A unified description of localization for application to large-scale tectonics

Laurent G. J. Montési and Maria T. Zuber
Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Received 16 May 2000; revised 14 August 2001; accepted 19 August 2001; published 7 March 2002.

[1] Localized regions of deformation such as faults and shear zones are ubiquitous in the Earth’s lithosphere. However, we lack a simple unified framework of localization that is independent of the mechanism or scale of localization. We address this issue by introducing the effective stress exponent, \( n_e \), a parameter that describes how a material responds to a local perturbation of an internal variable being tested for localization. The value of \( n_e \) is based on micromechanics. A localizing regime has a negative \( n_e \), indicating a weakening behavior, and localization is stronger for more negative \( 1/n_e \). We present expressions for the effective stress exponent associated with several mechanisms that trigger localization at large scale: brittle failure with loss of cohesion, elastoplasticity, rate- and state-dependent friction, shear heating, and grain-size feedback in ductile rocks. In most cases, localization does not arise solely from the relation between stress and deformation but instead requires a positive feedback between the rheology and internal variables. Brittle mechanisms (failure and friction) are generally described by \( n_e \) of the order of \(-100\). Shear heating requires an already localized forcing, which could be provided by a brittle fault at shallower levels of the lithosphere. Grain size reduction, combined with a transition from dislocation to diffusion creep, leads to localization only if the grain size departs significantly from its equilibrium value, because either large-scale flow moves rocks through different thermodynamic environments or new grains are nucleated. When shear heating or grain-size feedback produce localization, \( 1/n_e \) can be extremely negative and can control lithospheric-scale localization. INDEX TERMS: 8160 Tectonophysics: Rheology—general; 3220 Mathematical Geophysics: Nonlinear dynamics; 5104 Physical Properties of Rocks: Fracture and flow; 5120 Physical Properties of Rocks: Plasticity, diffusion, and creep; KEYWORDS: Localization, shear zones, rheology, earthquake, dynamics

1. Introduction

[2] Extensive geological and geophysical observations indicate that tectonic deformation in the Earth’s lithosphere is not uniform but is localized. Localized shear zones range in scale from microscopic cracks and foliation to brittle faults, ductile shear zones, even to entire plate boundaries. Whereas the microscopic processes leading to shear localization have been studied in the field and in the laboratory, the manner by which these processes influence large-scale tectonics has only rarely been assessed [Hobbs et al., 1990; Burg, 1999], partly because of the great variety of these localization processes and partly because of their complexity. To remedy this situation, we derive a general framework for the study of localization based on the effective stress exponent of a rheological system, \( n_e \). This quantity characterizes the non-linear behavior of a rheological system and provides a measure of localization efficiency. As this measure is defined independently of the actual variables involved during localization, it allows the importance of different localizing processes for large-scale tectonics to be evaluated. We will show how to compute the effective stress exponent for processes associated with localized shear zones or faults in the lithosphere.

[3] Defining a localized shear zone is subjective. In contrast to fluid-like behavior, it implies that some measure of deformation (e.g., strain, strain rate) is significantly heterogeneous in a given region, often being greatly enhanced within a narrow area. The scale of observation is important to define localization. For instance, the 1000-km-wide deformation area in the Central Indian Basin is a diffuse, or nonlocalized, plate boundary [Gordon, 2000] but deformation within it is concentrated on localized faults [Weissel et al., 1980]. Faults themselves reveal a fluid-like gouge [Sibson, 1977]. At yet smaller scales, localized slip surface and microcracks are apparent [Simpson, 1984; Scholz, 1990]. In this paper, we address localization from the point of view of large-scale tectonics. When deriving the effective stress exponent, we consider a micromechanical localization process, but we are interested only in its apparent behavior at larger scales. In this paper, the effective stress exponent characterizes a material with a fault or a shear zone in it, not the fault or the shear zone itself.

