \) for mode II cracks and shear wave velocity for mode III cracks (e.g., Freund, 1979; Kostrov and Das, 1988). Solutions become singular as \( V_l \rightarrow V_s \), in particular, the dynamic stress intensity factor and the energy release rate at the crack tip asymptotically vanish. Main fracture mechanics parameters of a steady-state elastodynamic rupture (such as the process zone size and the stress intensity factors) may be readily obtained by multiplying or dividing the respective results from quasistatic solutions by dimensionless coefficients that depend on rupture velocity only. The corresponding coefficients are

\[
f_{II} = (1 - \nu) \left( 1 - \frac{V_r^2}{V_s^2} \right) \sqrt{1 - \frac{V_r^2}{V_s^2} - \left( 2 - \frac{V_r^2}{V_s^2} \right)^2} \tag{41}
\]

\[
f_{III} = \sqrt{1 - \frac{V_r^2}{V_s^2}} \tag{42}
\]

for the mode II and mode III loading, respectively (e.g., Freund, 1998; Rice, 1980). In eqn [41], \( V_s \) is the P-wave velocity. Coefficients \( f_{II} \) and \( f_{III} \) monotonically decrease from unity at \( V_l = 0 \) to zero at \( V_l = V_s \). For instance, the length of the dynamic process zone \( R^* \) is given by \( R^*(V_l) = f_{II,m}(V_l)R \), and the dynamic stress intensity factor for a self-similar expanding crack is \( K_{II,m}(V_l) = f_{II,m}(V_l)K_{II,m}(0) \). Note that for a steady-state self-healing pulse with a fixed stress distribution in the reference frame of a moving pulse, the process zone shortens at high rupture velocity, similar to the case of a self-similar crack, but the dynamic stress intensity factor is independent of \( V_l \).
Rice et al. (2005) confirmed these results with full analytic solutions for a self-healing pulse with a linearly weakening process zone, propagating at a constant rupture speed. They found that the ratio of the dynamic process zone $R^*$ to the pulse length $L^*$ is independent of the rupture speed but is dependent on the ratio of stress drop to strength drop:

$$\frac{\sigma_0 - \sigma_d}{\sigma - \sigma_d} = \frac{\xi}{\pi} - \frac{\xi - \sin \xi}{2\pi \sin^2(\xi/2)}$$

where $\xi = 2 \arcsin \sqrt{R^*/L^*}$ (cf. eqn [32]). The velocity invariance of the relative size of the process zone gives rise to a somewhat unintuitive result that the pulse length $L^*$ vanishes as the rupture accelerates to a limiting speed. At the same time, the amount of slip produced by a self-healing pulse is also invariant with respect to the rupture velocity. Thus, the dynamically shrinking rupture size gives rise to a dramatic increase in the near-field coseismic strain. Figure 5 shows the dependence of the near-tip stress concentration on the rupture speed inferred from the Rice et al. (2005) model. As one can see from Figure 5, the extent of the off-plane yielding and damage substantially increases at high rupture speeds, so that the assumption of a thin in-plane process zone ceases to be valid. Similar results were reported for a semi-infinite crack by Poliakov et al. (2002). Note that the deduced areas of yielding likely underestimate the extent of off-fault damage, as stresses inside the shaded areas in Figure 5 are beyond the failure envelope. By explicitly allowing inelastic deformation, the excess stresses will be relaxed, and the yielding zone will further expand. The enhanced off-fault damage may be one of the mechanisms preventing extreme contractions of the rupture length at high propagation velocities. It may also appreciably modify the expenditure part of the earthquake energy balance, as discussed in the following section.

Elastodynamic slip instabilities with rupture velocity below the limiting speed $V_l$ are referred to as subsonic or subshear rupture. The majority of earthquakes for which high-quality measurements of the rupture speed are available appear to be subsonic. Theoretical models indicate that immediately above the limiting rupture speed $V_l$ the flux of mechanical energy into the crack tip region becomes negative for tensile and mode II shear cracks, effectively prohibiting self-sustained fracture (Broberg, 1999; Freund, 1998). However, physically admissible solutions do exist for shear cracks with rupture speeds spanning the interval between the S-wave and P-wave velocities; such mode of propagation is referred to as intersonic, transonic, or supershear rupture (Andrews, 1976b; Broberg, 1989; Burridge, 1973; Simonov, 1983). Although the mechanics of transition from the subsonic ($V_i < V_l$) to transonic

---

**Figure 5** Stress perturbation around the tip of a propagating slip pulse with a linear slip-weakening process zone (reproduced from Rice JR, Sammis CG, and Parsons R (2005) Off-fault secondary failure induced by a dynamic slip pulse. *Bulletin of the Seismological Society of America* 95: 109–134, their Figure 6). Light shading denotes areas where the Mohr–Coulomb failure envelope is exceeded (likely to fail in shear), and dark shading denotes areas of absolute tension (likely to fail by tensile cracking). Thick tick marks indicate the orientation of optimal planes for right-lateral slip, and thin tick marks indicate the orientation of optimal planes for left-lateral slip. $\Psi$ is an angle between the maximum compression axis and the fault plane. Process zone comprises 10% of the rupture length. Axes are normalized by the quasistatic process zone size.
(\(V_t < V_d\)) propagation is not fully understood, there is experimental (Samudrala et al., 2002; Xia et al., 2004) and seismologic (Archuleta, 1984; Bouchon and Vallee, 2003; Dunham and Archuleta, 2004; Wald and Heaton, 1994) evidence that transonic rupture speeds may be achieved under certain conditions. The spatial structure of the near-tip stress field and the radiation pattern in the transonic regime are markedly different from those due to subsonic ruptures (see Chapter 4.08).

