Dynamic fault weakening and strengthening by gouge compaction and dilatancy in a fluid-saturated fault zone

Evan Hirakawa1,2,3 and Shuo Ma1

1Department of Geological Sciences, San Diego State University, San Diego, California, USA, 2Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA, 3Now at Lawrence Livermore National Laboratory

Abstract Fault gouge deformation likely plays a significant role in controlling the strength of mature, large-displacement faults. Experiments show that intact gouge deforms in an overall ductile and stable manner, readily compacting, but dilates and experiences brittle failure under large strain rate. Inelastic gouge compaction and dilatancy are modeled here using a combined Mohr-Coulomb and end-cap yield criterion in a dynamic rupture model of a strike-slip fault with strongly velocity-weakening friction. We show that large shear stress concentration ahead of the rupture associated with the rupture front causes the gouge layer to compact (e.g., by structural collapse and comminution), leading to rapidly elevated pore pressure and significant weakening of the principal fault surface. Shortly after the rupture front passes, strong dilatancy during strength drop and rapid sliding reduces pore pressure and strengthens the fault, promoting slip pulses. Large strain localization in the gouge layer occurs as a result of rapid gouge dilatancy and strain softening. The combination of prerupture weakening from compaction and restrengthening from dilatancy hardening leads to a smaller-strength drop, and limits the stress concentration outside the gouge layer. This leads to a reduction of inelastic shear strain in the damage zone, which is more consistent with geological observations and high-speed frictional experiments. With the presence of well-developed fault gouge, the strength of mature faults may be limited by end-cap rather than Mohr-Coulomb failure; thus, their frictional strengths are significantly smaller than Byerlee friction.

1. Introduction

Resolving the strength of mature, large-displacement faults has been a longstanding problem in tectonics and earthquake mechanics. Multiple pieces of evidence suggest that these faults may operate at low shear stresses. Failure to detect a heat flow anomaly across the San Andreas Fault (SAF) implies that the dynamic frictional coefficient during past earthquake ruptures is only 0.1–0.2 or less [Brune et al., 1969; Lachenbruch and Sass, 1980]. Regional stress measurements also suggest that maximum compressive stresses are at a high angle to the SAF fault strike, indicating low shear stress on the fault [e.g., Zoback et al., 1987; Townend and Zoback, 2004]. In addition, a general absence of frictional melt products (pseudotachylytes) in the fault core implies relatively low frictional heating [Sibson, 1975, 2003]. Measurements at other tectonic settings seem to corroborate this view that mature faults are weak [e.g., Suppe, 2007; Fulton et al., 2013; Ujiie et al., 2013; Gao and Wang, 2014].

Yet laboratory studies have shown that for a wide variety of rock types, static friction of rocks is consistently 0.6–0.85, a result known as Byerlee’s law [Byerlee, 1978]. Regional stress measurements from a variety of tectonic settings suggest that the crust is critically stressed to incipient failure on preexisting microcracks with friction values consistent with Byerlee’s law and hydrostatic pore pressure [Townend and Zoback, 2000]. The inconsistency between the inferred low friction on mature faults and Byerlee’s law is the well-known “stress–heat flow paradox.” Current debate on this topic can be found in Scholz [2006] and Townend [2006]. Noda et al. [2009], Dunham et al. [2011a], and Fang and Dunham [2013] also thoroughly discussed this topic.

In this work we focus on one leading hypothesis to address the heat flow paradox: the statically strong and dynamically weak fault hypothesis [e.g., Lapusta and Rice, 2003; Rice, 2006; Noda et al., 2009; Dunham et al., 2011a, 2011b], in which both Byerlee’s law and heat flow constraints are satisfied. Low dynamic strength is well supported by high-speed frictional experiments [e.g., Di Toro et al., 2011], which demonstrate that rock friction plummets to extremely low values (~0.1) at coseismic speeds. Among various dynamic weakening mechanisms, flash heating [e.g., Goldsby and Tullis, 2011] and thermal pressurization [Sibson, 1973;
One possible shortcoming of most dynamic weakening mechanisms and the above hypothesis is that slip is required for the mechanisms to take effect and static friction cannot be reduced [Scholz, 2006]. If Byerlee’s law governs static friction, and dynamic friction is governed by dynamic weakening mechanisms, large strength drops (static friction minus dynamic friction) on the order of 100 MPa are inevitable at seismogenic depths. It may be questionable that the fault can sustain such large strength drops. In the model by Noda et al. [2009] that incorporates ash heating and thermal pressurization, large strength drop partly caused slip velocities greater than 300 m/s and fault-parallel strain of around ~0.1. Off-fault brittle shear failure can restrict these values toward those that are characteristic of earthquake ruptures [Dunham et al., 2011a]. However, the inelastic off-fault strain in these models may be oversimplified, as we will discuss below.

In view of the realization that large off-fault stress concentration associated with rupture propagation can cause material failure [e.g., Andrews, 1976; Poliakov et al., 2002; Rice et al., 2005], dynamic rupture models incorporating brittle off-fault shear failure have become common [e.g., Andrews, 2005; Templeton and Rice, 2008; Viesca et al., 2008; Duan and Day, 2008; Ma, 2008; Ma and Andrews, 2010; Dunham et al., 2011a, 2011b; Gabriel et al., 2013; Shi and Day, 2013; Xu and Ben-Zion, 2013; Kang and Duan, 2014] and are supported by geologic observations of complex fault zone architecture formed as a result of irreversible damage [e.g., Chester et al., 1993; Rempe et al., 2013]. These models have greatly advanced our understanding of coseismic damage generation and their resulting effects on rupture dynamics. The presence of well-developed fault gouge for mature faults, however, was not considered in these models.

Fault zone geology consists of a gradational deformation trend, where host rock has been subjected to irreversible damage with progressively increasing intensity toward the main fault trace (Figure 1). This largely comprises a region of mesoscopic brittle deformation that forms the outermost unit of fault zones known as the “damage zone” (~tens to hundreds of meters thick), but most relevant damage is localized to the highly deformed unit known as the “fault core,” which consists primarily of fault gouge. The fault core becomes especially localized at depth (on the order of tens of centimeters to meters [Sibson, 2003]). Further localization occurs on an extremely narrow principal fracture surface (hundreds of microns) within the gouge (Figure 1) (also see Rockwell and Ben-Zion [2007]). Laboratory experiments provide analogous observations where extreme slip localization develops in high-speed friction tests [e.g., Kitajima et al., 2010; Di Toro et al., 2011].

Based on particle size analysis and grain surface area estimates, it is inferred that fracture energy (a factor that significantly contributes to rupture dynamics) in the damage zone may be less than 10% of that in the fault gouge [e.g., Chester et al. 2005]. Consequently, the behavior of gouge material that is closest to the fault plane should have a much larger effect on rupture dynamics. Lachenbruch [1980] wrote, “… a more complete understanding of the earthquake process will probably require measurements of the permeability of fault...
zone materials, the width of the active shear zone, and studies of fault gouge dynamics." In this work, we consider the dynamics of fault gouge and investigate how they may contribute to the weakness of mature faults.

While the damage zone typically has the appearance of brecciated host rock or cataclastic rock, the centimeter-thick layer of ultracataclasite gouge is a narrow zone of highly deformed, low-permeability rock that has experienced large amounts of deformation and grain size reduction (comminution) from frictional wear and grain crushing. In the Punchbowl fault, the gouge consists of a very fine grained matrix, composed of grains around ~10 μm in diameter which constitute a combination of host rock particles and fragmented vein material [e.g., Chester et al., 1993].