[4] Faulting has been included in numerous geodynamical models but rarely in a self-consistent way. Frequently, faults are considered as predefined boundaries in otherwise elastic, viscous, or plastic models [Dahlen and Suppe, 1988; Melosh and Williams, 1989; Beaumont and Quinlan, 1994; Beekman et al., 1996; Zhong and Gurnis, 1996]. Although preexisting structures have a great importance in tectonic deformation, this approach lacks generality in that it concerns only a specific geometry and does not address the origin of these faults. In other studies, faulting is included only a posteriori, by applying a failure criterion to a continuum model [Anderson, 1905; Hafner, 1951; Comer et al., 1985; Zuber, 1995]. The results of these models are questionable beyond initial faulting as they neglect how the development of these faults influences the stress field [Buck, 1990; Schultz and Zuber, 1994; Gerbault et al., 1998]. Finally, a few studies applied slip line theory to lithospheric
deformation [Odé, 1960; Tapponnier and Molinar, 1976; Lin and Parmentier, 1990; Regenauer-Lieb and Petit, 1997]. In this approach, the trajectory of discontinuities of the flow field is predicted. However, slip line theory implies a continuum of slip lines rather than individual faults. Slip line solutions suffer from nonuniqueness and are restricted to rigid-plastic materials [Hill, 1950]. None of these approaches consider the dynamic evolution of stress field during the development of these faults.

Recently developed damage theories [Bercovici, 1998; Tackley, 2000] and elasto-visco-plastic models [Polisakov et al., 1994; Buck and Poliakov, 1998; Gerbault et al., 1999; Lavier et al., 1999; Branlund et al., 2000] do follow the dynamic evolution of localized deformation, but they require numerical methods to solve realistic problems. Hence the need is great for quasi-analytical analyses that provide physical foundations for more complex models. For instance, one- and two-dimensional shear zone models [Tackley, 1998] as well as a boundary layer theory of mantle convection including dynamic localization [Bercovici, 1993] were constructed using the self-lubricating rheologies introduced by Bercovici [1993]. The effective stress exponent defined in this study provides a general framework of localization that allows the comparison of analytical models, numerical models, and geological observations, although they may not use the same localization mechanism. A first quasi-analytical model that utilizes the concept of effective stress exponent is presented elsewhere [Montesi, 2002].

Before introducing the effective stress exponent, we define three categories of localizing behavior: inherited, imposed, and dynamic localization. The effective stress exponent is important in the context of dynamic localization. We present how the effective stress exponent relates to other approaches to localization, such as the bifurcation analysis [Rice, 1976;Needeleman and Tvergaard, 1992] and self-lubricating rheologies [Bercovici, 1993]. Then, after reviewing briefly the different microscopic mechanisms of localization, we present the mathematical expressions of the effective stress exponent for several of these mechanisms. We choose to group these mechanisms into (1) localization during failure, including the laws of elastoplasticity often studied in relation with the bifurcation approach to localization; (2) localization during fractional deformation, including rate- and state-dependent friction laws and processes linked to the granular aspects of fault gouge; and (3) localization in the ductile regime, including shear-heating or grain-size feedbacks, which might contribute to the formation of ductile shear zones at deeper levels in the lithosphere.

2. Effective Stress Exponent and Other Concepts of Localization

2.1. Types of Localization

We define three ways by which localized deformation may occur. In the first, termed inherited localization, the deforming medium is initially heterogeneous. Localization is controlled by the preexisting structure of the lithosphere, especially by preexisting zones of weakness such as alteration and damage zones, plutonic intrusions, sutures, or greenstone belts. The second, imposed localization, occurs when the medium is homogeneous, but its boundaries are not, resulting in local stress enhancements. Again, localization depends on the preexisting configuration. For instance, major faults in Tibet originate at the corners of the Indian tectonic indenter [Tapponnier and Molinar, 1976]. Finally, in dynamic localization, deformation localizes as the properties of the medium evolve the deformation. Dynamic localization requires that deformation is “easier” at locations where deformation has been enhanced. We quantify what “easier” means with the effective stress exponent. Dynamic localization does not require that the geometry of the localizing material or its surroundings be specified ad hoc. Certainly, the lithosphere is not a perfectly uniform medium, and it is not stressed uniformly, but we assume that at large scale these heterogeneities are sufficiently regular to appear as a uniform fabric. In that case, dynamic localization results in one of the embedded heterogeneities dominating the deformation.

2.2. Localization and Effective Stress Exponent

Dynamic localization occurs via a feedback between a material’s rheology and its deformation field. In general, the rheology relates a set of variables \( \{\chi_i\} \), the rheological system, that describe the state of a parcel of the material, to its strength \( \sigma \):

\[
\sigma = \sigma(\{\chi_i\}) .
\]

The most commonly used variables \( \chi_i \) include strain, strain rate, chemical composition, temperature, and pressure. They may also include variables describing the integrated history or, equivalently, the current physical or chemical state of the material element. Here, each \( \chi_i \) and \( \sigma \) are scalars which may be invariants of more general tensorial quantities.