### 4.03.7 Fracture Energy

The concept of fracture energy was originally introduced for tensile cracks by Griffith (1920) to quantify the irreversible work associated with breaking of the intermolecular bonds and creation of a stress-free crack surface. Griffith’s definition based on a global energy balance was subsequently shown to be equivalent to local definitions based on the LEFM and small-scale yielding models (Rice, 1968a; Willis, 1967). For example, for a Barenblatt-type process zone model, the fracture energy is the work spent against the cohesive stress \(\sigma_i\) in the process zone on separating the crack walls by the critical opening distance \(D_c\). An elegant demonstration of the equivalence of global and local definitions of fracture energy for tensile cracks was provided by Rice (1968b) in a form of the path-independent J-integral (also, see Cherepanov, 1968; Eshelby, 1956). Palmer and Rice (1973) extended this technique to the case of shear cracks and defined the shear fracture energy as work required to evolve shear stress on the slip interface from the yield stress (or static friction) \(\sigma_i\) to the residual dynamic friction \(\sigma_d\):

\[
G_c = \int_0^{D_c} (\sigma(D) - \sigma_d)\,dD
\]  

[44]

where \(\sigma(D)\) varies between \(\sigma_i\) and \(\sigma_d\) for \(0 < D < D_c\), respectively. A similar formulation was introduced by Ida (1972). Equation [44] allows a simple insight into the fracture process and has been widely used for interpretations of seismic data. However, several factors may limit its application to the analysis of earthquake ruptures. First, the displacement-weakening model assumes that all inelastic deformation is limited to the slip plane. Both theoretical models (Andrews, 2005; Rice et al., 2005; Rudnicki, 1980) (Figure 5) and field observations (Chester et al., 2005; Fialko, 2004b; Fialko et al., 2002; Li et al., 1998) suggest that the earthquake-induced damage likely extends well off of the fault plane, and the energy dissipated in the fault damage zone may be quite significant (e.g., Andrews, 2005; Ben-Zion and Shi, 2005; Wilson et al., 2004). Second, the fracture energy given by eqn [44] has a clear physical interpretation if the residual dynamic stress \(\sigma_d\) is constant (or at least if the along-fault variations in \(\sigma_d\) are small compared to the strength drop, \(\sigma_i - \sigma_d\)). The second point can be illustrated by considering a traditional representation of the earthquake energy budget:

\[
\Delta U_p = U_f + U_t + U_C
\]  

[45]

where \(\Delta U_p\) is the change in the total potential energy (which includes changes in the elastic strain energy \(\Delta U_e\) and gravitational potential energy); \(U_f\) is the energy radiated in seismic waves; \(U_t\) is the energy dissipated on the well-slipped portion of the fault due to friction, comminution, phase transitions, and other irreversible losses; and \(U_C\) is the fracture energy spent on overcoming high resisting stresses near the crack tip (Dahlen, 1977; Kostrov, 1974; Rivera and Kanamori, 2005; Rudnicki and Freund, 1981). A significant part of \(U_t\) is believed to be ultimately converted into heat (Fialko, 2004a; Sismondo, 1980). Under the approximation of the displacement-weakening model, \(U_C \approx \sigma_j D_c \) and \(U_t \approx \sigma_d D_m\). Assuming that the residual friction is of the order of the peak strength, \(\sigma_d \sim O(\sigma_i)\), and the critical slip-weakening distance is much smaller than the cohesive offset, \(D_c \ll D_m\), the fracture energy is negligible compared to frictional losses in the earthquake energy balance equation [45]. However, if the fault friction progressively decreases with slip, as suggested by the experimental observations and theoretical inferences of the dynamic weakening (Abercrombie and Rice, 2005; Di Toro et al., 2004; Fialko and Khazan, 2005; Goldsby and Tullis, 2002; Tsutsumi and Shimamoto, 1997), the effective slip-weakening distance \(D_c\) is expected to scale with the slip magnitude, and the fracture energy \(U_C\) may not be small compared to \(U_t\). Because neither the slip-weakening distance \(D_c\) nor the residual friction \(\sigma_d\) in this case are material properties (in particular, they may depend on the details of slip history, thickness and permeability of the slip zone, etc.), a distinction between \(U_C\) and \(U_t\) terms in the earthquake energy balance equation [45] becomes somewhat arbitrary. Note that for a rupture on a preexisting fault, there is little physical difference between \(U_C\) and \(U_t\) as both terms represent spatially and temporally variable frictional losses associated with fault slip; both \(U_C\) and \(U_t\) ultimately contribute to wearing and heating on the slip interface. The situation is further complicated if the dynamic friction is a nonmonotonic function of slip (e.g., Brune and Thatcher, 2002; Hirose and Shimamoto, 2005; Rivera and Kanamori, 2005; Tinti et al., 2005). While a formal distinction between the frictional and fracture losses associated with shear ruptures may be problematic, the entire amount of work spent on inelastic deformation of the host rocks during the crack propagation is unambiguous and can be readily quantified. For simplicity, here, we consider the case of a quasistatic crack growth. Formulation presented in the preceding text can be also generalized to the case of dynamic cracks.