Apart from textural differences, fault gouge and rock samples from the surrounding damage zone have distinctly different deformational behavior [Chester and Logan, 1986; Scott et al., 1994]. Damage zone rocks deform in a typical elastic-brittle manner accompanied by dilatancy and strain localization, while fault gouge readily compacts and experiences continued compaction and strain hardening in triaxial tests (Figure 2). Although the intact gouge has relatively low initial porosity, structural collapse and comminution under the increase of shear and normal stresses can still cause gouge to compact. The fact that gouge readily compacts leads to an important deformation mode that we consider in this work.

Sleep and Blanpied [1992] proposed interseismic fault zone compaction and the resultant fluid overpressure as a possible mechanism for the weakness of the SAF. Segall and Rice [1995] noted the contradiction with this model, in that extreme pore pressures can drastically reduce effective normal stress and leave faults in the frictionally stable regime, prohibiting earthquakes from nucleating. This problem is avoided through dynamic gouge compaction, as we describe in this study. Compaction of fault gouge has also been studied in Daub and Carlson [2008], Van der Elst et al. [2012], and Lieou et al. [2014]; in these works, the effects of pore fluids were not considered.

Another important characteristic of gouge deformation is its sensitivity to strain rate. Frictional experiments [e.g., Morrow and Byerlee, 1989; Marone et al., 1990] show that gouge compacts during slow frictional sliding, but dilates when slip rate on the fault increases. Rudnicki and Chen [1988] modeled dilatancy hardening due to pore pressure reduction assuming no rate dependence. Segall and Rice [1995] modeled the rate dependence by formulating porosity as a state variable that evolves toward a slip rate-dependent steady state value and coupled it to pore pressure change; this formulation is widely used in models of earthquake cycles and slow slip events [e.g., Segall et al., 2010; Liu and Rubin, 2010]. Recent developments on gouge dilatancy can be seen in Rice et al. [2014].

Here we build on the above works and construct a dynamic model of gouge deformation incorporating both compaction and dilatancy. We will show that the distinct rheology of fault gouge together with undrained

![Figure 2. Comparison of deformation characteristics in triaxial experiments for rocks from the Punchbowl Fault under different confining pressures (from Chester and Logan [1986], with permission from Springer). The sandstone from the damage zone experiences typical brittle failure, while intact gouge compacts during the entire experiment. The strength increase is due to strain hardening. Gouge compaction is an important characteristic that we consider in this study.](image-url)
fluid response in dynamic earthquake rupture likely exerts a fundamental control on the strength of mature faults. In particular, large shear stress concentration ahead of the rupture front causes gouge to compact (shear-enhanced compaction), which leads to increases in pore pressure and reduction of static friction on the fault (also strength drop). During the rapid stress breakdown process at the rupture front, strong gouge dilatancy reduces pore pressure and strengthens the fault, promoting self-healing slip pulses, a slip mode that is thought to operate during earthquakes as inferred from seismic data [Heaton, 1990]. In our model, slip pulse generation can occur at higher background shear stress levels than predicted by the theory of Zheng and Rice [1998]; their theory does not include pore fluid effects. Strong dilatancy and softening in the gouge localize shear strain within a narrow zone in the gouge layer with less generation of inelastic shear strain in the damage zone due to smaller strength drop, which is more consistent with geological observations and high-speed frictional experiments.

With well-developed fault gouge in mature faults we suggest that fault strength at the rupture front is likely controlled by end-cap failure, not Mohr-Coulomb failure; thus, this strength along with lower dynamic frictional strength (governed by dynamic weakening mechanisms) is significantly less than the prediction by Byerlee’s law.

2. Constitutive Modeling

Compaction of porous rocks has been extensively studied in laboratory experiments [e.g., Wong et al., 1997; Baud et al., 2006; Wong and Baud, 2012]. Results show that when the effective confining pressure is low, brittle failure occurs in shear, which often accompanies strain softening and dilatancy. The failure envelope can be well described by the Mohr-Coulomb or the Drucker-Prager criterion, where shear strength increases linearly with effective confining pressure. At intermediate pressures, shear strength is at a maximum and volumetric strain is low. At relatively high confining pressure the shear strength begins to decrease with increasing confining pressures. In this regime, the material fails by compaction and strain hardening. Graphically, an ellipse fits the high confining pressure portion of the failure envelope. Under this yield criterion, compactant failure can in fact occur with the increase of shear stress (shear-enhanced compaction). Stefanov et al. [2011] provides a good review of these different regimes.

The characteristics of fault gouge that allow it to readily compact require a yield criterion other than the Mohr-Coulomb or Drucker-Prager criteria which is commonly used to model the damage zone. As a well-constrained gouge yield strength profile for a range of stress regimes is not available, we seek a yield criterion that is qualitatively similar to the strength envelopes found for porous rock deformation, as described above. The new yield surface needs to close at its end to model yielding under isotropic compression. One example of this type of yield surface is the well-known Cam-clay model [Roscoe et al., 1958], which was the first criterion with such a feature and is widely used in soil mechanics. We use a combined yield criterion where the Mohr-Coulomb criterion (a special case of the Drucker-Prager criterion in plane strain) is used to model brittle shear failure (accompanying strain softening and dilatancy) in the low confining pressure regime. At high confining pressure, an elliptical cap is used to model compaction and strain hardening, where the shear strength decreases with increasing confining pressure. This combined yield surface provides a good fit to the strength of porous rocks as found in experiments [Wong et al., 1997; Wong and Baud, 2012] and has been used to model rock compaction [Andrews, 2007; Sleep, 2010; Stefanov et al., 2011]. The combined yield criterion that we use is shown in Figure 3.

The mathematical form of the yield surface is given by

\[
\tau = c \cos \phi - \left( \frac{\sigma_{lk}}{3} + P \right) \sin \phi, \quad - \left( \frac{\sigma_{lk}}{3} + P \right) \leq S_0
\]

\[
\left( \frac{\sigma_{lk}}{3} - P - S_0 \right) a_{\text{CAP}} + \left( \frac{\tau}{b_{\text{CAP}}} \right)^2 = 1, \quad - \left( \frac{\sigma_{lk}}{3} + P \right) \geq S_0
\]

(1)

where \( \tau \) is the square root of the second invariant of the deviatoric stress tensor in the Drucker-Prager criterion (or the maximum shear stress over all possible orientations in the Mohr-Coulomb criterion), \( \sqrt{\sigma} \) is the mean stress, \( P \) is pore pressure, \( c \) is cohesion, \( \phi \) is the internal friction angle, and \( S_0 \) is the effective mean pressure at the center of the ellipse, and \( a_{\text{CAP}} \) and \( b_{\text{CAP}} \) are the major and minor semiaxes of the ellipse, respectively. The continuum mechanics stress convention is used in this paper (i.e., compression is negative). We use the term
“pressure” in this paper to denote compression as positive for simplicity. In this model the effective mean pressure \( S_0 \) coincides with the intersection of the Drucker-Prager line and the elliptical cap. The pressure \( S_1 \) at the right end of the cap denotes yielding with no shear, i.e., the crushing pressure. Strain hardening associated with compaction is represented by cap expansion. We use a combined hardening rule where the cap expands with both increasing inelastic shear and volumetric strains. Specifically, the cap expands by

\[
dS_0 = -h_S K d\varepsilon_{p,kk} + h_S G d\eta,
\]

\[
dS_1 = h_S dS_0,
\]

where \( K \) is the drained bulk modulus, \( G \) is the shear modulus, \( d\varepsilon_{p,kk} \) is the increment of the inelastic volumetric strain (repeated indices indicate summation), \( d\eta \) is the equivalent inelastic shear strain increment (\( d\eta = \sqrt{2 \frac{d\varepsilon_{p,k}^2}{d\varepsilon_{p,j}^2}} d\varepsilon_{p,j} \), where \( d\varepsilon_{p,j}^2 \) is the increment of the inelastic deviatoric strain), and \( h_S, h_S \), and \( h_S \) are hardening parameters. The parameters used in this work are shown in Table 1.