One of these variables, \( \chi_0 \), is identified as the localization quantity. Dynamic localization of \( \chi_0 \) occurs if, when \( \chi_0 \) is perturbed from an equilibrium value by \( d\chi_0 \), the system of internal variables adjusts in such a way that an additional increment of \( \chi_0 \) is generated, which pushes \( \chi_0 \) even further from its original equilibrium value. In particular, if both \( \sigma \) and \( \chi_0 \) are positive, the response of the material that brings localization is a weakening one; the sign of the stress changes, \( dr \), is opposite to the sign of \( d\chi_0 \).

For the case where \( \chi_0 \) is the plastic strain \( \varepsilon_p \), Drucker [1952] proposed that a material is stable if the work done by an increment of strain is positive, \( d\varepsilon_p d\sigma_p > 0 \). Localization occurs otherwise. This criterion was subsequently extended by [Hill, 1958] and Drucker [1959]. On the other hand, Mandel [1966] favored the use of the tangent modulus \( A \), a fourth-order tensor that relates increments of strain and stress, in a criterion for localization. Stability is ensured if \( A \) is positive definite. A stable material in the sense of Mandel is also stable in the sense defined by Drucker [1952, 1959]. However, Mandel’s definition implies that a material is unstable if any eigenvalue of \( A \) is negative, even if Drucker’s criterion for stability is verified. Identifying localization with material instability, Mandel’s criterion is a necessary condition for localization, whereas Drucker’s criterion is sufficient for localization. Rudnicki and Rice [1975] and Rice [1976] extended the definition of \( A_0 \) to include the effects of a planar discontinuity imbedded in the deforming medium. When \( |A| = 0 \), the deformation field undergoes a bifurcation expressed by different deformation states across the discontinuity. Localization arises through the coupling between different components of the strain and stress increment tensors by the discontinuity, although the constitutive behavior of the material is stable.

These analyses indicate that not all the components of the strain and stress tensors become unstable under the same conditions and that the coupling between the variables involved in the rheology is important. Building on the work by Mandel [1966] and Rudnicki and Rice [1975], we consider that two scalar measures of the strain and stress tensors localize when the tangent modulus relating them, considering coupling of other internal variables or geometrical relations such as potential discontinuity, indicates weakening. For instance, we refer to the localization of the shear strain \( \varepsilon_{xy} \) with the normal stress \( \sigma_{nn} \), both being positive values if \( d\varepsilon_{xy} d\sigma_{nn} < 0 \). Coupling is implicit in the use of the total derivative if more than one variable appears in equation (1). Moreover, the use of strain in the localization criterion is generalized to any quantity \( \chi_0 \) of the rheological system. Although we present only examples where \( \chi_0 \) is the strain or the strain rate in this paper, it is conceptually any variable, such as the temperature or the chemistry of a pore fluid.

The generalized tangent modulus, \( A = d\varepsilon/d\chi_0 \), is computed from the rheology (equation (1)), where \( \sigma \) is either a component or an
invariant of the stress tensor. Rephrasing Mandel’s [1966] criterion, which assumes \( \sigma > 0 \), quantities \( \chi_0 > 0 \) for which \( A < 0 \) localize.

[13] Beyond a general criterion for localization, we seek a measure of localization efficiency to compare different localization processes. Hence we scale the generalized tangent modulus by the current state of deformation, \( \sigma \) and \( \chi_0 \), and define the effective stress exponent, \( n_\varepsilon \), by

\[
\frac{1}{n_\varepsilon} \equiv \frac{\chi_0}{\sigma} \frac{d\sigma}{d\chi_0}.
\]  

(2)

The effective stress exponent gives the relative importance of the dynamic evolution of strength with respect to its current values. Therefore dynamic localization occurs for negative \( n_\varepsilon \) and is strongest for more negative \( 1/n_\varepsilon \).