Consider an equilibrium mode II crack in a medium subject to initial stress \(\sigma_0\). The medium is elastic everywhere except inside the crack and within a finite process zone near the crack tips (Figure 6(a)). The inelastic zone is demarcated by a surface \(S_i\). Let external forces do some work \(\delta W\) on a medium, in the

![Figure 6](image)
result of which the crack acquires a new equilibrium configuration. In the new configuration, some area ahead of the crack front undergoes inelastic yielding and joins the process zone (see an area bounded by surfaces $\Delta S_1$ and $S_t$ in Figure 6(a)). At the trailing end of the process zone, slip exceeds $D_c$, and some fraction of the process zone (bounded by surface $\Delta S_2$ in Figure 6(a)) joins the developed part of the crack. The external work $\delta W$ is spent on changes in the elastic strain energy $\delta U_e$ and irreversible inelastic deformation $\delta U_C$ (which includes friction, breakdown, and comminution):

$$\delta W = \delta U_e + \delta U_C$$ [46]

Changes in the elastic strain energy are given by (e.g., Landau and Lifshitz, 1986; Timoshenko and Goodier, 1970)

$$\delta U_e = \frac{1}{2} \left( \sigma_{ij}^1 \epsilon_{ij} - \sigma_{ij}^0 \epsilon_{ij} \right)$$ [47]

where $\sigma^1_{ij}$ and $\epsilon^0_{ij}$ are stresses and strains, respectively, in the elastic part of a medium after the crack extension. The assumption of linear elasticity [9] implies that

$$\sigma^1_{ij} \epsilon^0_{ij} = \sigma^1_{ij} \epsilon^0_{ij}$$ [48]

for any $\sigma^0_{ij}$ and $\epsilon^0_{ij}$. The identity eqn [48] allows one to write eqn [47] as follows:

$$\delta U_e = \frac{1}{2} \left( \sigma_{ij}^1 \epsilon_{ij} - \sigma_{ij}^0 \epsilon_{ij} \right) - \frac{1}{2} \left( \sigma_{ij}^0 \epsilon_{ij} + \sigma_{ij}^1 \epsilon_{ij} \right) \delta u_i$$ [49]

where $\delta u_i = u^r_i - u^d_i$ is the displacement field produced by crack propagation. Expressions [47] and [49] assume that the strains are infinitesimal, so that the relationship between strain and displacement gradients is given by eqn [8]. Also, expression [49] makes use of the fact that under quasistatic conditions, the divergence of stress is zero, $\sigma_{ij,x} = 0$ (see the equilibrium equations [10]; note that the body forces may be excluded from consideration by incorporating the effects of gravity in prestress). Using the Gauss theorem along with a condition that the crack-induced deformation must vanish at infinity, from the energy balance equation [46], one obtains the following expression for the work done on inelastic deformation:

$$\delta U_C = \delta W - \delta U_e = \int_{\Delta S_1} \sigma_{ij}^0 \epsilon_{ij} \delta u_i dS - \frac{1}{2} \int_{\Delta S_1 + \Delta S_2} \left( \sigma_{ij}^0 + \sigma_{ij}^1 \right) \epsilon_{ij} \delta u_i dS$$ [50]

In the limit of an ideally brittle fracture, the area of inelastic yielding has a negligible volume (i.e., $S_t \rightarrow 0$), so that the first integral on the right-hand side of eqn [50] vanishes. In the second integral, the surface $\Delta S_2$ also vanishes, while the surface $\Delta S_1$ becomes the crack length increment $\delta L$ (Figure 6(b)). Within $\delta L$, the stresses are weakly singular, $\sigma_{ij} \propto 1/\sqrt{r}$, where $r$ is distance to the crack tip, and the crack wall displacements scale as $\delta u_i \propto \sqrt{r}$, so that the product of stresses and displacements is of the order of unity, and the corresponding integral in eqn [50] is of the order of $\Delta S_1 \propto \delta L$. This is the well-known LEFM limit, for which the fracture energy is given by eqn [2].

