Thermo- and Hydro-mechanical Processes along Faults during Rapid Slip

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:10226780

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Thermo- and hydro-mechanical processes along faults during rapid slip

James R. Rice & Eric M. Dunham
Department of Earth and Planetary Sciences and School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Hiroyuki Noda
Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

ABSTRACT: Field observations of maturely slipped faults show a generally broad zone of damage by cracking and granulation. Nevertheless, large shear deformation, and therefore heat generation, in individual earthquakes takes place with extreme localization to a zone <1–5 mm wide within a finely granulated fault core. Relevant fault weakening processes during large crustal events are therefore likely to be thermal. Further, given the porosity of the damage zones, it seems reasonable to assume groundwater presence. It is suggested that the two primary dynamic weakening mechanisms during seismic slip, both of which are expected to be active in at least the early phases of nearly all crustal events, are then as follows: (1) Flash heating at highly stressed frictional micro-contacts, and (2) Thermal pressurization of fault-zone pore fluid. Both have characteristics which promote extreme localization of shear. Macroscopic fault melting will occur only in cases for which those processes, or others which may sometimes become active at large enough slip (e.g., thermal decomposition, silica gelation), have not sufficiently reduced heat generation and thus limited temperature rise. Spontaneous dynamic rupture modeling, using procedures that embody mechanisms (1) and (2), shows how faults can be statically strong yet dynamically weak, and operate under low overall driving stress, in a manner that generates negligible heat and meets major seismic constraints on slip, stress drop, and self-healing rupture mode.

1 INTRODUCTION

There has been a surge of activity in recent years towards increased physical realism in description of the earthquake process. That includes insightful geological characterization of the fine structure of fault zones, new laboratory experiments that reveal response properties in rapid or large slip, and new theoretical concepts for modeling dynamic rupture. The purpose here is to review some of those new perspectives and their impact on how we think about earthquake rupture dynamics.

1.1 Fault zone structure, friction and a quandary in seismology

Field observations of maturely slipped faults show a generally broad zone of damage by cracking and granulation (Chester et al., 1993), but nevertheless suggest that shear in individual earthquakes takes place with extreme localization to a long-persistent slip zone, <1–5 mm wide, within or directly bordering a finely granulated, ultracataclastic fault core (Chester and Chester, 1998; Chester et al., 2003, 2004; Heermance et al., 2003; Wibberley and Shimamoto, 2003).

On the other hand, the shear strength along a fault may be represented by

$$\tau = f \sigma$$

where

$$\sigma = \sigma_n - p_f$$

Here \( f \) is the friction coefficient, \( \sigma \) is the effective normal stress, \( \sigma_n \) is the total normal stress clamping the fault shut, and \( p_f \) is the pore pressure along it. It is well known that lab estimates of \( f \) for rocks (under sliding rates of, say, \( \mu \) m/s to mm/s) are usually high, \( f \approx 0.60–0.85 \) (e.g., Byerlee (1978)).

Given that fault slip zones seem to be so extremely thin, one must conclude that if those \( f \) prevail during seismic slip, with \( p_f \) that is much closer to hydrostatic than lithostatic, we should find the
following: (a) measurable heat outflow near major faults, and (b) evidence of extensive melting along exhumed faults. However, neither effect (a) or (b) is generally found.

1.2 Weak faults, vs. statically strong faults that dynamically weaken

There are two general lines of explanation that have been explored to resolve this quandary. One line of explanation simply postulates that major faults are weak. That could be because fault core materials are simply different from most rocks and have very low $f$, e.g., like documented for some clays and talc. Alternatively, it could be because $f$ is not necessarily low, but pore pressure $p_f$ is high and nearly lithostatic over much of the fault, especially down-dip where $\sigma_n$ is large.

It is not the purpose here to argue against such weak-fault lines of explanation, but rather to explore an alternative which we are led to by recent observations. That is that major faults are statically strong but dynamically weaken during seismic slip. Owing to the extreme thinness of slip zones, the relevant fault weakening processes during large crustal events are likely to be thermal and, given the damage zones and geologic evidence of water-rock interactions within them, it seems reasonable to assume fluid presence. Of the various dynamic weakening processes thus far identified, it has been argued that two should be singled out as being essentially universal, in that they are expected to be active and important from the start of seismic slip in crustal events (Rice, 2006; Rice and Cocco, 2007). These are as follows:

1. Flash heating and hence shear weakening of frictional micro-asperity contacts, a process which reduces $f$ in rapid slip, and
2. Thermal pressurization of pore fluid, which reduces the effective stress; $p_f$ increases, because the highly granulated fault gouge is of low permeability and the thermal expansion coefficient of water is much greater than that of the rock particles.