[14] When localization is weak (\( 1/n_\varepsilon \to 0 \)), it has little effect on the stress field; localized shear zones may be traced a posteriori on a stress field derived from continuum mechanics. If, on the other hand, localization is strong (\( 1/n_\varepsilon \to -\infty \)), localization dominates the deformation. That regime is not accessible by continuum mechanics. The perfectly plastic limit, where strength is invariant of \( \chi_0 \), corresponds to \( 1/n_\varepsilon = 0 \). It bridges the localizing and nonlocalizing regimes. The condition \( 1/n_\varepsilon = 0 \) also indicates the bifurcation points of the deformation history, where two deformation states coexist, one being continuous, the other may have an active discontinuity [Rudnicki and Rice, 1975; Rice, 1976; Needleman and Tvergaard, 1992]. Hobbs et al. [1990] pointed that in that case, localization requires dynamic weakening of the deformation field that includes an active discontinuity, i.e., \( 1/n_\varepsilon < 0 \) for the relationship between the loading and strain when the discontinuity is active.

[15] When \( 0 < 1/n_\varepsilon < 1 \), the ratio \( \sigma/\chi_0 \) decreases with \( \chi_0 \). This ratio correspond to the apparent modulus or apparent viscosity if \( \chi_0 \) is a strain or a strain rate, respectively. Therefore the material softens where \( \chi_0 \) is increased, so that further increases of stress result in enhanced deformation in that region. Changing \( \chi_0 \) by \( d\chi_0 \) creates a heterogeneity that can bring inherited localization (section 2.1) even though the material strengthens and dynamic localization is not possible. We call the softening-related localization that is possible at \( 0 < 1/n_\varepsilon < 1 \) progressive localization. It may result in a localized shear zone, but unlike dynamic localization, the stress in that shear zone is higher than in its surroundings. The stress in some natural ductile shear zones is indeed higher than in its surrounding [Jin et al., 1998], and localization in the ductile field may be progressive rather than dynamic, as our micromechanics-based approach also indicates (section 7).

[16] Although Smith [1977] did not consider the possibility of \( n_\varepsilon \) being negative, he showed that a general nonlinear viscoelastic rheology is characterized by an effective stress exponent. If \( \eta \equiv \sigma/\varepsilon \) is the (secant) viscosity at a given strain rate, small perturbations of strain rate behave as if they obey an effective viscosity \( \eta n_\varepsilon \) [Smith, 1977]. If \( n_\varepsilon < 0 \), the negative effective viscosity of these perturbations results in a weaker perturbed material.

[17] The use of the total derivatives \( d\sigma/d\chi_0 \) in (2) indicates that although we address the localization of a particular variable \( \chi_0 \), the response of the whole system of internal variables \( \{\chi_i\} \) is considered (Figure 1). We define the apparent rheology as the relation between \( \chi_0 \) and \( \sigma \) that takes into account the response of the full system \( \{\chi_i\} \) to changes in \( \chi_0 \). In practice, it is often required to truncate the system to a limited number of suitably chosen variables. Previously, Poirier [1980] and Hobbs et al. [1990] emphasized the difference between the direct response of the strength to a variation of \( \chi_0 \) and the total response of the system \( \{\chi_i\} \). In most of the cases presented below, deformation does not localize solely from the direct relation between \( \chi_0 \) and \( \sigma \), which is usually strengthening, but requires the feedback of other internal variables (Figure 1). Some of the coupling may result from the geometry of the material considered, especially when a planar discontinuity is embedded in the material [Rudnicki and Rice, 1975; Rice, 1976].

[18] In fact, only a subset of the system of internal variables depends on \( \chi_0 \). For those, it is possible to write \( \chi_1 \) as a function of \( \chi_0 \) and to define \( d\chi_1/d\chi_0 \). The other variables cannot participate in the localization of \( \chi_0 \) and may be considered constant as the perturbation is applied. However, these variables may be important in determining the strength \( \sigma \) and therefore the intensity of localization. As an example, let us consider a system where the strength depends on the strain rate, temperature, and pressure and examine its stability with respect to the strain rate. If the strain rate increase, the temperature increases as well because a fraction of the mechanical work is converted into heat, but the pressure may be buffered. Therefore the derivative in (2) includes the variation of strength with strain (a strengthening term) and temperature.