For more realistic models that explicitly consider failure at the crack tip, the stresses are finite everywhere, so that the second integral on the right-hand side of eqn [50] is of the order of $\sigma_{ij}^0 \delta u_i (\Delta S_1 + \Delta S_2)$, that is, negligible compared to the integral over the finite inelastic zone $\sim O(\sigma_{ij}^0 \delta u_i S_t)$. Provided that the displacement field associated with the crack extension can be represented as $\delta u_i = \delta L \delta u_i / \delta L$, eqn [50] allows one to introduce the fracture energy $G_c$ as the total inelastic work per increment of the crack length or the energy release rate:

$$G_c = \frac{\delta U_C}{\delta L} = -\frac{1}{2} \int_{\Delta S_1} \sigma_{ij}^0 \epsilon_{ij} \delta u_i dS$$ [51]

Factor of 1/2 in eqn [51] stems from the assumption of a bilateral crack propagation. Equation [51] in general cannot be readily evaluated analytically because the size and geometry of the inelastic zone $S_t$ are not known in advance and have to be found as part of a solution. Further insights are possible for special cases. For example, assuming that all yielding is confined to a crack plane (Figure 6(b)), eqn [51] reduces to

$$G_c = \int_{\gamma} \sigma(x) \frac{\partial D(x)}{\partial L} dx$$ [52]

where we took into account that the total offset between the crack walls is $D(x) = 2u_x$, and the sense of slip is opposite to that of the resisting shear tractions acting on the crack walls. The integral equation [52] is still intractable for an arbitrary loading, as the derivative $\partial D(x)/\partial L$ must be calculated along the equilibrium curve (e.g., see eqn [31]). Closed form analytic solutions can be obtained for limiting cases of small-scale ($L \gg R$) and large-scale ($L \approx R$) yielding. First, consider a crack that is much longer than the critical size $L_c$ [36] and has a complete stress drop, $\sigma(x) = 0$ for $D > D_c$. As shown in Section 4.03.2, for such a crack the size of the process zone is independent of the crack length, $R = L_c$, and the crack propagation does not modify the slip distribution within the process zone in the reference frame of a propagating crack tip. In this case,

$$\partial D(x)/\partial L = -\partial D(x)/\partial x$$ [53]

so that eqn [52] gives rise to

$$G_c = \int_{L-R}^{L} \sigma(x) \frac{\partial D(x)}{\partial L} dx = \int_{0}^{\Delta L} \sigma(D) dD$$ [54]

Expression [54] is analogous to the result obtained using the $J$-integral technique (Rice, 1968b). An expression for the $J$-integral contains a derivative of the crack wall displacement with respect to the integration variable, rather than the crack length. As noted by Khazan and Fialko (2001), the two derivatives coincide (up to a sign) in a limiting case of a very long crack (for which the $J$-integral was derived), but the difference may be significant if the small-scale yielding approximation does not hold (also, see Rice, 1979).

For a case of a constant, but nonvanishing, residual friction, $\sigma(x) = \sigma_d$ for $D_c < D < D_m$, evaluation of integral equation [52] gives rise to

$$G_c = \sigma_d (D_m - D_c) + \int_{0}^{\Delta L} \sigma(D) dD$$ [55]

thanks to self-similarity of the along-crack displacement $D(x)$ over the interval $0 < x < L - R$ for long cracks, such that the
relationship [53] still holds. For a constant yield stress within the process zone, \(\sigma(x) = \text{const} = \sigma_s\) (e.g., Figure 4), eqn [35] gives rise to a simple expression \(G_c = (\sigma_s - \sigma_d)D_c + \sigma_d D_{\text{av}}\), in which the first and second terms can be recognized as the traditionally defined fracture energy \(U_c\) and the frictional work \(U_f\), respectively (cf. eqn [45]).

In case of a large-scale yielding, relationship [53] is generally not applicable. Assuming \(\sigma_s = \text{const}, \) expression [52] can be integrated by parts to yield

\[
G_c = \sigma_s \frac{\partial}{\partial L} \int_0^L \! D(x) \, dx - \sigma_s \int_0^L \! \frac{\partial D}{\partial L} \, dx \tag{56}
\]

In the developed part of the crack with the full stress drop \((x < F)\), slip is essentially constant and equals to \(D_c\), so that the second integral on the right-hand side of eqn [56] can be neglected. Taking advantage of expressions [33] and [35] to evaluate the first term in eqn [56], one obtains (Khasan and Fialko, 2001)

\[
G_c = \frac{\pi F}{2D_c} \sigma_s D_c \tag{57}
\]

Equation [57] indicates that the fracture energy for the case of small cracks (or large-scale yielding, \(F \ll L_c\)) is substantially different from the fracture energy for large cracks (or small-scale yielding, \(F \gg L_c\)). In case of large-scale yielding, the fracture energy is not constant even if both the yield strength \(\sigma_s\) and the critical slip-weakening displacement \(D_c\) are material constants independent of loading conditions. In particular, \(G_c\) is predicted to linearly increase with the size of the developed part of the crack \(F\). A linear scaling of fracture energy implies increases in the apparent fracture toughness \(K_o\) proportional to a square root of the developed crack length, \(K_o \propto \sqrt{F}\) (see eqn [2]), in the large-scale yielding regime. Increases in the apparent fracture toughness for small cracks are well known from laboratory studies of tensile fracture (e.g., Bazant and Planas, 1998; Ingraffea and Schmidt, 1978).