Other thermal weakening processes may set in at large enough slip or large enough rise in temperature $T_f$ along the fault. These include the following:

3. Thermal decomposition at large rise in $T_f$ in lithologies such as carbonates, thus liberating a fluid product phase at high pore pressure,
4. Formation of a gel-like layer at large slip in wet silica-rich fault zones, or some related process relying on the presence of silica and water, and
5. The ultimate thermal weakening mechanism, formation of a macroscopic melt layer along the fault at large enough slip and rise of $T_f$, if the above set has not limited the actual increase of $T_f$ to levels lower that that for such bulk melting.

While we focus on processes (1) and (2), it is very important to understand (3), (4), and (5) and others not yet identified. Still, in a sense the latter are secondary, because one expects that some significant earthquake slip, and fault weakening, will already have occurred before they can become activated. Preliminary estimates (Rempel and Rice, 2006) of when (5) would set in suggest that with hydrostatic $p_f$ and representative material parameter ranges, (1) and (2) are sufficiently effective at shallow fault depths that slip in significant earthquakes could often be accommodated without an onset of macroscopic melting, but that deeper in a fault zone, where the initial $\sigma_n - p_f$ (which scales the rate of heat input), and the initial $T_f$, are higher, melt onset should occur during increasing slip of typical surface-breaking earthquakes. For quantification of the slips and parameter ranges involved, see Rempel and Rice (2006).

Flash heating, weakening process (1), is a mechanism that has been advanced to explain high speed frictional weakening in metals (Bowden and Thomas, 1954; Archard, 1958/59; Ettles, 1986; Lim and Ashby, 1987; Lim et al., 1989; Molinari et al., 1999). It is only relatively recently that it has been considered as a process active during earthquake slip (Rice, 1999, 2006; Beeler and Tullis, 2003; Beeler et al., 2008; Tullis and Goldsby, 2003a,b; Hirose and Shimamoto, 2005; Noda et al., 2006; Noda, 2008). Because of the relatively low thermal conductivity of most rocks, and the relatively high shear stresses which they support at frictional micro-contacts, they are in fact susceptible to weakening by flash heating starting at sliding rates as low as 0.1 to 0.3 m/s, which is well less than the average slip rate of $\sim1$ m/s (Heaton, 1991) inferred from seismic
Thermo- and hydro-mechanical processes along faults during rapid slip

2 DYNAMIC RUPTURE FORMULATION

2.1 Elastodynamic methodology

Noda et al. (2006, 2009) and Dunham et al. (2008) have begun to integrate weakening by flash heating and thermal pressurization into elastodynamic numerical methodology for spontaneous rupture development. The problems thus far addressed are of rupture along a planar fault zone within an effectively unbounded and homogeneous solid. For those, the implementation of an elastodynamic boundary integral equation (BIE), with a spectral basis set for the slip and stress distributions (Perrin et al., 1995; Geubelle and Rice, 1995) is extremely efficient and accurate. Results have been obtained for 2D anti-plane or in-plane strain. In the formulation, with the x axis passing along the fault plane, the shear stress \( \tau(x,t) \) along the fault and the slip \( \delta(x,t) \) are related by

\[
\tau(x,t) = \tau_0(x,t) - (\mu/2c_s)V(x,t) + \phi(x,t)
\]

where \( V(x,t) = \delta(x,t)/\partial t \) is slip rate, \( \mu \) is the shear modulus, \( c_s \) is the shear wave speed, and the functional \( \phi(x,t) \) is given as a linear space-time convolution of an elastodynamic kernel, dependent on \( x - x' \) and \( t - t' \), with the slip \( \delta(x',t') \) for all \( x',t' \) within the wave cone with vertex at \( x,t \). Here \( \tau_0(x,t) \) is some specified loading stress on the fault; it is the stress that would have been induced by the applied loadings if the fault had been constrained against any slip. We prescribe \( \tau_0(x,t) \) as a uniform background stress \( \tau^b \) for all time, plus some localized overstress applied at \( t = 0 \) to nucleate rupture. \( \delta(x,t) \) and \( \phi(x,t) \) are expanded in a Fourier basis set, so the convolution is expressed by

\[
\begin{bmatrix} \delta(x,t) \\ \phi(x,t) \end{bmatrix} = \sum_{n=-N/2}^{N/2} \begin{bmatrix} D_n(t) \\ \Phi_n(t) \end{bmatrix} \exp(in\hat{k}x) \quad \text{with} \quad \Phi_n(t) = \int_0^t C_n(t-t')D_n(t')dt'
\]

Here \( N \) is a large even integer, the \( C_n(t) \) are known real functions (Perrin et al., 1995; Geubelle and Rice, 1995), \( \hat{k} = 2\pi/X \) where the periodic repeat length \( X \) of the truncated Fourier series is chosen large enough that waves from the periodic replications of the rupture event do not arrive to neighboring replications in the time of interest, and the \( D_{-n}(t) \) are complex conjugates of the \( D_n(t) \).
with $D_\theta$ and $D_{N,2}$ being real. Eq. (3) is equivalent to a real Fourier cosine and sine series truncated at N terms (with no sine term when $n = N/2$). Through FFT procedures, the histories $D_n(t)$ are determined by $\delta(x,t)$ at N sample points, equally spaced by $\Delta x = X/N$, i.e., by the histories $\delta(j\Delta x - X/2,t)$ for $j = 0, 1, 2, \ldots, N - 1$, and conversely. That greatly speeds calculations. The $\Phi_n(t)$ and $\phi(j\Delta x - X/2,t)$ are similarly related.