**Figure 1.** Yield surface \( \sigma(\chi_0,\chi_1) \), where \( \chi_0 \) and \( \chi_1 \) are two internal variables that represent for instance strain rate and temperature. The strength increases if only \( \chi_0 \) is perturbed (dashed arrow, \( d\sigma/d\chi_0 > 0 \)), but the system response may include also a change of \( \chi_1 \) induced by \( \chi_0 \), so that the total response is weakening (thick arrow, \( d\chi_1/d\chi_0 < 0 \)). The darker region of the yield surface indicate negative stress exponent, or localization.
(a weakening term) but not with pressure, which is to a large extent uncoupled from the strain rate. Localization is possible if the temperature change dominates over the strain rate change, but the intensity of localization depends on the initial strength of the material, which depends on the pressure as well as the initial temperature and strain rate.

2.3. Arresting Localization

[19] Using an effective stress exponent, \( n_e \), to characterize the behavior of a rheological system suggests that its apparent rheology can be approximated by a power law relation \( \sigma \propto \dot{\varepsilon}^{n_e} \), ignoring the details of the coupling between the internal variables. Such a rheology follows the same spirit as, for instance, the dislocation creep law, a power law between stress and strain rate that does not explicitly the details of dislocation motions. While this approach can be useful in infinitesimal perturbation analyses [Smith, 1977; Montesi, 2002], one must exert caution when the system is locally far from its original configuration. Indeed, a power law with a negative stress exponent predicts that localization continues until one of the two singularities \( \sigma = \infty \) or \( \sigma = 0 \) is reached [Melosh, 1976; Tackley, 1998], resulting in one infinitely thin shear zone.

[20] However, the concepts of tangent modulus and effective stress exponent imply that the perturbation of the variable undergoing localization is infinitesimal. Therefore the evolution of \( n_e \) as localization progresses must be considered if the end result of localization is sought. For instance, the thickness of a shear band or a plate boundary may be dictated by the point where \( n_e \) becomes positive within the localized deformation zone. Once nonlinearities in the system are taken into account, the effective stress exponent of a localizing system certainly becomes positive before the unrealistic singularities mentioned above are reached. In addition, other parameters that may be considered fixed at the onset of localization might become important when localization is ongoing. For instance, grain growth [Kameyama et al., 1997] and volatile incorporation [Regenauer-Lieb, 1999] might stabilize shear zones formed by shear heating. Fault rotation may stabilize frictional sliding [Sibson, 1994]. The stress exponent must be negative only over a specific domain of the \( \chi_0 - \sigma \) space.

2.4. Self-Lubricating Rheology

[21] At the largest scale relevant for Earth sciences the global deformation field is localized at plate boundaries, which appear weaker than plate interiors. The associated viscosity heterogeneities allow for toroidal motion as well as the poloidal motion driven by convection in the mantle [Bercovici et al., 2000]. To produce dynamically the weakening associated with a plate boundary, Bercovici [1993, 1995] introduced a self-lubricating (modified Carreau) rheology:

\[
\eta = (\gamma + \dot{\varepsilon})^{\frac{(\gamma + 1)}{z}},
\]

where \( \dot{\varepsilon} \) is the second invariant of the strain rate tensor, \( \eta \) is the viscosity, and \( \gamma \) and \( r \) are constitutive parameters. The strength of such a material is \( \sigma = 2\eta \dot{\varepsilon} \) from which we compute

\[
\frac{d\sigma}{d\dot{\varepsilon}} = \frac{\sigma}{\dot{\varepsilon}} + \frac{(1/(\gamma - r))\eta^2}{\gamma + \dot{\varepsilon}^2}.
\]

Then the inverse effective stress exponent for self-lubricating rheologies is

\[
\frac{1}{n_e} = \frac{\gamma/z^2 + 1/r}{\gamma/z^2 + 1}.
\]

Dynamic localization occurs when \( 1/n_e < 0 \), or \( -\gamma/\dot{\varepsilon}^2 < 1/r < 0 \) (Figure 2).