A simple nonassociated plastic flow rule is used when the yield criterion is violated. In the end-cap regime, the direction of the stress adjustment back to the yield surface, which corresponds to a dilatancy angle \( \theta \), is given by the line between the temporary elastically updated stress and a point on the horizontal axis halfway between the temporary elastically updated effective mean stress and the center of the ellipse (Figure 3). The enforcement of this flow rule is an ad hoc consideration, as we are not aware of well-documented plastic flow behavior of fault gouge material. Laboratory studies aiming to quantify this behavior will be useful for more realistic future models. Nevertheless, this flow rule qualitatively reproduces the behavior of porous rock deformation. Namely, that the dilatancy angle is a function of effective mean stress and causes a transition from dilation (in the low-stress Mohr-Coulomb or Drucker-Prager regime) to compaction (in the high-stress end-cap regime), with no inelastic volumetric strain at the intermediate stress region, near \( S_0 \) [Wong et al., 1997; Stefanov et al., 2011].

During rapid earthquake rupture, the fluid diffusion time is much longer than the travel time of stress waves; thus, the undrained condition applies. In the following, we follow closely Viesca et al. [2008] to implement yielding with undrained fluid response.

During plastic flow, the inelastic strain increment can be written as

\[
d\varepsilon_{p,i} = d\eta \left( \frac{\frac{s_{ij}}{2\tau} + \frac{\beta_i}{3} \sigma_j} \right),
\]
Table 1. Calculation Parameters in the Baseline Case

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>( \rho )</td>
<td>2700 kg/m(^3)</td>
</tr>
<tr>
<td>( P ) wave velocity</td>
<td>( V_P )</td>
<td>6000 m/s</td>
</tr>
<tr>
<td>( S ) wave velocity</td>
<td>( V_S )</td>
<td>3464 m/s</td>
</tr>
<tr>
<td>Gouge width</td>
<td>( w )</td>
<td>20 cm</td>
</tr>
<tr>
<td>Skempton’s coefficient</td>
<td>( B )</td>
<td>0.6</td>
</tr>
<tr>
<td>( \alpha )</td>
<td></td>
<td>0.45</td>
</tr>
<tr>
<td>Off- and on-fault normal stresses</td>
<td>( \sigma_{xx} = \sigma_{yy} = \sigma_n )</td>
<td>(-126 ) MPa</td>
</tr>
<tr>
<td>Off- and on-fault shear stresses</td>
<td>( \sigma_{xy} = \tau_b )</td>
<td>(35 ) MPa</td>
</tr>
<tr>
<td>Direct effect parameter</td>
<td>( a )</td>
<td>0.016</td>
</tr>
<tr>
<td>State variable evolution parameter</td>
<td>( b )</td>
<td>0.2</td>
</tr>
<tr>
<td>State variable evolution distance</td>
<td>( L )</td>
<td>(1.3717 \times 10^{-4} ) m</td>
</tr>
<tr>
<td>Reference friction</td>
<td>( t_0 )</td>
<td>0.7</td>
</tr>
<tr>
<td>Reference slip velocity</td>
<td>( V_0 )</td>
<td>1.0 ( \mu )m/s</td>
</tr>
<tr>
<td>Weakened friction coeff.</td>
<td>( t_w )</td>
<td>0.13</td>
</tr>
<tr>
<td>Weakening velocity</td>
<td>( V_w )</td>
<td>0.17 m/s</td>
</tr>
<tr>
<td>Initial state variable</td>
<td>( \eta_0 )</td>
<td>0.4367</td>
</tr>
<tr>
<td>Characteristic state evolution region</td>
<td>( R_0 )</td>
<td>16. ( \Delta )x</td>
</tr>
<tr>
<td>Internal friction</td>
<td>( \tan \phi )</td>
<td>0.85</td>
</tr>
<tr>
<td>Cohesion</td>
<td>( c )</td>
<td>0 MPa</td>
</tr>
<tr>
<td>Mohr-Coulomb dilatancy angle</td>
<td>( \theta )</td>
<td>( \tan^{-1}(0.5 \sin \phi) )</td>
</tr>
<tr>
<td>Closeness to cap failure</td>
<td>( C_{CAP} )</td>
<td>0.95</td>
</tr>
<tr>
<td>Ellipse aspect ratio</td>
<td>( \alpha_{CAP}/\delta_{CAP} )</td>
<td>2</td>
</tr>
<tr>
<td>Ellipse hardening parameters</td>
<td>( h_{s}, h_{s}, h_{s} )</td>
<td>0.5, 0.3, 0.2</td>
</tr>
<tr>
<td>Steady state dilatancy factor</td>
<td>( \xi )</td>
<td>(1.5 \times 10^{-5} )</td>
</tr>
<tr>
<td>Dilatancy evolution velocity factor</td>
<td>( \zeta )</td>
<td>0.1</td>
</tr>
<tr>
<td>Velocity evolution standard deviation</td>
<td>( \delta )</td>
<td>1 cm</td>
</tr>
<tr>
<td>Element Size</td>
<td>( \Delta x )</td>
<td>1 cm</td>
</tr>
<tr>
<td>Time step</td>
<td>( \Delta t )</td>
<td>1.25 ( \mu )s</td>
</tr>
<tr>
<td>Calculation time</td>
<td></td>
<td>10 ms</td>
</tr>
</tbody>
</table>

where \( s_i \) is the deviatoric stress, the superscript asterisk (*) denotes the elastically updated temporary stress, \( \beta_s \) is the undrained dilatancy factor \( \beta_s = (1 - B) \beta \) (\( \beta \) is the drained dilatancy factor, \( \beta = \tan \theta \)) that gives the ratio of inelastic volumetric strain to inelastic shear strain, i.e., \( \Delta \varepsilon_v = \beta_s \Delta \varepsilon_s \), and \( B \) is Skempton’s coefficient. According to Hooke’s law, for isotropic elastic material the corresponding stress adjustment (equivalent to volume density of seismic moment increment) is

\[
dm_{ij} = 3K_u \delta_{kk} \delta_{ij} + 2G \left( \frac{d\sigma_p}{3} - \frac{d\sigma_p}{3} \right), \tag{5}
\]

where \( K_u \) is the undrained bulk modulus \( K_u = \frac{K}{1 - \nu} \) and \( \alpha \) is Biot's coefficient. We further write isotropic and deviatoric components of the stress adjustment:

\[
dm_{kk} = 3K_u \delta_{kk} \delta_{kk} = 3K_u \beta_s \delta_{kk} \delta_{kk}, \tag{6}
\]

\[
dm_{ij} = \frac{Gd\eta}{\tau} s_i. \tag{7}
\]

The adjustment in mean stress and inelastic volumetric strain can cause pore pressure change. This important effect can be calculated by

\[
dP = -B \frac{dm_{kk}}{3} - \frac{KB}{\alpha} \beta d\eta. \tag{8}
\]

To adjust stress back onto the yield surface we substitute equations (6)–(8) into the yield criterion (equation (1)) and solve for the only unknown, \( d\eta \). For the Drucker-Prager yield criterion this can be written as

\[
\sqrt{\frac{1}{2} \left( \sigma_{ij}^{e} - dm_{ij} \right) \left( \sigma_{ij}^{e} - dm_{ij} \right)} = c \cos \phi - \left( \frac{\sigma_{ij}^{e} - dm_{ij}}{3} + P^* + B \frac{dm_{kk}}{3} - \frac{KB}{\alpha} \beta d\eta \right) \sin \phi, \tag{9}
\]

where \( P^* \) denotes the elastically updated temporary pore pressure. We obtain the solution

\[
d\eta = \frac{P^* - \sigma^{\text{yield}}}{G + K_u \beta_s (1 - B) \sin \phi + \frac{KB \Delta \varepsilon_v}{\alpha}}. \tag{10}
\]
where \( \tau \) is the yield stress, which is identical to the solution of Viesca et al. [2008] except that we ignore the hardening during each inelastic strain increment.