2.2 Friction law with weakening by flash heating and thermal pressurization

The simple flash heating model reviewed here (Rice, 1999, 2006; Beeler and Tullis, 2003; Beeler et al., 2008) was intended to approximately determine the expression for $f$ of eq. (1) for conditions of sustained sliding at some speed $V$. The $f$ so derived must be regarded as a steady state value, written here as $f_{ss}$ and is regarded as a function of slip rate, $f_{ss} = f_{ss}(V)$, although it is also a function of the spatially averaged (over patches of fault area large enough to include many contacts) temperature $T_f$ of the fault plane, which evolves with ongoing slip and time. In the model (Rice, 2006) it is assumed that contact temperature $T_c = T_f$ when a contact pair first forms, but then as the contact slides during its brief lifetime (which is $D/V$ for a contact asperity of diameter $D$; see Fig. 1), $T_c$ rises substantially above $T_f$. The rise is due to the intense localized heat generation at the contact, at rate $\tau, V$, where $\tau_c$ is the contact shear strength (typically of order 0.1 times shear modulus $\mu$ at low $T_c$; see discussion in Rice (2006)). $\tau_c$ is assumed to have negligible variation as $T_c$ increases, but then to abruptly decrease to a much lower weakened value $\tau_w$ when $T_c$ reaches a “weakening” temperature $T_w$. Within the model, based on 1D heat conduction at the sliding contact like in Archard (1958/59), the slip rate such that an asperity of diameter $D$ would begin to weaken only just as it is slid out of existence is then (Rice, 1999, 2006)

$$V_w = (\pi a_{th}/D)(\rho c(T_w - T_f)/\tau_c)^2$$  (4)

Here $\rho c$ is volumetric specific heat and $a_{th} = K/\rho c$ is thermal diffusivity ($K$ is thermal conductivity). Estimates based on rock properties and assumed $D$ of order 10 $\mu$m, as well as comparison to experimental results of Tullis and Goldsby (2003a, b) and Beeler et al. (2008) (by rotary shear of a rock annulus in an Instron frame), suggest that $V_w$ is of order 0.1 to 0.3 m/s for rocks such as quartzite, feldspar, granite and gabbro when $T_f = $ room temperature. $V_w$ is expected to be less at higher $T_f$. Thus the model takes contacts to be strong for all their lifetime at low slip rates, $V < V_w$, but to be strong for only a fraction of their lifetimes at high slip rates, $V > V_w$, that fraction being $V_{w}/V$.

Neglecting the actual statistical distribution of contact diameters, and taking $D$ as a representative value, these concepts lead to the steady-state friction coefficient

$$f_{ss}(V) = \begin{cases} f_w + (f_0 - f_w)V/V & V \leq V_w \\ f_0 & V \geq V_w \end{cases}$$  (5)

where $f_0$ is the low-speed friction coefficient and $f_w = f_0\tau_w/\tau_c$ is the value to which it would reduce if all contacts were in the weakened state. Note that despite the notation $f_{ss}(V)$, the dependence

Figure 1. Simple representation of asperity contacts and their strength loss for flash heating model.
Thermo- and hydro-mechanical processes along faults during rapid slip

mentioned on ambient fault temperature enters from the dependence of \( V_w \) on \( T_f \), eq. (4). Rough estimates from the Tullis and Goldsby (2003a, b) experiments, which covered a limited velocity range, up to slightly less than \( V = 0.4 \) m/s, are that \( f_0 \approx 0.64 \) and \( f_w \approx 0.12 \) for quartzite, \( f_0 \approx 0.82 \) and \( f_w \approx 0.13 \) for granite, and \( f_0 \approx 0.88 \) and \( f_w \approx 0.15 \) for gabbro. However, the \( f_w \) involve significant extrapolation and, in experiments of Yuan and Prakash (2008) on quartzite from the same source and annular configuration, \( f \) was only very slightly below 0.20 at slip rates as high as 2–4 m/s. Eq. (5) and the small \( V_w \sim 0.1 \) m/s for quartzite, suggest that \( f \) should nearly coincide with \( f_w \) at such rates, thus that \( f_w \approx 0.18–0.20 \).