![Figure 2. Inverse effective stress exponent for self-lubricating rheology with \( r = 100 \) (solid line), \( r = 1 \) (dashed line), \( r = -1 \), (dotted line), \( r = -100 \) (dash-dotted line). Circles show the limit of localizing domain when \( r \) is negative (\( \dot{\varepsilon}^2/\gamma = -r \)).](image)

[22] The self-lubricating rheology is actually an apparent rheology that approximates the weakening due to shear heating [Whitehead and Gans, 1974; Bercovici, 1993] or damage accumulation [Bercovici, 1998; Tackley, 1998, 2000]. An alternative effective rheology could be a power law relation between strain rate and stress, with stress exponent \( n_e \). However, self-lubricating rheologies have an advantage over such power laws in that they strengthen at low strain rate [Bercovici, 1995; Tackley, 1998]. Neither the power law nor the self-lubricating rheology are valid if, on the other hand, coupling is external, \( \chi = \chi_{nl} \), with \( \chi_{nl} \) as \( \chi_0 \). The effective stress exponent for the total system is:

\[
\frac{1}{n_e} = \frac{\gamma_0}{\chi_0} + \frac{\chi_{nl}}{\chi_0}.
\]

The effective stress exponents for the localizing and nonlocalizing parts of a rheological system are \( n_{nl} < 0 \) and \( n_{nl} > 0 \), respectively, defined by (2) using successively \( \chi_{nl} = \chi_{nl} \) and \( \chi_{nl} = \chi_{nl} \) as \( \chi_0 \). The effective stress exponent for the total system is:

\[
\frac{1}{n_e} = \frac{\gamma_0}{\chi_0} + \frac{\chi_{nl}}{\chi_0}.
\]

If, on the other hand, coupling is internal, \( \chi_{nl} = \chi_{nl} \), and the total stress of the material is the sum of a contribution from the
localizing system $\sigma' l$ and a contribution from the nonlocalizing system $\sigma'^{nl}$:

$$\sigma = \sigma'^{nl} + \sigma' l.$$  

Then the effective stress exponent becomes

$$\frac{1}{n_e} = \frac{1}{n'^{nl}} + \frac{1}{n'^{l}},$$  

[25] In particular case where $\chi_l$ is the strain the nonlocalizing part of the system may be elastic. The elastic strain $e_l$ is related to the strength by

$$\sigma = Ge_l.$$  

where $G$ is an apparent rigidity modulus appropriate for the loading conditions. This gives $n'^{el} = 1$. When coupling is internal, the total strain, $e$, is the sum of $e_l$ and a strain from the localizing part of the rheological system, $e_l$. Equation (7) becomes

$$n_e = 1 + \frac{G}{e_l} (l'^{el} - 1) = n'^{el} + \frac{\sigma}{G} (1 - n'^{el}) .$$  

The inverse effective stress exponent $1/n_e$ is more negative for finite Gihan as it is the rigid limit ($G \to +\infty$): When coupled internally, elasticity enhances localization. Physically, the weakening associated with an increase of $e_l$ decreases the supported stress and with it the elastic strain. As the total stress is the localizing variable, this is compensated by higher $e_l$ and further weakening: The system is more unstable than in the absence of elastic effects.

When elasticity is coupled externally to a localizing system, $e = e_l = e'^{el}/G$, with $σ'$ the stress supported by the elastic system. The localizing system supports a stress $σ' l$. The total strength of the coupled system is $σ = σ' + σ' l$, and its effective stress exponent is

$$\frac{1}{n_e} = 1 + \frac{σ'}{σ} \left( \frac{l'^{el} - 1}{n'^{el} - 1} \right) = \frac{1}{n'^{el}} + \frac{Ge_l}{σ} \left( 1 - \frac{1}{n'e_l} \right).$$  

Although (12) is similar to (11) for an internal coupling, the coupled system does not localize if $G$ is large, i.e., when the elastic subsystem is too stiff to accommodate the enhanced deformation. Indeed, stiff machines have been used in experimental rock mechanics to determine the full-strain-stress curves around yield in spite of the instability related to strain weakening [Cook, 1981].

3. Microscopic Localization Mechanisms

[27] Many micromechanical processes occur in association with localized shear zones and faults in natural examples and laboratory studies. In section 3.1–3.4 we present the mathematical expressions for the effective stress exponents of these mechanisms. In many of these examples, localization does not arise solely from the rheology but also from the coupling of several internal variables. Therefore it is useful to first review how these processes lead to localization. We group the localization mechanisms according to whether they occur with brittle failure, elastoplastic deformation, frictional sliding, or ductile creep. Additional feedback processes can be imagined, such as enhancing localization, such as the development of anisotropy with strain [Poitier, 1980] or the growth of ductile voids [Bercovici et al., 2001b], others reducing it, such as strain hardening [Burg, 1999] or fault rotation [Sibson, 1994]. However, we limit this study to the processes discussed below. The parameters and variables involved in the localization processes considered herein are compiled in Table 1.