The apparent scaling of the earthquake fracture energy with the earthquake size has been inferred from seismic data (e.g., Abercrombie and Rice, 2005; Hussein, 1977; Kanamori and Heaton, 2000). Arguments presented in the preceding text indicate that several mechanisms may be responsible for the observed increases in the seismically inferred fracture energies with the rupture length, in particular, (1) off-fault damage that scales with the rupture length (larger ruptures are expected to produce broader zones of high stress near the rupture fronts, presumably advancing the extent of off-fault damage (see Figure 5 and eqn [51])); (2) a continuous degradation of dynamic friction on a fault plane, for example, due to thermal pressurization or any other slip-weakening mechanism; and (3) rupture propagation under conditions of large-scale yielding (eqn [57]), although it remains to be seen whether elastodynamic instability can occur when the process zone comprises a substantial fraction of the crack length (see Section 4.03.4). These mechanisms are not mutually exclusive and may jointly contribute to the observed scaling \(G_c \propto F\). For example, the third mechanism might be relevant for small earthquakes, while the first and second ones perhaps dominate for large events. Note that the second mechanism is ultimately limited by a complete stress drop, beyond which no further increase in fracture energy is possible. The same limit may also apply to the first mechanism, as the size of the dynamic damage zone scales with the quasistatic one for a given displacement-weakening relationship (Rice et al., 2005).

Establishing the relative importance of contributions of various mechanisms to the effective fracture energy is an important but challenging task. In particular, if contributions from the off-fault yielding are substantial, interpretations of seismic data that neglect such yielding may systematically overestimate the magnitude of the effective slip-weakening distance \(D_c\). Unfortunately, distinguishing between different contributions to the overall value of \(G_c\) can be unlikely accomplished based on the seismic data alone.

In summary, the fracture energy defined by eqn [51] is analogous to Griffith’s concept for tensile cracks, provided that the stress drop is complete \((\sigma_d = 0)\) and the small-scale yielding condition is met. For a nonvanishing friction on the crack surface, eqn [51] combines inelastic work spent against residual friction and work spent on evolving the shear stress on a fault to a residual level (i.e., the traditionally defined fracture energy). Separation between these two contributions is justified if small-scale yielding condition applies but may be ill-defined otherwise. For models with a continuous strength degradation, there is a continuous repartitioning of the energy budget, such that the effective fracture energy increases at the expense of a diminishing frictional dissipation.

4.03.8 Coupling between Elastodynamics and Shear Heating

One of the factors that can strongly affect the dynamic friction on the slipping interface and, thereby, the seismic radiation, efficiency, and stress drop is the coseismic frictional heating. Rapid slip during seismic instabilities may substantially raise temperature on a fault surface. The dependence of dynamic friction on temperature may stem from several mechanisms, including thermal pressurization by pore fluids (Lachenbruch, 1980; Mase and Smith, 1987; Sibson, 1973), frictional melting (Jeffreys, 1942; Maddock, 1986; McKenzie and Brune, 1972; Sibson, 1975), and flash heating of contact asperities (Rice, 2006). Recent experimental measurements confirm appreciable variations in the dynamic friction at slip velocities approaching the seismic range of order of a meter per second (Brown and Fialko, 2012; Di Toro et al., 2004; Goldsby and Tullis, 2002, 2011; Hirose and Shimamoto, 2003; Reches and Lockner, 2010; Spray, 2005). The documented variations are significantly larger than predictions of the rate-and-state friction (Dieterich, 1979; Ruina, 1983) extrapolated to seismic slip rates. The likely importance of the thermally induced variations in friction warrants a quantitative insight into the dynamics of fault heating during seismic slip. Major questions include the following: What are the dominant mechanisms of fault friction at high slip rates? How do increases in temperature affect the dynamic shear stress on a slipping interface? How robust is the thermally activated weakening? Is dynamic friction a monotonically decaying function of temperature? If not, what are the mechanisms, conditions, and significance of the thermally activated strengthening? When and where is the onset of the thermally induced weakening or strengthening
likely to occur on a slipping interface? And what are the implications for the dynamics of earthquake ruptures?

In a simple case of the LEFM crack with a constant residual shear stress, the frictional heating problem admits closed analytic solutions (Fialko, 2004a; Richards, 1976). Consider a mode II (plain strain) crack rupturing bilaterally at a constant speed $V_t$ (Figure 7), so that $L(t) = tV_t$. The thickness of the gouge layer that undergoes shear, and the shear strain rate across the gouge layer are assumed to be constant (Cardwell et al., 1978; Mair and Marone, 2000). Because the thickness of the gouge layer $2w$ is negligible compared with any other characteristic length scale in the problem (e.g., the rupture size $2L$ and the amount of slip $D$), the temperature evolution in the gouge layer and in the ambient rock is well described by the one-dimensional diffusion equation with a heat source:

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial y^2} + \frac{Q}{c_p}$$ [58]

where $y$ is the crack-perpendicular coordinate and $Q$ is the rate of frictional heat generation within the slipping zone,

$$Q(x, y, t) = \left\{ \begin{array}{ll} \frac{\sigma_d(x) \partial D(x, t)}{2w(x)} & , \ t > 0, |y| < w \\
0 & , \ |y| > w \end{array} \right.$$ [59]