In fact, we cannot simply assume \( f \) to be a decreasing function of \( V \) in eq. (1) because that makes the problem of sliding between elastic continua ill-posed (there is a short wavelength divergence in response to small initial perturbations from steady sliding). The problem is remedied (Rice et al., 2001) mathematically when we look to experiments and embed the description of variations of \( f \) in rate and state friction concepts. Thus, with a “slip” version of state evolution, \( f \) is assumed to be given by the form (Rice, 1983)

\[
\frac{df}{dt} = \frac{a}{V} \frac{dV}{dt} - \frac{V}{L}(f - f_{ss}(V))
\]

(6)

where \( a \) is the direct effect coefficient in rate and state modeling (\( a \approx 0.01 \) at room \( T \) for quartzite and granite, and it scales in approximate proportion to absolute \( T \)), and \( L \) is a slip distance adequate to renew the asperity contact population, typically taken as 5–20 \( \mu \)m in our studies, as guided by observed state evolution slip distances in rate and state friction experiments. The smallness of \( L \) makes the simulations extremely challenging in terms of present-day computers and, as of the writing, the longest rupture lengths simulated are \( \sim 30 \) m (Dunham et al., 2008; Noda et al., 2009).

In the numerical simulations (Noda et al., 2009; Dunham et al., 2008) we replace the constant term \( f_0 \) in eq. (5) with \( f_{LV}(V) \), which is the weak logarithmic function of \( V \) describing slow-rate friction at steady-state. That is, \( f_{LV}(V) = f_0 - \left( b - a \right) \ln(V/V_0) \) in the standard notation, where \( b - a \approx 0.002–0.004 \) for rate-weakening frictional surfaces in granite. Here \( V_0 \) can be chosen arbitrarily in that low-speed regime (say, 1 \( \mu \)m/s) with corresponding adjustment of \( f_0 \) to correspond to the correct \( f_{ss} \) at that rate. Fig. 2, taken from Noda et al. (2009), plots the resulting \( f_{ss} \) using parameters which are thought to be representative for granite, at conditions corresponding

![Figure 2](image-url)

Figure 2. From Noda, Dunham and Rice (2008), their figure 1. Plot of \( f_{ss} \) based on Tullis and Goldsby (2003a, b) parameters for granite, adjusted to conditions at mid-seismogenic depth, \( \sim 7 \) km, for crustal strike slip earthquakes. Plot is based on assumed ambient effective normal stress \( \sigma_0 = 126 \) MPa and temperature \( T_0 = 210^\circ \)C; as slip develops, heating increases \( T_f \) over \( T_0 \), and pore fluid pressurization will reduce \( \sigma \) from \( \sigma_0 \). The Zheng and Rice (1998) stress level \( \tau_{\text{pulse}} \), important to understanding whether rupture takes the self-healing pulse or classical shear crack mode, is also shown.
to a mid-seismogenic zone depth for crustal strike-slip earthquakes. Also, while irrelevant in our simulations, for some purposes the $a/V$ term in eq. (6) should be regularized near $V = 0$ in a manner consistent with the thermal activation basis for the direct effect (Rice et al., 2001).

2.3 Inclusion of thermal pressurization of pore fluid in friction formulation

In order to close our system of governing equations which, so far, consists of eqs. (1), (2), (3) and (6), it is necessary to relate the histories of $p_f$ and $T_f$ to that of slip and stressing along the fault. That is because $p_f$ directly enters eq. (1) and $T_f$ is a parameter on which $f_a(V)$ (see eqs. (5) and (4)). The new ingredients which close the system are the equations of conservation of energy, with inclusion of conductive heat transfer, and of conservation of fluid mass with inclusion of Darcy fluid seepage and poro-thermo-elastic considerations. These amount to a pair of coupled PDEs for the fields of pore pressure $p$ and temperature $T$ near and on the fault. The values of $p$ and $T$ thus determined at the fault surface itself are the respective $p_f$ and $T_f$ that we seek.

In writing the conservation laws we neglect certain apparently minor terms (e.g., advective heat transfer by moving fluid), and recognize that the gradients of pore pressure $p$ and temperature $T$ very near the fault are generally very much larger in the direction perpendicular to the fault (the $z$ direction) than in the $x$ or $y$ directions which are parallel to it (such may sometimes not be a valid assumption immediately at the moving rupture tip). Thus (e.g., Rice (2006)) we have

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \left( \rho c \beta \frac{\partial T}{\partial z} \right) + \frac{\partial p_{pl}}{\partial t}$$

Here $\dot{\gamma}^p$ is the inelastic fault-parallel shear strain rate and $\dot{p}_{pl}$ is the inelastic dilatancy rate ($n$ itself is the volume of pore space per unit aggregate volume of porous material, that aggregate volume being measure in some reference state before the deformation episode considered). Also, $\rho c$ is the specific heat per unit volume, $\alpha_{sh} = K/\rho c$ is the thermal diffusivity, and $K$ is thermal conductivity; $\beta$ is a porous medium storage coefficient under the particular mechanical constraints near a fault zone (see Rice (2006)), $\alpha_{sh} = k/\beta \eta_f$ is the hydraulic diffusivity, $k$ is permeability, and $\eta_f$ is viscosity of the pore fluid; and $\Lambda$ is a parameter representing $dp/dT$ due to heating under undrained, elastically reversible conditions. Rice (2006) and Rempel and Rice (2006) compile estimates of these various parameters at mid-seismogenic depths in the crust, based principally on data of Wilberley (2002) and Wilberley and Shimamoto (2003) for gouge of the Median Tectonic Line Fault (Japan) under a range of confining stresses, and on tabulated thermophysical data for water and minerals.