Similarly, for the end-cap yield criterion, we have

\[
\left( \frac{\sigma_{kk} - \alpha_{kk} \delta \varepsilon_{kk} - \nu \sigma_{kk} \delta \varepsilon_{kk}}{a_{CAP}} - \frac{\nu \sigma_{kk} \delta \varepsilon_{kk}}{a_{CAP}} - \frac{\kappa \delta \varepsilon_{kk}}{a_{CAP}} \right)^2 + \left( \frac{\nu \sigma_{kk} \delta \varepsilon_{kk}}{b_{CAP}} \right)^2 = 1.
\]

This results in a quadratic equation of \( \delta \varepsilon_{kk} \). We solve it analytically and the minimum root gives the solution for \( \delta \varepsilon_{kk} \). A similar derivation can be seen in Stefanov et al. [2011], which, however, only considered the dry case. Once \( \delta \varepsilon_{kk} \) is obtained the stresses and pore pressure are adjusted by using equations (6)--(8). The plastic work done can be calculated by using equation (1) or (2) of Ma and Hirakawa [2013].

To model rate dependence of gouge dilatancy, we follow the formulation of Segall and Rice [1995] for porosity evolution during frictional sliding. When the stress is within the yield surface, we evolve inelastic volumetric strain via the following equations:

\[
\dot{\epsilon}_{kk}^P = -\frac{V_{ev}}{\lambda} \left( \dot{\epsilon}_{kk}^P - \dot{\epsilon}_{kk,ss}^P \right),
\]

\[
\dot{\epsilon}_{kk,ss}^P = \zeta \sinh^{-1} \left( \frac{V_{ev}}{2V_0} \right), \text{ and}
\]

\[
V_{ev} = \zeta V \exp \left( -\frac{y^2}{2\theta^2} \right).
\]

Assuming that the solid skeleton has a much lower compressibility, inelastic volumetric strain is equivalent to porosity change [Viesca et al., 2008]. Equation (12) is similar to the slip law for state variable evolution in the rate-and-state friction law (shown in the next section in equation (18)). Equation (13) provides an expression for steady state inelastic volumetric strain \( \dot{\epsilon}_{kk,ss}^P \) that is well posed at zero velocity, where \( V_0 \) is the reference slip velocity used in the rate-and-state friction law, and \( V_{ev} \) is evolution velocity. Evolution velocity \( V_{ev} \) (equation (14)) is a function of slip velocity, \( V \), and also a Gaussian function of distance from the fault (with \( y \) being fault-perpendicular distance and the standard deviation). A similar concept was used by Sleep [1997]. Segall and Rice [1995] used slip velocity directly for the porosity evolution, i.e., \( \zeta = 1 \). The parameters \( \zeta \) and \( \zeta \) control the amount of gouge dilatancy during evolution. In the simulations we choose \( \zeta = 1.5 \times 10^{-5}, \zeta = 0.1, \) and \( \delta = 1 \) cm as the baseline case, and will also discuss the effects of varying these parameters.

We solve equations (12)--(14) to obtain the inelastic volumetric strain increment \( \delta \varepsilon_{kk,ss}^P \) during dilatancy evolution. A dilatancy angle \( \theta = \max \left[ \tan^{-1} \left( \frac{0.577}{\delta \varepsilon_{kk,ss}^P} \right), 5^\circ \right] \) is specified to obtain the inelastic shear strain increment \( \delta \varepsilon_{ss}^P \):

\[
\eta = \frac{\alpha_{ss}^P \delta \varepsilon_{ss}^P}{P} = \frac{\alpha_{ss}^P}{P} \delta \varepsilon_{kk,ss}^P.
\]

We then use equations (6)--(8) to apply the corresponding stress and pore pressure adjustments. This dilatancy angle controls the amount of inelastic shear strain and shear strength reduction (strain softening) for a given inelastic volumetric strain increment; smaller angles lead to larger strain softening. We see that as the strength \( (\tau) \) softens, this angle decreases, which is consistent with typical rock failure experiments. A minimum angle of \( 5^\circ \) is used to limit the amount of softening. Our introduction of equations (12)--(14) inside the yield surface is different from standard plasticity models in which the region inside the yield surface corresponds to an elastic regime. Here these equations allow us to kinetically model the effect of gouge dilatancy in a rate-and-state framework.

### 3. Dynamic Rupture Model

We model dynamic rupture on a 2-D planar, right-lateral strike-slip fault embedded in a narrow gouge layer that is surrounded by a damage zone (Figure 4). Off-fault material is poroelastoplastic. The gouge layer obeys the combined yield criterion described earlier. The damage zone obeys the standard Mohr-Coulomb criterion only. The width of the gouge layer is chosen as 20 cm for most cases to be consistent with geological observations (Figure 1); the effect of varying this parameter will also be shown. The fault length is sufficiently long that the rupture never reaches the end of the fault. Elastic material properties are homogeneous, where \( P \)
wave velocity, $S$ wave velocity, and density are $V_p = 6000$ m/s, $V_s = 3464$ m/s, and $\rho = 2670$ kg/m$^3$, respectively. An element size of 1 cm is used in all the simulations to ensure a good resolution of the gouge layer.

The regional stress field is homogeneous with the maximum compressive stress oriented at 45° to the fault. Initial stresses on the fault are selected to be representative of conditions at seismogenic depth (~7 km). The initial effective normal stress on the fault is $-126$ MPa.

Zheng and Rice [1998] defined a critical background stress $\tau_{\text{pulse}}$ given by the intercept of the radiation damping line that is tangent to the steady state friction curve. They showed that propagation of crack-like ruptures is not possible when the background shear stress is lower than $\tau_{\text{pulse}}$, and pulse-like ruptures are only possible in a narrow range of initial background stresses around $\tau_{\text{pulse}}$. For our chosen parameters (Table 1), $\tau_{\text{pulse}}$ is 30.64 MPa. In order to produce a range of slip modes (i.e., arresting rupture, crack-like rupture, and pulse-like rupture), we use a range of background shear stresses on the fault between 30 and 40 MPa. Our stress states are similar to the ones used in Dunham et al. [2011a].