$\partial D/\partial t$ being the local slip velocity. A solution to eqn [58] subject to the initial condition $T(x, y, 0) = T_0$, where $T_0$ is the temperature of the host rocks prior to faulting, is (Fialko, 2004a)

$$T - T_0 = \frac{1}{4c_p \sqrt{\pi w}} \exp \left( -\frac{y + w}{2 \sqrt{\kappa (t - \tau)}} \right) - \exp \left( -\frac{y - w}{2 \sqrt{\kappa (t - \tau)}} \right)$$

$$\frac{\partial D(x, \tau)}{\partial \tau} \sigma_d(x) d\tau$$ [60]

Dimensional arguments suggest the following similarity variables:

- Nondimensional along-fault coordinate $\chi = \frac{x}{L(t)}$ [61]
- Nondimensional fault thickness $\bar{\pi} = \sqrt{\frac{2}{\kappa t}}$ [62]

For the LEFM crack with a constant stress drop, the along-crack displacement profile $D(x, t)$ is self-similar in that it may be expressed in terms of a single similarity variable $\chi = \chi(x, t)$:

$$D(x, t) = L(t) \epsilon \sqrt{1 - \chi^2}, \ t > 0, |\chi| < 1$$ [63]

where $\epsilon$ is the characteristic shear strain due to the crack, $\epsilon = D(0, t)/L(t) = 2(1 - v)(\sigma_0 - \sigma_d)/\mu$. Here, $\epsilon$ is taken to be independent of $L$, as the earthquake stress drops $(\sigma_0 - \sigma_d)$ do not exhibit any scale dependence across a wide range of earthquake magnitudes (Abercrombie, 1995; Kanamori and Anderson, 1975; Scholz, 2002). The local slip rate in terms of new variables is

$$\frac{\partial D}{\partial \tau} = \frac{V_t \epsilon}{\sqrt{1 - \chi^2}}$$ [64]

Equations [60] and [64] suggest the following similarity variable for temperature:

Nondimensional temperature $\theta = \frac{T - T_0}{\bar{T}}$ [65]

where

$$\bar{T} = \frac{\sigma_d V_t \epsilon}{c_p} \sqrt{\frac{t}{\pi \kappa}}$$ [66]

is a characteristic temperature scale for frictional heating assuming a perfectly sharp fault contact.

Substituting eqn [64] into eqn [60] and making use of the similarity variables [62] and [65], one obtains the following expression for the along-crack temperature distribution in the middle of the slip zone ($\gamma = 0$):

$$\theta(\chi) = \sqrt{\frac{\pi t}{\bar{\pi}^2}} \exp \left( \frac{\bar{\pi}^2}{2 \sqrt{2(1 - \xi)}} \right) \frac{\bar{\pi} \xi}{\sqrt{\xi^2 - \chi^2}}$$ [67]

Solutions to eqn [67] are shown in Figure 8. A family of curves in Figure 8 illustrates a spatiotemporal evolution of temperature on the slipping fault surface. For faults that are

\[\text{Figure 8} \quad \text{Variations of the nondimensional excess temperature } \theta(\chi) \text{ along a mode II crack propagating at a constant rupture speed under constant frictional stress. Labels denote the nondimensional thickness of a slipping zone } \bar{\pi} \text{ (see eqn [62]).}\]

\[\text{Figure 7} \quad \text{A schematic view of a dynamically propagating mode II crack. The crack has a thickness } 2w \text{ and is rupturing bilaterally at a constant velocity } dL/dt = V_t.\]
thicker than the thermal diffusion length scale or at early stages of rupture (i.e., $\mathcal{W} > 1$), the temperature increase along the fault is proportional to the amount of slip. For thin faults or later during the rupture ($\mathcal{W} \ll 1$), the temperature is maximum near the crack tip and decreases from the tip to the crack center. However, the instantaneous temperature maximum near the crack tip does not imply cooling of the crack surface behind the tip; at any given point, the temperature on the crack surface at the tip is zero for cracks having finite thickness ($\mathcal{W} > 0$). For cracks that are much thinner than the conductive boundary layer is zero, while the slip velocity is infinite (see eqn [64]). Nonetheless, the excess temperature at the tip is zero for cracks having finite thickness ($\mathcal{W} > 0$). For cracks that are much thinner than the conductive boundary layer ($\mathcal{W} \ll 1$), the temperature field develops a shock-like structure, with the tip temperature exceeding the temperature at the crack center by about 10% (Figure 8). Assuming that the thickness of the slip zone is constant during an earthquake, eqn [67] predicts that the maximum temperatures are initially attained at the center of a cracklike shear instability. As the earthquake rupture expands, the temperature maximum may migrate toward the rupture fronts. For the thermal diffusivity of the ambient rocks $\kappa = 10^{-6}$ m$^2$ s$^{-1}$ and rupture durations of $t = 1$–10 s (corresponding to the rupture sizes of $\sim 5$–50 km), this transition will occur for faults that have thickness of the order of $\sqrt{2\kappa t} \sim 2$–5 mm or less. The critical fault thickness may be larger still if the heat removal from the fault involves some advective transport by the pressurized pore fluids, and the in situ hydraulic diffusivity exceeds the thermal diffusivity $\kappa$.