Rempel and Rice (2006) also compare fully non-linear solutions of eqs. (7) to the linearized versions

$$\frac{\partial T}{\partial t} = \alpha_{sh} \frac{\partial^2 T}{\partial z^2} + \frac{\tau}{\rho c} \frac{\partial \dot{\gamma}^p}{\partial t}$$

(with coefficients $\rho c$, $\alpha_{sh}$, $\beta$ and $\alpha_{sh}$ considered constant) for cases with a priori specified histories of $\partial \dot{\gamma}^p/\partial t$ and $\partial \dot{p}_{pl}/\partial t$, e.g., representing earthquake slip at a specified constant rate in time, with the aim of estimating (using eq. (1) but assuming a constant low $f$ as motivated by flash heating) the relation between stress $\tau$ and slip $\delta$ during seismic rupture. They find that the procedure adopted by Rice (2006), of iteratively choosing the constant values of coefficients in the linearized PDEs of eqs. (8), as certain path averages in $p$, $T$ space of those same coefficients, when regarded as known functions of $p$, $T$, along the $p$, $T$ path predicted by eqs. (8), gives a tolerable match to results for $\tau$ versus $\delta$ based the full non-linear solutions to eqs. (7). In our spontaneous elastodynamic analyses presented here we have used eqs. (8), without that iterative choice of coefficients, but rather with coefficients based on the ambient $p$, $T$ at the mid-seismogenic depth ($\sim 7$ km) considered.

The conceptually (but not computationally) simplest case for eqs. (7) or (8) is that of slip on a mathematical plane. In that case $\partial \dot{\gamma}^p/\partial t = V(x, t) \delta_{Dir}(z)$, where $V$ is the local slip rate and $\delta_{Dir}(z)$ is the Dirac function. However, actual shear zones, when highly localized to a prominent slip...
surface, can nevertheless distribute deformation over regions that may extend over a few 10s to a few 100s of μm. Because the difference between such small but finite thicknesses and zero thickness is sometimes not negligible on the seismic time scale in our modeling, a Gaussian distribution of shear heating $\tau \frac{\partial \gamma^{pl}}{\partial t}$, like in Andrews (2002), over a zone of nominal thickness $2w$ is assumed. We write

$$\tau \frac{\partial \gamma^{pl}}{\partial t} = \tau(x,t)V(x,t)\frac{z^2}{2\pi w} \exp\left(-\frac{z^2}{2w^2}\right)$$

(9)

which reduces to $\tau \frac{\partial \gamma^{pl}}{\partial t} = \tau(x,t)V(x,t)\delta_{Dir}(z)$ as $w \to 0$.

In the numerical studies (Noda et al., 2008; Dunham et al., 2008) of spontaneous rupture based on this formulation (which have thus far taken $n^{pl} = 0$), explicit finite difference (in $z$ and $t$) versions of eqs. (8) are solved at each elastodynamic gridpoint location $x_j = -X/2 + jX/N$ along the rupture, ultimately to give, with help of the other governing equations, $T_f(x_j,t)$ and $p_f(x_j,t)$. For what we think to be appropriate ranges $\alpha_{th} \sim 1$ mm$^2$/s and $\alpha_{hy} \sim 1$–10 mm$^2$/s (Rice, 2006; Rempel & Rice, 2006), it turns out that when the elastodynamic time steps $\Delta t$ are already short enough to resolve the state evolution of eq. (6) with $L \sim 5–20$ μm, explicit finite difference solution of eqs. (8) requires significantly shorter time steps, to meet the requirement is that diffusion grid spacing be sufficiently small that the error associated with the spatial discretization of the diffusion equations is comparable the error in the elastodynamic system. Thus Noda et al. (2008) devised a procedure based on a quadratic interpolant of $\phi(x,t)$ within an elastodynamic time step $\Delta t$, to achieve second-order accuracy. The interpolant is constructed from $\phi(x,t-\Delta t)$, $\phi(x,t)$, and $\phi(x,t+\Delta t)$ for use between $t$ and $t+\Delta t$, so that the condition of eq. (2), $\tau = \tau_0 - (\mu/2c_s)\dot{V} + \phi$, along with eq. (1), eq. (6) with eqs. (4) and (5) (with $f_0$ replaced by $fLV(V)$), and eqs. (8) are satisfied, to numerical accuracy, within each of the smaller time steps for diffusion. This is highly accurate but very demanding computationally because of the small but presumably realistic $L \sim 5–20$ μm and realistic diffusivities, $\sim 1$–10 mm$^2$/s, that we use.