The Mohr-Coulomb parameters are the same in both the fault gouge and the damage zone, assuming zero cohesion and an internal friction of 0.85. Ellipse parameters are set so that the initial stress state brings the gouge close to compactant failure. To satisfy this condition, we define a parameter $C_{\text{CAP}}$, to be the ratio of the initial shear stress to the shear strength,

$$C_{\text{CAP}} = \frac{\tau_{\text{init}}}{\sigma_{\text{init}} b_{\text{init}}^2}$$

where the superscript “init” indicates that these are initial stress values. The parameter $C_{\text{CAP}}$ is analogous to the closeness-to-failure parameter used in studies with Drucker-Prager plasticity [e.g., Templeton and Rice, 2008]. As we will show that compaction is induced mostly by the shear stress increase ahead of rupture front (with negligible change in normal stress), $C_{\text{CAP}}$ is defined by only considering the relative vertical distance to the yield cap. The initial aspect ratio of the ellipse is chosen to be 2 in all simulations, which is motivated by the data for porous rocks [Wong and Baud, 2012]. With these parameters $C_{\text{CAP}}$ alone completely determines the shape and position of the elliptical cap. We use $C_{\text{CAP}} = 0.95$ in the baseline case (we also have results for two other cases, $C_{\text{CAP}} = 0.75$ and 0.85). Parameterization of compactant strength by this method is justified by the observation that gouge readily compacts at different confining pressures (Figure 2). This implies that the ellipse parameters in this study do not represent actual material parameters, but can be thought of as a means to represent the history of the gouge deformation that has brought the stress state close to failure.

A friction law is necessary to govern evolution of fault strength during rupture. Recently, progress has been made in the development of frictional models for fault gouge that employ flash heating for rapid weakening. Elbanna and Carlson [2014] implemented flash heating and temperature-dependent viscoplasticity into a framework of shear transformation zone theory, which takes into consideration the granular nature of fault gouge and links microscopic to macroscopic heating processes. Platt et al. [2014] used numerical models of friction in gouge with dilatancy and frictional rate strengthening to study localization zone thickness. Though these recent models are insightful, we use the rate-and-state friction law with strong velocity weakening which mimics flash heating of microasperities, to facilitate comparison with similar studies of dynamic

---

**Figure 4.** Model geometry. A right-lateral strike-slip fault is embedded in a narrow layer of fault gouge, which obeys the combined Mohr-Coulomb-end cap yield criteria (Figure 3). The surrounding damage zone obeys the Mohr-Coulomb criterion only. A constant dilatancy angle is used in the damage zone, $\theta = \tan^{-1}(0.5 \sin \phi)$. Maximum principal stress is oriented at 45° to the fault.
rupture. We follow closely the formulation of Dunham et al. [2011a] (see also Rice [2006], Noda et al. [2009], Rojas et al. [2009], and Shi and Day [2013]).

In this formulation, shear stress on the fault is always equal to the strength of the fault

$$\tau = f'(-\sigma_n - p) = -f\sigma_n$$

(16)

where $\sigma_n$ and $\tau$ are the total normal stress and shear stress on the fault, respectively, $p$ is on-fault pore pressure, and $\sigma_n'$ is the effective normal stress on the fault.

The friction coefficient, $f$, on the fault is governed by

$$f(V, \psi) = a\sinh^{-1}\left[\frac{V}{2V_0}\exp\left(\frac{\psi}{a}\right)\right].$$

(17)

where $V$ is slip velocity, $V_0$ is a reference slip velocity, $a$ is the direct effect parameter, and $\psi$ is the state variable that evolves according to the slip law:

$$\dot{\psi} = -\frac{V}{L}(\psi - \psi_{ss}),$$

(18)

$$\psi_{ss} = a\ln\left[\frac{2V_0}{V}\sinh\left(\frac{f_{ss}}{\sigma}\right)\right].$$

(19)

The steady state variable $\psi_{ss}$ is a function of the steady state friction $f_{ss}$ that satisfies

$$f_{ss} = f_w + (f_{LV} - f_w)\left[1 + \left(\frac{V}{V_w}\right)^{8/3}\right],$$

(20)

$$f_{LV} = f_0 - (b - a)\ln\left(\frac{V}{V_0}\right).$$

(21)

where $f_{LV}$ is steady state value at conventional slow slip velocities, $V_w$ is the weakening velocity beyond which flash heating takes effect, and $f_w$ is the weakened friction coefficient. All friction parameters are shown in Table 1 and similar to those in Dunham et al. [2011a].

In this work, we require the pore pressure on the fault (equation (16)). Calculation of on-fault pore pressure from pore pressures in the volume is not trivial when there is inelastic strain [Viesca et al., 2008; Viesca and Rice, 2009]. We assume the on-fault pore pressure to be the average of the pore pressures on both sides of the fault in this work. The implications of this assumption will be discussed later in the paper.

Ruptures are nucleated by applying an instantaneous Gaussian perturbation in shear stress. The standard deviation of the Gaussian function is the characteristic extent of the state evolution region $R_0$ (16 cm) and the amplitude is such that the sum of the peak perturbation and the background stress equals 0.7 $\sigma_n'$ at the center of the fault. Due to the extremely small element size (1 cm) each simulation is run for 10 milliseconds (ms), corresponding to a rupture distance of ~30 m, except for one simulation that we ran for 30 ms (corresponding to a rupture distance of ~90 m) to study the off-fault damage profile after a long propagation distance.

4. Results

We show inelastic shear strain, inelastic volumetric strain, and pore pressure change in the medium after 8 ms of rupture propagation in the baseline case (Figure 5). Slip velocity and on-fault pore pressure change curves are superimposed on these figures. The dashed lines show the location of the rupture front. Prior to the arrival of rupture, the fault plane experiences shear stress increase from two contributions, both of which lead to shear-enhanced compaction owing to the condition that fault gouge is close to compactant failure. The first is the arrival of the $S$ wave, which occurs a few meters ahead of the rupture front. This causes a simultaneous increase in pore pressure and an implicit decrease in shear strength on the fault. Following this, the gouge experiences a much larger increase in shear stress due to the approach of the mode II crack tip, an essential concept in fracture mechanics [Freund, 1998]. This crack tip stress dominates the prerupture shear-enhanced compaction and consequent pore pressure increase; the compaction from the $S$ wave arrival is minor in comparison.
$S$ wave compaction generates a plateau in the on-fault pore pressure distribution. Because the fault has not slipped, there is no evolution in inelastic volumetric strain, so the pore pressure stays constant until the arrival of the rupture front. Shortly after the rupture front arrives, the on-fault pore pressure drops rapidly. Gouge starts to dilate due to the evolution of inelastic volumetric strain toward a positive steady state value (equation (13)). Large slip rate during strength drop gives rise to a large evolution velocity ($V_{ev}$), which leads to rapid dilatancy and increase of inelastic shear strain due to softening. The Gaussian function of evolution velocity (equation (14)) limits the dilatancy and inelastic shear strain within 3 standard deviations to the fault, which can be seen as a narrow localized zone behind the rupture front. Outside this localized zone, most of the gouge layer remains compacted because of the negligible evolution effect. The inelastic shear strain outside the localized zone is mostly due to plastic yielding during the prerupture compaction by the large stress concentration carried by the rupture front.

The strong gouge dilatancy during strength drop rapidly reduces pore pressure and strengthens the fault, which can be seen in the on-fault shear stress distribution. The strengthening locks the fault, generating slip pulses. The fault strengthening due to undrained dilatancy is in addition to the rate strengthening (equations (18) and (19)) and is more efficient to quench slip.