3.1. Failure-Related Processes

[28] Failure of intact rocks is associated with localized features such as cracks, faults, and plastic shear bands. Upon failure, a rock may lose some of its strength, either by growth of microcracks in low-porosity rocks or breakdown of a diagenetic matrix for less consolidated sedimentary rocks [Paterson, 1978; Lockner, 1995]. In particular, the cohesion of the rock is reduced. Weakening in that type of brittle failure is explicit, and it may seem redundant to compute effective stress exponents in this case. We will do so only for comparison with other mechanisms and because the mathematical simplicity of this example provides a good introduction to the procedure involved in computing effective stress exponents.

3.2. Localization in Elastoplastic Materials

[29] Upon failure, a material may deform by a combination of elastic and plastic strains. Even if plastic flow does not reduce the yield strength, the combination of elastic and plastic deformation can produce a bifurcation in the deformation field and result in localized deformation [Rudnicki and Rice, 1975; Needleman and Tvergaard, 1992]. We will show that an elastoplastic material may have negative effective stress exponent even though its yield strength is not reduced.

Localization in elastoplastic materials occurs because the volume changes associated with loading are different from that provided by increments of plastic strain. The plastic volume change and shear strain are related by the dilation angle $\varphi$, whereas the loading shear stress and pressure are related through the yield criterion by the angle of internal friction, $\varphi$. Unlike the majority of metals considered in mechanical engineering, $\varphi \neq \psi$ for rocks [Mandl, 1988] (however, Mandl [1988] noted that conjugate faulting brings an additional degree of freedom, so that an apparent dilation angle at large scale may be such that $\psi = \varphi$). As $\psi$ is smaller than $\varphi$, shearing on a failure plane does not produce enough dilation, resulting in elastic unloading of the normal stress [Vermeer and de Borst, 1984]. Therefore plastic shearing leads to more failure, and deformation localizes. The occurrence and efficiency of localization depend on the elastic properties of the material as well as on the plastic flow laws.

[31] Because the plastic strain is the integral of strain increments that occur when the yield criterion is verified [Hill, 1950; Lubliner, 1990], the quantities $\sigma$ and $\chi_l$ tested for localization depend on the loading history of the material and on the angle $\theta$ between the principal directions of stress and the plane along which deformation localizes. Fractures in rocks are observed at $\theta = 0$ (axial splitting) at low confining pressure [Griggs and Handin, 1960; Paterson, 1978] and at $\theta_0 = \pi/4 - \varphi/2$, which is the angle where the failure criterion is maximum [Anderson, 1951; Sibson, 1994]. However, in laboratory experiments using sand, shear bands are also observed at the angles $\theta_0 = \varphi - \psi/4$ and $\theta_0 = \pi/4 - (\varphi + \psi)/4$, which correspond to the directions where the strain increment in the shear band is collinear to the loading direction and where the postbifurcation macroscopic weakening is maximum, respectively [Vermeer, 1990]. Recently, compaction bands ($\theta = \pi/2$) have been observed in the field and in the laboratory [Antonellini et al., 1994; Olsson, 1999]. We will show that the effective stress exponent either changes sign or passes through extreme at each of these angles.

[32] Bifurcation analysis predicts that discontinuities in an elastoplastic material arise spontaneously at given points in the loading [Rudnicki and Rice, 1975; Rice, 1976]. However, to be visible at large scale, the discontinuous state must be weaker that...
Table 1. Parameters and Variables Used