For sufficiently large ruptures, a model of a self-healing slip pulse may be a better approximation than the cracklike models (e.g., Beroza and Mikumo, 1996; Heaton, 1990; Kanamori and Anderson, 1975; Olsen et al., 1997). A self-healing mode II pulse having a constant length $L$ (Freund, 1979) generates a temperature field that is steady state in the reference frame of a moving rupture front (Fialko, 2004a). The appropriate similarity variables are

\[
\begin{align*}
\text{along-fault coordinate } & \chi = \frac{x - tV_L}{L} + 1 \\
\text{nondimensional fault thickness } & \mathcal{W} = \sqrt{\frac{2V_L}{L}} \mathcal{W} \\
\text{nondimensional temperature } & \theta = \frac{T - T_0}{T} \\
\text{dimensionless slip } & \dot{T} = \frac{\sigma \dot{\epsilon}_f}{c \rho} \sqrt{\frac{LV_L}{\pi \kappa}}
\end{align*}
\]

The coseismic displacements and the rate of slip are assumed to have the LEFM-like characteristics at the rupture front ($\chi = 1$) and a nonsingular healing at the trailing edge ($\chi = 0$):

\[
D(\chi) = L \sqrt{1 - \chi^2}
\]

\[
\frac{\partial D}{\partial \chi} = \frac{V \epsilon_f}{\sqrt{1 - \chi^2}}
\]

As before, we assume a constant dynamic friction on the slipping interface. Upon nondimensionalization using variables $[71]$, eqn [60] gives rise to the following expression for the along-fault temperature variations in the middle of the slip zone ($y = 0$):

\[
\theta(\chi) = \sqrt{\frac{\pi}{2}} \int_{\chi}^{1} \text{erf} \left( \frac{\mathcal{W}}{2\sqrt{2}(\chi - \xi)} \right) \frac{\xi d\xi}{\sqrt{1 - \xi^2}}
\]

Solutions to eqn [74] are shown in Figure 9. The near-tip structure of the temperature field due to a steady-state pulse is similar to that due to a self-similar expanding crack (cf. Figures 8 and 9). At the leading edge of an infinitesimally thin shear pulse, there is a thermal shock of amplitude $^*$ (eqn [71]). The fault temperature monotonically decreases toward the healing front, where the temperature falls to about one-half of the maximum value (Figure 9). For ‘thick’ pulses ($\mathcal{W} \gg 1$), the fault temperature increases toward the healing front proportionally to the amount of slip. For intermediate nondimensional fault thicknesses of the order of unity, the fault surface initially heats up to a maximum temperature and then cools down before the arrival of the healing front. This behavior is qualitatively different from that on a surface of an expanding crack, which indicates a progressive heating at every point along the crack as long as the rupture continues. The inferred cooling toward the healing front of the steady-state pulse for $\mathcal{W} > 5$ is caused by a decreasing heat generation due to a vanishing slip velocity and efficient removal of heat by thermal diffusion. For the characteristic rise times $L/V_L$ of the order of seconds, the steady-state LEFM pulses need to be thinner than $5 \sqrt{2\kappa L/V_L} \approx 1$ cm to experience maximum temperatures at the rupture front.

Several factors may accentuate the tendency for the temperature peaks near the rupture front. First, higher stresses in the process zone near the rupture tip imply enhanced heating. Numerical simulations indicate that the thermal effect of the
process zone can be significant even under conditions of small-scale yielding; in particular, for thin faults, the instantaneous temperature increase within the process zone is predicted to be a factor of \( \sigma_u/\tau_d \) greater than the temperature increase on the rest of the slipping interface (Fialko, 2004a). Second, in the presence of a continued dynamic weakening (e.g., Abercrombie and Rice, 2005), the frictional heating is expected to progressively decay behind the rupture front, further suppressing the excess temperature. In the context of thermal weakening, such coupling between the coseismic heating and dynamic friction may be conducive to self-sustained slip pulses (Noda et al., 2009; Perrin et al., 1995; Zheng and Rice, 1998) and to a transition to a pulse-like behavior for ruptures that initially propagate in a cracklike mode.

Numerical simulations that allow for full coupling between the shear stress, shear heating, pore fluid pressure, and elasticity indicate a rich variety of slip behaviors and important controls exerted by the effective thickness of the fault slip zone \( w \). For example, Garagash (2012) showed that in the presence of thermal pressurization, spontaneous fault slip can occur in the form of slip pulses, with the rupture speed \( V_r \) inversely correlated with the thickness of the slip zone. For the laboratory values of hydraulic and thermal properties of fault gouge, the model predicts earthquake-like behavior for slip zones having thickness of the order of centimeters (or thinner), an aseismic episodic slip with characteristics similar to those of the episodic slow slip events observed in subduction zones worldwide (Obara, 2002; Rogers and Dragert, 2003), assuming width of the slip zone of the order of 1 m (Garagash, 2012).