Figure 5. Snapshot of inelastic shear strain, inelastic volumetric strain, and pore pressure change is mapped at 8 ms simulation time. Slip velocity (magenta curves) and on-fault pore pressure change (solid black curves) on the fault are superimposed on the top and bottom figures, respectively. The numbers next to the arrows give the peak values of these curves. The portion shown is a region near the rupture front to the right of the nucleation point (distance along strike in the x axis is in meters from the nucleation point). The dashed lines mark the location of the rupture front. The combination of the $S$ wave and the intense shear stress in front of the crack tip causes compaction prior to slip. This increases the pore pressure before rupture arrives. When the fault begins sliding, dilatancy evolution begins to take place, and the gouge becomes dilated and softened, and pore pressure is reduced. This leads to a narrow zone of dilated, highly strained gouge (see also Figure 6) and a short-duration slip pulse on the fault.
The distribution of inelastic shear strain is distinctly different from what is found in off-fault plasticity models using the Mohr-Coulomb or Drucker Prager yield criteria. We provide one example by modeling rupture in the absence of the gouge (i.e., all parts of the domain obey only the Mohr-Coulomb yield criterion and on-fault pore pressure remains zero), but with otherwise identical parameters (Figure 6). The damage zone characteristic of previous models emerges when the gouge layer is ignored, where a triangular pattern of inelastic shear strain outlines a damage zone that grows with propagation distance. However, in the case with the gouge layer, we see highly localized shear strain within the gouge layer and slightly less inelastic shear strain in the damage zone. The substantially large amount of shear failure in the gouge in comparison with the damage zone is in agreement with geological observations. The decrease in shear failure in the damage zone itself in comparison with the Mohr-Coulomb model is due to the reduction in strength drop on the fault (Figure 7), and is supported qualitatively by the notion that off-fault damage should be small, and shear failure is highly localized to the fault core [e.g., Chester et al., 1993; Brune, 2001; Sibson, 2003; Rockwell and Ben-Zion, 2007].

We show time histories of slip velocity, shear stress, and on-fault pore pressure at a fault node 15 m from the hypocenter (Figure 7). We compare the results of our end-cap model with simulations with either elastic or purely Mohr-Coulomb off-fault response and no on-fault pore pressure change. Other than the yield surfaces used (or the lack of a yield surface in the elastic case) and lack of on-fault pore pressure change, the three

---

**Figure 6.** (a) Comparison of inelastic shear strain distributions between our end-cap model (top) and a Mohr-Coulomb (MC) model. The MC model shows a typical triangular region of damage, while the end-cap model has the most deformation occurring in the gouge with highly localized failure near the fault. (b) Plots of inelastic shear strain are shown in cross sections at two distances along the fault strike. The MC case has more damage outside of the gouge but far less damage inside of the gouge zone.
simulations are identical. In the end-cap model, the initial pore pressure increase from $S$ wave compaction is visible shortly before 4 ms. Subsequent pore pressure increase from compaction due to the arrival of the rupture front is much more significant. At the rupture front the pore pressure increase is already $\sim 27.2$ MPa. The maximum pore pressure increase is 34.94 MPa, because a small amount of continued compaction occurs on the compressional side of the fault after the rupture front arrives (see Figure 8). The peak frictional strength is 25.53 MPa lower than that in the elastic case. This corresponds to a decrease in static friction from $\sim 0.68$ in the elastic case to $\sim 0.48$ in the end-cap case. This reduction in strength drop in conjunction with plastic dissipation and dilatancy hardening reduces both the peak slip velocity and rupture velocity in comparison to both the elastic and Mohr-Coulomb cases.

Stage I encompasses the time span before the rupture front arrives. At this stage, shear stress increase first drives the stress point to move vertically upward in stress space, until the stress state reaches the yield cap. Then, yielding causes compaction and drives the stress to the left of stress space, increasing pore pressure and decreasing the effective mean pressure. The first increase and subsequent drop is the $S$ wave arrival, and the second corresponds to the large shear stress concentration before the arrival of the rupture front (defined as when slip velocity exceeds $10^{-3}$ m/s). The cap expands as a result of strain hardening. During this stage, the stress state remains nearly the same on both sides of the fault. Slight deviations between stress paths can be seen because the two points are at 0.5 cm (half the element size) from the fault and also there are small radiations by $P$ and $S$ waves from the inelastic strain (corresponding to seismic potency) behind the rupture front.

Stage II encompasses the time between the arrival of the rupture front and the time that shear stress drops to its residual value. This is an extremely brief period of time in comparison with stages I and III; however, it predictably features the most drastic variations in the stress paths. On the compressional side, the mean pressure increases during compression and causes further compaction. On the extensional side, the mean pressure is reduced, which causes the stress state to reach the Mohr-Coulomb line and the gouge experiences shear dilatancy. After these brief surges in mean stress change, both sides of the fault experience a large shear stress drop (coinciding with strength drop on the fault) and elastic mean stress relaxation (compressional side becomes more tensile and extensional side more compressive). The sudden onset of sliding during this period causes strong dilatancy evolution within 3 standard deviations of the evolution velocity to the fault, which leads to additional adjustments on mean and shear stresses.

Figure 7. Time histories at $x = 15$ m for an elastic case, a case with only the Mohr-Coulomb criterion, and a case with the Mohr-Coulomb-end cap model. All three cases are identical except for the yield surface used (or lack of yield surface for the elastic case) and the on-fault pore pressure calculation. The dashed line at 5.4056 ms marks the rupture front for the end-cap case. With the cap model, on-fault pore pressure increase due to prerupture compaction lowers static friction and strength drop, resulting to a smaller slip velocity. Pore pressure decrease from dilatancy hardening can also be clearly seen.
Finally, in stage III, a period of continued dilatancy evolution and rate strengthening occurs. The evolution of inelastic dilatancy reduces pore pressure and increases the effective mean pressure on both sides of the fault. The plastic flow rule that we used corresponds to strain softening inside the yield surface, which is a shear stress decrease. Shear stresses on both sides of the fault also increase due to dilatancy hardening on the fault and rate strengthening in the rate-and-state friction law, because shear traction is continuous across the fault. The interaction of these effects causes the small oscillations in the stress paths in stage III. The kink in the slip velocity time history marks the start of stage III, because the combination of these effects all cause the slip velocity to decay at a different (likely faster) rate. Interestingly, the shape is similar to the slip rate function used in some kinematic rupture models to simulate broadband ground motion [e.g., Liu et al., 2006]. Additional increases in shear stress occur as a result of variations of fault-parallel stresses in material that has experienced plastic yielding [Templeton and Rice, 2008].

We compare slip contours, inelastic volumetric strain and pore pressure change distributions for three CF cases (0.75, 0.85, and 0.95) in Figure S1 in the supporting information. As expected, the degree of compaction and the resultant pore pressure elevation are less for the cases with lower CF. Although weakening due to the S wave arrival is not seen in the two lower CF cases, significant overpressure is present in all three

Figure 8. (a) Stress paths at two points (x = 15.5 cm) on either of the fault, located at half the element size away from the fault plane. Three stages of deformation can be identified, marked as I, II, and III on stress paths and time histories. The black dot marks the initial stress state. The red and blue dots marked as RF are the stress states at the rupture front arrival, which mark the start of stage II. The red and blue dots marked τ, show the stress state at the time when shear stress on the fault reaches its minimum value (the residual shear stress). This marks the start of stage III. (b) Relevant time histories are plotted at the same points. The three stages are noted as well. Note that the final pore pressure change is negative (more dilated) in the compressional side and positive (more compressional) in the extensional side of the fault. This is because the fault-parallel normal stress is more compressive on the extensional side and more extensional on the compressional side to reach final equilibrium [Templeton and Rice, 2008].
cases due to large shear stress increase ahead of the rupture front. On-fault pore pressure, shear stress, and slip velocity time histories for the three cases are similar (Figure S2), suggesting that extreme closeness to failure is not a necessary condition for the prerupture compaction weakening to occur. Rupture velocity is the largest for the case $C_{F\text{CAP}} = 0.95$. Higher $C_{F\text{CAP}}$ induces a larger amount of fracture energy that should lower rupture velocity, but at the same time the fault is weakened from the elevated pore pressure. In this case prerupture weakening by compaction is more effective in facilitating rupture propagation than the tendency of plastic dissipation to limit rupture propagation.