2.4 Theoretical background on strong rate-weakening and self-healing slip pulses

The adopted friction description involves strong rate-weakening (see Fig. 2). To provide background for understanding when strong rate-weakening will lead to rupture in the mode of a self-healing slip pulse, versus a classical enlarging shear crack, we digress here to review results from Zheng and Rice (1998), in part following a recapitulation in Rice (2001), discussed with reference to Fig. 3. Some representative results from Dunham et al. (2008) and Noda et al. (2009), based on the formulation in the earlier parts of this Section 2, are presented in Section 3 to follow.

Studies by Cochard and Madariaga (1994, 1996), Perrin et al. (1995), and Beeler and Tullis (1996) had shown that strong rate-weakening could lead to rupture in the self-healing mode. The

![Figure 3. Friction with strong rate dependence. It is assumed here, to simplify, that the steady-state friction $f_{ss}(V)$ is a function of slip rate $V$ only, and that the effective normal stress $\sigma$ is constant; then the steady-state shear stress $\tau_{ss}(V) = f_{ss}(V)\sigma$. (Modified from Rice (2001)).](image-url)
objective in Zheng and Rice (1998) was then to establish guidelines for when this type of rupture mode would in fact result. To simplify enough that sharp results could be established, it was assumed in that development that the steady state shear strength \( \tau_{ss}(V) \) was a function only of slip rate \( V \), which could, e.g., result if \( \overline{\sigma} \) is constant and \( f_{ss}(V) \) is a function only of \( V \). That, \( \tau_{ss}(V) = f_{ss}(V)\overline{\sigma} \). However \( \tau_{ss}(V) \) does not give the expression for \( \tau \) for variable \( V \), because \( \tau = \overline{f}\overline{\sigma} \) and \( f \) evolves according to eq. (6). Thus the response to a sudden change in \( V \) is as depicted on the left in Fig. 3, a response which also assures well-posedness in problems of frictional sliding between elastic continua (Rice et al., 2001).

Now consider a fault surface which we treat as the boundary \( z = 0 \) between two identical half spaces (Fig. 3, at right). An initial shear stress \( \tau_0(x, y) = \tau_{0b}^b \), a constant level too small to cause failure, acts everywhere on \( S_\infty \) (the entire \( x, y \) plane) except in small nucleation region \( S_{out} \) which will be overstressed to start the rupture. The stress level \( \tau_{pulse} \) (the same as what Noda et al. (2009) denoted \( \tau_{pulse} \) in their diagram shown here in Fig. 2) is marked in Fig. 3. \( \tau_{pulse} \) is defined as the largest value that \( \tau_0^b \) could have and still satisfy \( \tau_0^b - \mu V/2c_s \leq \tau_{ss}(V) \) for all \( V > 0 \).

Suppose that \( \tau_0^b < \tau_{pulse} \). As will now be seen, that effectively precludes the possibility that rupture could occur on \( S_\infty \) in the form of an indefinitely expanding shear crack. Note that

\[
\tau_0^b < \tau_{pulse} \quad \text{implies that} \quad \tau_{ss}(V) - (\tau_0^b - \mu V/2c_s) > 0 \quad \text{for all} \quad V > 0. \tag{10}
\]

Use is made of an elastodynamic conservation theorem (Zheng and Rice, 1998)

\[
\int \int_{S_\infty} [\tau(x, y, t) - \tau_0(x, y) + \mu V(x, y, t)/2c_s] \, dx \, dy = 0 \tag{11}
\]

which holds throughout the rupture; \( \tau_0(x, y) \) is the stress field at \( t = 0 \) when rupture is nucleated.

Assume that, with \( \tau_0^b < \tau_{pulse} \), rupture has been locally nucleated and grows on \( S_\infty \) in the form of an indefinitely expanding shear crack. Such an assumption can be shown, as follows, to lead to a definite contradiction in the case of mode III (anti-plane) slip, and to seemingly implausible result in general, meaning that we must reject the assumption that an indefinitely expanding crack-like rupture is possible when \( \tau_0^b < \tau_{pulse} \). To see why, note that the integrand in eq. (11) everywhere along the rupturing surface \( S_{rupt}(t) \) where \( \tau \approx \tau_{ss}(V) \), except for \( S_{nuc} \) and for small regions at the rupture front affected by the rate/state regularization (so that \( \tau \) may depart significantly from \( \tau_{ss}(V) \) in those regions; they are small because \( L \) is in the range of a few to a few tens of microns; see discussions in Zheng and Rice (1998) and Noda et al. (2009) for further quantification), is equal to

\[
\tau_{ss}(V) - \tau_0^b + \mu V/2c_s = \tau_{ss}(V) - (\tau_0^b - \mu V/2c_s) > 0 \tag{12}
\]