If thermal weakening mechanisms that operate at relatively small increases in the average fault temperature, such as thermal pressurization and flash melting (e.g., Andrews, 2002; Goldsby and Tullis, 2011; Lachenbruch, 1980; Lee and Delaney, 1987; Rice, 2006), do not give rise to substantial reductions in the dynamic friction, a continued dissipation and heating due to a localized slip will result in macroscopic melting on the slip interface and a transition from the asperity-contact friction to viscous rheology. High transient stresses associated with shearing of thin viscous films of melt have been considered as one of possible mechanisms of thermally induced strengthening (Fialko, 2004a; Koizumi et al., 2004; Tsutsumi and Shimamoto, 1997). However, recent laboratory, field, and theoretical studies suggest that transient viscous braking may not be an efficient arresting mechanism, especially in the lower part of the brittle layer (Brown and Fialko, 2012; Di Toro et al., 2006; Fialko and Khazan, 2005). If so, earthquakes that produced macroscopic melting and pseudotachylites (e.g., Davidson et al., 2003; Sibson, 1975; Swanson, 1992; Wenk et al., 2000) must have been accompanied by nearly complete stress drops. More generally, one may argue that highly localized slip zones are an indicator of nearly complete stress drops for sizeable earthquakes, regardless of whether macroscopic melting took place. This stems from the fact that melting of a narrow slip interface could have been prevented only if the dynamic friction were already low (Fialko, 2004a). Some theoretical arguments and experimental data (Brown and Fialko, 2012; Di Toro et al., 2006; Fialko and Khazan, 2005; Mitchell et al., 2013) also raise a question about the mechanism of rupture arrest below the brittle–ductile transition. A currently prevailing view is that the bottom of the seismogenic layer represents a rheological transition between the velocity-weakening friction and velocity-strengthening friction (e.g., Marone, 1998; Scholz, 1998). While this transition may well be caused by the temperature-controlled onset of stable creep (although recent laboratory data revealed a possibility of stick–slip behavior in granite at temperatures well above 300 °C, at least under dry conditions; see, e.g., Mitchell et al., 2013), it is unclear whether the velocity strengthening is relevant for the arrest of seismic ruptures that propagate into the ductile substrate from the brittle upper crust. For highly localized ruptures, the onset of melting or some other thermally activated mechanism may occur immediately behind the rupture front (aided by high ambient temperature and stress concentration), potentially overwhelming the effect of rate-and-state friction. In this case, the rupture arrest may require either delocalization of seismic slip near the rupture front (thereby limiting the temperature rise and increasing the effective fracture energy) or low deviatoric stress below the brittle–ductile transition. Discriminating between these possibilities may provide important insights into long-standing questions about the effective mechanical thickness and strength of tectonically active crust and lithosphere (England and Molnar, 1997; Jackson, 2002; Lamb, 2002; Takeuchi and Fialko, 2012).

### 4.03.9 Conclusions

Shear tractions on a slipping fault may be controlled by a multitude of physical processes that include yielding and breakdown around the propagating rupture front, rate- and slip-dependent friction, acoustic fluidization, dynamic reduction in the fault-normal stress due to dissimilar rigidities of the fault walls, thermally activated increases in the pore fluid pressure, flash heating of the asperities, and macroscopic melting. A traditional distinction between the fracture energy consumed near the rupture front and frictional losses on a well-slipped portion of a fault may be ill-defined if the slipping interface undergoes a substantial yet gradual weakening behind the rupture front. Both theoretical models and experimental data indicate that a continuous degradation of the dynamic fault strength is an expected consequence of a highly localized slip. In this case, earthquake ruptures violate a basic assumption of the LEFM, namely, that the zone of the material breakdown is small compared to the rupture size. The assumption of small-scale yielding is also violated if the extent of the off-fault damage is not negligible compared to the characteristic length of a slipping region. Field observations of kilometer-wide damage zones around major crustal faults (Ambraseys, 1970; Chen and Freymueller, 2002; Fialko, 2004b; Fialko et al., 2002) are evidence that the off-fault yielding may constitute a significant fraction of the earthquake energy balance (Wilson et al., 2004).

Both the dynamically induced variations in shear stress along the slip interface and distributed inelastic deformation off the primary slip surface affect the distribution of slip and slip rate, dynamic stress drop, and radiated seismic energy. Inferences from seismic data about constitutive properties of the fault zone material, such as the peak breakdown stress and critical slip-weakening distance, may be biased if the
to large in situ may attribute the dynamic (e.g., thermally induced) weakening to large in situ values of $D_i$ interpreted as an intrinsic rock fracture property. Constitutive relationships governing evolution of the dynamic strength at seismic slip velocities are difficult to formalize, as temperature, pore fluid pressure, and shear tractions on the fault interface are sensitive to a number of parameters such as the effective thickness of the slip zone, fluid saturation, dynamic damage and changes in the fault wall permeability, and ambient stress. Many of these parameters are also poorly constrained by the available data. Nonetheless, further insights into the dynamics of earthquakes may require fully coupled elastodynamic–thermodynamic models of seismic rupture. Development of such models is a significant challenge for future work.

Acknowledgments

I thank Jim Rice, Allan Rubin, and Hiroo Kanamori for their insightful reviews. Jim Rice kindly provided results of his calculations of off-fault damage (Figure 5). This work was supported by NSF through grants EAR-0338061, EAR-0450035, and EAR-0944336.

References

Tectonophysics 9: 143–165.