To highlight the importance of gouge dilatancy evolution, we compare cases with different values for the dilatancy parameter $\zeta$, a scaling parameter between evolution velocity ($V_{ev}$) and slip velocity (equation (14)). When $\zeta = 0$, $V_{ev}$ is also zero and $\varepsilon_{p,kk}^s$ does not evolve during sliding. Pore pressure on the fault remains high during and after slip, the fault remains weak, and the rupture is crack-like (Figures 9 and 10). The moderate value of $\zeta = 0.01$ allows dilatancy to evolve but not as rapidly as in the baseline case ($\zeta = 0.1$). In this case, the rupture initially appears crack-like but becomes pulse-like in later stages. The baseline case with $\zeta = 0.1$ allows a typical pulse-like rupture. When $\zeta = 1$, rapid dilatancy prohibits rupture propagation.

The effects of the dilatancy parameter, $\zeta$, are illustrated in Figures S3 and S4. The parameter $\zeta$ governs steady state inelastic volumetric strain $\varepsilon_{p,kk}^s$. Increasing $\zeta$ corresponds to evolution toward a more dilatant state. For $\zeta = 0$, $\varepsilon_{p,kk}^s$ does not evolve as severely so the pore pressure remains high and keeps the fault relatively

---

**Figure 9.** Comparison of results for cases with varying dilatancy parameter $\zeta$. (left column) Slip contours every 0.5 ms, (middle column) snapshots of inelastic volumetric strain at 8 ms, (right column) and pore pressure change distribution at 8 ms are illustrated. The color scales are the same as in Figure 5. The number in each panel denotes the peak value in the medium. When $\zeta = 0$, the gouge does not experience dilatancy evolution during sliding, so the pore pressure remains relatively high. This causes the fault to remain weak, and fault rupture is crack-like. The narrow zone that appears to have lower pore pressure is due to elastic reloading on the compressional side after experiencing inelastic failure [Templeton and Rice, 2008]. As $\zeta$ increases, dilatancy during sliding becomes strong and begins to make rupture more pulse-like, until it does not allow rupture at large values.
weak during sliding. On-fault pore pressure change remains well above zero even after slip stops (Figure S4). As $\zeta$ increases, evolution toward higher dilatancy causes pore pressure to drop more rapidly and toward a lower absolute value. Final pore pressure change is negative in the case with $\zeta = 2.5 \times 10^{-5}$. Cases with larger $\zeta$ have shorter slip durations, larger residual strengths, smaller strength drops, slip velocities, and slips (Figure S4); these are all manifestations of the larger dilatancy effect. For $\zeta = 3.5 \times 10^{-5}$, strong dilatancy prevents rupture from propagating (Figure S3).

The standard deviation ($\delta$) of the Gaussian function for evolution velocity strongly controls the width of the intense inelastic shear strain zone (Figure S5). The decrease of $\delta$ lessens the restrengthening effect from dilatancy and leads to a larger slip velocity (Figure S6). This causes enhanced weakening ahead of the rupture front, as high-stress concentrations from larger stress drops lead to more compaction. The actual slip zone can be as narrow as hundreds of microns (Figure 1). For such narrow shear zones we expect that the large dilatancy parameters that Segall and Rice [1995] used ($\xi = 1$ and $\zeta = 1.7 \times 10^{-4}$) can be used to counteract the extreme prerupture weakening while pulse-like ruptures can still be obtained. These results suggest that extreme strain localization in geological observations, weakness of mature faults, and pulse-like ruptures are likely interrelated. In our model, this interrelation is embodied in the undrained gouge compaction and dilatancy driven by the rupture front stress field.

Figure 11 shows how background shear stress affects the rupture mode (i.e., crack-like or pulse-like) when undrained gouge compaction and dilatancy operate. For comparison, we show slip contours for cases with background stresses between 32 and 40 MPa and equivalent cases with only Mohr-Coulomb failure and no on-fault pore pressure change. For both simulations, ruptures with background stress below $\tau_b = 33$ MPa are quickly arrested. Ruptures begin to take the form of growing slip pulses at $\tau_b = 33$ MPa. In the Mohr-Coulomb case, rupture mode transitions to crack-like at $\tau_b = 37$ MPa. In contrast, the inelastic models maintain pulse-like ruptures for all background stresses above 33 MPa, with slip distributions maintaining qualitative similarity for all cases. This reflects the capacity of undrained dilatancy to more effectively restrengthen the fault and terminate slip, even for high background stresses.

Dynamic weakening from prerupture gouge compaction should allow rupture propagation at even lower background stress levels than $\tau_b = 33$ MPa. However, in our model dilatancy evolution starts immediately when the fault is slipping, which strengthens the fault in the nucleation zone and prevents rupture propagation. This is an artifact. When we reduce the amount of dilatancy in the nucleation region, we do observe rupture propagation at lower background stress levels (results not shown here).

5. Discussion

Gouge compaction of varying degrees has been reported in recent high-speed frictional experiments [e.g., Kitajima et al., 2010; Ujiie and Tsutsumi, 2010; Faulkner et al., 2011; Ujiie et al., 2013; French et al., 2014].
Kitajima et al.'s [2010] study may be the most descriptive. Their high-speed rotary shear experiments on ultracataclasite from the surface exposure of the Punchbowl Fault revealed four distinct gouge units that developed under varying amounts of slip and slip rate. Units 1 and 2, developed under low slip and slip rate, are slightly compacted and in some cases foliated but appear to largely maintain the fabric of the undeformed gouge with low shear strain. Units 3 and 4, formed under larger velocities and displacements, consist of a more random fabric, are less compacted, and define regions of highly localized slip. Unit 4 exists only in a very narrow region near the frictional surface, while the rest of the gouge consists of less intensely deformed Units 1–3. Overall, those findings suggest a process in which gouge deformation initially (under low strains and strain rates) occurs via distributed compaction throughout the majority of the gouge layer with overall maintenance of the original gouge fabric, followed by a period of intense localized and dilatant shearing which only occurs in the immediate vicinity of the fault discontinuity. The overall deformation features in these
experiments are similar to our model results, where a compacted region of little shear strain surrounds an extremely narrow dilatant region of intense inelastic shear strain (Figures 5 and 6). These laboratory studies do not feature rupture propagation due to their small laboratory scales and experimental configurations. Therefore, they do not allow observation of the resulting weakening effect that we model here, which is due to compaction ahead of a propagating rupture. Future laboratory experiments may be able to verify this effect.

In contrast to other dynamic weakening mechanisms, this mechanism lowers the apparent static friction. Large shear stress concentration associated with the rupture front is a physical concept and must exist. Such stress concentrations are probably the largest stresses that the gouge layer experiences during the entire earthquake cycle, and it should not be surprising that it can cause comminution and compaction although the gouge may be overconsolidated during the interseismic period.

Microstructural observations of comminution in the fault gouge provide evidence that this occurs. Chester et al. [1993] wrote: “Because of continual reworking, the ultrafine-grained matrix material of the ultracataclastic rocks may be derived, to a large extent, from recombination of neomineralized material and alteration products.” The implication is that comminution is an occurrence in each successive earthquake, and the currently existing gouge microstructures likely represent only the deformation from the last event, rather than the entire fault history.

Fluid overpressure was observed in the high-speed frictional experiments cited above. Most authors attributed it to thermal pressurization, with the exception of Ujiie and Tsutsumi [2010] who suggested that pore pressurization could be due to shear-enhanced compaction. Our model here does not include thermal pressurization. The weakening is due to gouge compaction ahead of rupture front and flash heating during rapid sliding. We show that strong gouge dilatancy during strength drop and rapid sliding reduces pore pressure and promotes slip pulses. Depending on the relative contribution between thermal pressurization and mechanical effects, dilatancy has the potential to diminish the effect of thermal pressurization.