where the inequality follows from eq. (10). Thus, denoting by \( S_{out}(t) = S_\infty - S_{rupt}(t) \) the region of \( S_\infty \) lying outside the rupture at time \( t \), and noting that \( V = 0 \) there, we must by eq. (11) then have, for any sufficiently large rupture,\n
\[
\Delta F_{out}(t) = \int \int_{S_{out}(t)} [\tau(x, y, t) - \tau_0^b] \, dx \, dy < 0 \quad \text{(and} \quad \Delta F_{out}(t) \rightarrow -\infty \text{as} \quad S_{nuc}(t) \rightarrow \infty) \tag{13}
\]

where \( \Delta F_{out}(t) \) is the change in total shear force (positive in the direction of initial shear stressing) supported outside the ruptured zone. Inequality (13), however seems implausible: We expect ruptures to result in an increase \( \Delta F_{out}(t) \) in the net force carried outside themselves, or at least to not decrease the force, and in the 2D case of mode III rupture it is provably the case that \( \Delta F_{out}(t) \) can never correspond to a decrease.

We therefore conclude that an indefinitely expanding rupture in crack-like mode cannot occur if \( \tau_0^b < \tau_{pulse} \). And that any ruptures which do then occur would have to be of another type for which not all of the ruptured surface was slipping for all time. i.e., they would have to be of the
Thermo- and hydro-mechanical processes along faults during rapid slip

self-healing type. That is consistent with a range of calculations for different stress levels and forms of \( \tau_\text{ss} V \); crack-like ruptures are not found, only self-healing slip pulses, when \( \tau_\text{b} < \tau_\text{pulse} \).

The guidelines just outlined on when self-healing versus crack-like ruptures occur are also consistent with laboratory studies in which both types of ruptures were generated in different conditions (Lykotrafitis et al., 2006; Lu et al., 2007).

Thus, imagine a fault on which tectonic stress is being increased slowly, and on which, at various times, ruptures nucleate from localized patches of the fault where the ratio \( \tau/(\sigma_\text{n} - p_\text{f}) \) is high (consistent with the fault material being “statically strong”). Such nucleating regions could occur because \( \tau \) is locally high from a stress concentration at a boundary between creeping and locked lithologies, or because \( p_\text{f} \) is locally high, or because \( \sigma_\text{n} \) is low due, e.g., to a dilational twist in the fault trace or to an intersection with a secondary fault that slipped in the past and locally reduced \( \sigma_\text{n} \). In general, when the average stress on the fault is low, we expect that none of these events will lead to a rupture that spreads over the entire fault. But ultimately, as average stress on the fault is gradually increased, one may do so. The first type of rupture encountered which can propagate without limit, once nucleated, will inevitably be of the self-healing type. Thus the presently reviewed results argue strongly for an association of strong rate-weakening on major tectonic faults with the occurrence of large earthquakes on those faults in the self-healing rupture mode (Lapusta and Rice, 2003), a mode which is generally thought consistent with seismic observations. A relevant quote from Zheng and Rice (1998) is as follows:

“The Gutenberg-Richter frequency versus size statistics of earthquakes tells us that for every 1 earthquake that achieves, say, magnitude 3 size but does not arrest and, rather, grows to magnitude 4, there are approximately 10 that do arrest at that smaller size, and so on for other magnitudes, at least within the range for which there is power law scaling with \( b \approx 1 \). An interpretation of such results is to say that faults are chronically understressed, so that most ruptures simply fail to become large. Such understressing is likely to be very heterogeneous and not like the uniform \( \tau_\text{b} \) considered here. Nevertheless, if natural faults are indeed velocity weakening at seismic slip rates, so that our present analysis applies, then it is plausible to make the association that these faults are lightly stressed and perhaps understressed in the precise meaning of the term here. In such case, then, we could understand that the self-healing mode of rupture would be a pervasive one, because the stresses are too low on average to allow the cracklike mode and can do so only at places of local stress concentration where rupture nucleates.”

They also add the caution that strong rate-weakening is not the only feasible explanation of self-healing ruptures:

“We emphasize that this study has been on understanding the rupture mode in the presence of velocity-weakening friction on a fault of spatially uniform properties between identical linear elastic solids. As explained in the Introduction, other mechanisms of self-healing pulse generation exist and involve, for example, strong spatial nonuniformity of frictional weakening properties within the fault zone or dissimilarity of elastic properties across the fault.”