On-fault pore pressure plays an important role in dynamic fault weakening and strengthening in this model. Rudnicki and Rice [2006] and Dunham and Rice [2008] show that on-fault pore pressure can be represented by a weighted average between the pore pressures on both sides of the fault, and the weights are given by a function of permeabilities and storage coefficients. In this work, we assume that these hydraulic parameters do not change with inelastic deformation (or change by the same amount). The weighted average reduces to a simple average of the pore pressures on either side of the fault.

This should be a valid assumption at least during the prerupture weakening phase of the simulation. Ahead of the rupture front, shear stress increases on both sides of the fault, and normal stress change on the fault is nearly zero. Inelastic strains are nearly identical on both sides of the fault. Permeability and storage coefficient changes due to inelastic strain should be nearly symmetric on both sides of the fault, making our assumption valid. However, after rupture commences, the two sides of the fault experience different inelastic deformation. Calculating the inelastic storage coefficient is not trivial, but a solution was given by Viesca and Rice [2009]. The change in permeability should be much more significant as small changes in porosity can lead to permeability change by orders of magnitude. For example, assuming that permeability increases with porosity, more dilatant failure on the extensional side of the fault should more heavily weight its pore pressure contribution; we may underestimate the restrengthening effects from dilatancy. A more rigorous approach to dynamically updating hydraulic parameters with plasticity may be a topic of future work; however, this will be difficult due to the complicated nonlinear relationship between stresses, elastic and inelastic strains, and permeability [e.g., Zhu et al., 2007].

The microstructural observations of Chester et al. [1993] along with other petrological and geochemical findings [e.g., Evans and Chester, 1995; Morton et al., 2012; Colby and Girty, 2013] show that fault zone processes may be intimately involved with fluid flow. Silson et al. [1988] showed evidence for fault-valving behavior, where low-permeability seals are broken during rupture and drive trapped overpressured pore fluids up a fault zone that acts as a conduit. Our model suggests that fluid overpressure can also be generated in the gouge itself, which can drive postseismic fluid flow into the damage zone that is coseismically dilated and more permeable (Figures 5 and 6). The flow of overpressured fluid from the compacted gouge layer into the central narrow dilated shear zone may also contribute to afterslip (T. Yamashita and T. Suzuki, personal communication, 2016).
Our model also offers an alternative explanation for the overthrust paradox [Hubbert and Rubey, 1959; Price, 1988]. Large thrust sheets can move in a piecemeal fashion with transient fluid overpressure on the fault from gouge compaction driven by the dynamic stress field of the rupture front. Pore pressure needs not to be permanently high in the fault zone although we do not argue against its existence.

Zheng and Rice [1998] showed that the critical background stress, \( \tau_{\text{pulse}} \), represents a stress level below which it is not possible to host crack-like ruptures. The dynamic rupture models of Dunham et al. [2011a] with Drucker-Prager plasticity support this but showed that the different rupture mode regimes shift to higher background stresses due to plastic dissipation (i.e., maintenance of slip pulses requires higher background stresses). Conversely, Noda et al. [2009] showed that the inclusion of pore fluid increase via thermal pressurization shifted the regimes to lower background stresses (i.e., maintenance of slip pulses and transition to crack-like ruptures can occur at lower shear stress). Our results show that pulse-like ruptures can be maintained for a wide range of background stresses. Dynamic restrengthening from dilatancy hardening limits ruptures from becoming crack-like in comparison to the Mohr-Coulomb models. The weakening ahead of the rupture front by compaction also has the ability to host ruptures at lower background stresses than an equivalent Mohr-Coulomb model.

### 6. Conclusions

Mature fault zones contain a well-developed layer of ultracataclasite gouge. Distinctly different from brittle shear failure of rocks in the surrounding damage zone, gouge readily compacts and exhibits velocity-strengthening behavior under slow loading but dilates when strain rate is high, showing a strong rate dependence.

We introduce a combined Mohr-Coulomb and end-cap yield criterion to model gouge compaction and dilatancy. The rate dependence of gouge dilatancy is incorporated using a formulation similar to that of Segall and Rice [1995], allowing inelastic volumetric strain to evolve toward a steady state value that depends on slip rate and width of the shear zone. Undrained fluid response is incorporated in this plasticity framework.

Our dynamic rupture simulations with inelastic gouge deformation show that large shear stress concentration ahead of the rupture front causes gouge to compact (e.g., by structural collapse and comminution). This dynamic stress concentration carried by the rupture front is expected to be much larger than the stresses that the gouge layer experiences during the interseismic period. Compaction increases the pore pressure and reduces static friction. Shortly after the rupture front passes, gouge dilatancy during rapid stress breakdown and sliding quickly relieves the fluid overpressure developed during the prerupture compaction and restrengthens the fault, promoting slip pulses. Pulse-like rupture mode is possible at background stresses higher than those found in the Mohr-Coulomb case. Undrained compaction weakening also allows growing slip pulses at lower background stresses.

Our model produces highly localized shear deformation. While the gouge layer experiences distributed compaction, the inelastic shear strain is highly localized onto a narrow zone within the gouge, consistent with geological observations and high-speed frictional experiments. Reduction in strength drop limits the stress concentration outside the gouge layer, generating less inelastic shear strain in the surrounding damage zone. This suggests that dynamic rupture models with Mohr-Coulomb or Drucker-Prager plasticity may overestimate the effect of off-fault failure in the damage zone while underestimating the degree of damage very close to the fault plane.

The shear zone widths in geological observations (approximately hundreds of microns) are much smaller than the centimeters that we model here. As the shear zone narrows, it limits the strengthening effect from dilatancy hardening, and we expect a larger stress concentration ahead of rupture front and more significant prerupture compaction weakening. Similarly, the width of active slipping zone in large earthquakes (approximately kilometers or more) is much larger than the meter scale modeled here, which can also induce much wider stress concentration ahead of rupture front and cause more significant weakening. The frictional strength at the rupture front and the strength drop can be further reduced, leading to more localized slip. With well-developed fault gouge we suggest that the fault strength is likely limited by the end-cap failure, not the Mohr-Coulomb failure. Therefore, the strength of mature faults is significantly smaller than the prediction from the Byerlee’s law.
Acknowledgments

We are grateful for insightful comments from the Associate Editor and two anonymous reviewers that significantly improved the paper. James Brune’s seminar at SDSU in 2014 motivated us to consider the slip localization. Discussion with Steve Day, Gary Girty, and Tom Rockwell was helpful. This work was supported by the Southern California Earthquake Center (contribution 6296). SCEC is funded by NSF Cooperative agreement EAR-1033462 and USGS Cooperative agreement G12AC0038. The numerical simulations were conducted at the Center for High-Performance Computing at the University of Southern California. The data for this paper are available by contacting the corresponding author at hirakawa1@lrl.gov.

References

Andrews, D. J. (2007), Compaction can limit peak vertical velocity at Yucca Mountain, AGU, Fall Meeting 2007, abstract #514C-07.


Sibson, R. H. (1973), Interactions between temperature and pore-fluid pressure during earthquake faulting and a mechanism for partial or total stress relief, Nature, 243(126), 66–68.


Sibson, R. H., F. Robert, and K. H. Poulsen (1988), High-angle reverse faults, fluid-pressure cycling, and mesothermal gold-quartz deposits, Geology, 16(6), 551–555.