3 SUMMARY OF SOME REPRESENTATIVE RESULTS AND THEIR IMPLICATIONS

3.1 Rupture modes, phase boundaries, and (low) background stress levels to sustain rupture

Figures 4 and 5 show results from Noda et al. (2009), for the same model based on granite properties at high slip rates as shown here in Fig. 2. The parameter \( r \) in Figs. 4 and 5, and in Fig. 6 next, measures how much we assume that damage induced by the high off-fault and fault-parallel stresses and strains very near the rupture tip (beyond the elastic range) alter the near-fault poromechanical properties. Parameter \( r = 0 \) corresponds to poromechanical properties of intact fault gouge at temperature and effective stress conditions corresponding to 7 km depth (mid-seismogenic depth), as estimated from various sources by Rice (2006). Parameter \( r = 1 \) corresponds to the “highly damaged” fault walls case of Rice (2006), for which it was estimated, rather arbitrarily, that damage at the passing rupture front causes the near-fault materials to have a ten-fold increase in permeability and a doubling of a poroelastic compressibility measure.
Figure 4. From Noda, Dunham and Rice (2009), their figure 4; the figure numbers printed on the diagrams correspond to figures which follow. (a) Rupture mode identified as a function of background stress and width of the shear zone, for a given set of hydro-thermal properties. (b) Rupture mode identified as a function of background stress and hydro-thermal properties (parameterized by \( r \); see text) for a given width of the shear zone. The horizontal bar at the top of (a) corresponds to a case for which \( T \) and \( p \) variations are neglected so that the theory of Section 2.4 applies with \( \tau_{\text{pulse}} \) shown as the dashed vertical line. The solid lines in (a) and (b) are a plots of a parameter devised by Noda et al. (2009) that approximately generalizes the \( \tau_{\text{pulse}} \) concept to the realistic cases with variable \( T \) and \( p \).

Figure 5. From Noda, Dunham and Rice (2009), their figure 2. Illustration of the three rupture modes, (a) enlarging shear crack, (b) growing self-healing pulse, and (c) arresting self-healing pulse. Initial development of slip (vs. distance along fault at successive times, half of fault shown) during dynamic rupture, nucleated by localized overstressing, for different cases from Fig. 4, all at the same background stress level but with different widths of the shear zone.

Figure 6 based on Noda et al. (2009) shows results for the longest rupture it has thus far (as of April 2009) been possible to simulate. A companion case of rupture of the same length, showing crack-like rupture, has also been done; it shows much greater slip at a given rupture length.

### 3.2 Some implications

The results are preliminary, but due to current day computational limitations it will likely be much time before substantially longer ruptures can be treated routinely. At least that is so in studies following our aim to strictly adhere to parameters (including friction state transition length \( L \), shear zone thickness \( 2w \), and those characterizing extreme rate-weakening at rapid slip rates) which we think represent reality as shown in the laboratory and in geological characterizations of natural faults.

Nevertheless, our results show definitively that faults modeled in that way can operate at low overall shear stress levels compared to Byerlee friction thresholds, yet produce runaway ruptures. For those ruptures, Byerlee-like conditions are achieved fleetingly at the moving rupture front, before dynamic weakening takes command.

Our results are in general consistency with the low overall fault stress levels inferred in tectonic modeling of strike slip faults in California (e.g., Bird and Kong (1994)). We typically find that
 faults can begin to sustain runaway ruptures at background stress levels corresponding to \( \tau/\bar{\sigma}_0 \approx 0.20–0.25 \). That compares quite favorably to the estimate of Hickman and Zoback (2004), by direct stress measurements in the SAFOD pilot borehole near the San Andreas Fault at Parkfield, that at the deepest location measured, 2.1 km depth, a ratio \( \tau/\bar{\sigma}_0 \approx 0.2–0.3 \) is inferred for the nearby San Andreas Fault.

As Fig. 6 shows, the results out to 30 km overall rupture length suggest an approximate self-similarity (like in a simpler self-healing pulse model by Nielsen and Madariaga (2003), and as conjectured from natural observations by Manighetti et al. (2005)). If we optimistically assume that such scaling would extend approximately to the multi-km dimensions of large surface-breaking ruptures, then we would predict a ratio of slip to rupture length of order 14 cm/km, which is quite reasonable in terms of earthquake phenomenology. Also, while the shear stress on the fault in Fig. 6 is low as slip develops (high dynamic stress drop), the final static stress drop is \( \sim 3 \) MPa, which is quite in the range of such stress drops for crustal earthquakes. The stress on the fault can be seen in fig. 6 to gradually decrease as slip accumulates. That is essentially all due to thermal pressurization (which would seem to break strict self-similarity), but it means that for much longer ruptures, with correspondingly greater slip, that the average ratio \( \tau/\bar{\sigma}_0 \) during sliding in large events is expected to be in a range of \( \sim 0.1 \) or less, which will not produce a perceptible heat-flow anomaly. And of course, our prediction, for reasons elaborated in Section 2.4, is that large ruptures will normally be in a self-healing mode, given the strong rate-weakening of friction.

While only results from laboratory rock observations and geological characterizations of fault zones were used to construct the model, the resulting dynamic rupture predictions made by it have no obvious inconsistencies with what can be learned from seismology.

ACKNOWLEDGMENTS

The study was supported at Harvard University by NSF- EAR Award 0510193 and by the Southern California Earthquake Center as funded by Cooperative Agreements NSF EAR-0106924 and USGS 02HQAG0008 (the SCEC contribution number for this paper is 1276).
REFERENCES


Chester, F.M., and J.S. Chester (1998), Ultracataclasite structure and friction processes of the Punchbowl fault, San Andreas system, California, Tectophys., 295, 199–221.