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Symbol</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasticity and brittle failure</td>
<td>µ</td>
<td>coefficient of friction</td>
</tr>
<tr>
<td>Friction</td>
<td>µ</td>
<td>coefficient of friction</td>
</tr>
<tr>
<td></td>
<td>d, s</td>
<td>slip</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>time</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>gouge volume</td>
</tr>
<tr>
<td></td>
<td>σ</td>
<td>shear stress</td>
</tr>
<tr>
<td></td>
<td>σn</td>
<td>normal stress</td>
</tr>
<tr>
<td></td>
<td>pγ</td>
<td>pore fluid pressure</td>
</tr>
<tr>
<td></td>
<td>σe</td>
<td>effective normal stress</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>shear-induced heat</td>
</tr>
<tr>
<td>Ductile regime</td>
<td>σ</td>
<td>stress</td>
</tr>
<tr>
<td></td>
<td>ε</td>
<td>total strain</td>
</tr>
<tr>
<td></td>
<td>εD</td>
<td>strain rate accomodated by dislocation creep</td>
</tr>
<tr>
<td></td>
<td>εG</td>
<td>strain rate accomodated by diffusion creep</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>grain size</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>equilibrium grain size</td>
</tr>
<tr>
<td></td>
<td>η</td>
<td>viscosity</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>strain</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>temperature (with and without dimension, respectively)</td>
</tr>
<tr>
<td></td>
<td>H, T</td>
<td>heat generation (with and without dimension, respectively)</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>time</td>
</tr>
<tr>
<td>Parameters</td>
<td>γ</td>
<td>characteristic square strain rate</td>
</tr>
<tr>
<td>Self-lubrication</td>
<td>r</td>
<td>weakening coefficient</td>
</tr>
<tr>
<td>Loss of cohesion</td>
<td>σi</td>
<td>initial strength</td>
</tr>
<tr>
<td></td>
<td>σf</td>
<td>final strength</td>
</tr>
<tr>
<td></td>
<td>Δσ</td>
<td>stress loss</td>
</tr>
<tr>
<td></td>
<td>εc</td>
<td>characteristic plastic strain for cohesion loss</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>elastic modulus</td>
</tr>
<tr>
<td>Elastoplasticity</td>
<td>φ</td>
<td>internal friction angle</td>
</tr>
<tr>
<td></td>
<td>c</td>
<td>cohesion</td>
</tr>
<tr>
<td></td>
<td>v</td>
<td>dilation angle</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>ratio of Lamé parameters</td>
</tr>
<tr>
<td></td>
<td>v</td>
<td>ratio of first to second invariant of strain increment</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>ratio of first to second invariant of stress increment</td>
</tr>
<tr>
<td></td>
<td>θ</td>
<td>angle between the reference axes (plane of discontinuity) and principal strain (stress) direction</td>
</tr>
<tr>
<td>Friction</td>
<td>µv</td>
<td>static coefficient of friction</td>
</tr>
<tr>
<td></td>
<td>µd</td>
<td>dynamic coefficient of friction</td>
</tr>
<tr>
<td></td>
<td>Ds, D</td>
<td>characteristic sliding distances</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>external stiffness</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>velocity dependence coefficient</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>state dependence coefficient</td>
</tr>
<tr>
<td></td>
<td>c</td>
<td>steady state velocity weakening coefficient</td>
</tr>
<tr>
<td></td>
<td>Kf</td>
<td>fluid capacity</td>
</tr>
<tr>
<td></td>
<td>σt</td>
<td>heat transfer coefficient</td>
</tr>
<tr>
<td></td>
<td>td</td>
<td>heat diffusion timescale</td>
</tr>
<tr>
<td>Shear heating</td>
<td>B</td>
<td>preexponential factor</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>stress exponent</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>activation energy</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>gas constant</td>
</tr>
<tr>
<td>Grain-size feedback</td>
<td>n</td>
<td>stress exponent (dislocation creep)</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>grain-size dependence exponent (dislocation creep)</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>stress exponent (dislocation creep)</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>preexponential factor (dislocation creep)</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>stress exponent of the equilibrium grain size</td>
</tr>
<tr>
<td></td>
<td>D0</td>
<td>preexponential factor of equilibrium grain size</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ratio of diffusion over dislocation creep</td>
</tr>
<tr>
<td></td>
<td>R0</td>
<td>R for grain size d0</td>
</tr>
<tr>
<td></td>
<td>εs</td>
<td>characteristic strain for grain-size evolution</td>
</tr>
</tbody>
</table>

3.3. Friction and Gouge Processes

Once a fault is formed, it slides at a stress level dictated by the laws of frictional sliding. With its long geological history, the brittle outer layer of the Earth, or schizosphere [Scholz, 1990], is riddled with faults to the point that it may be viewed as a continuum obeying the laws of friction at the tectonic scale [Brace and Kohlstedt, 1980]. However, not all faults are equally active and localization for a frictional material is expressed as an increase of the deformation taken by a given fault. Thus localization is insured by the relative weakness of the most active faults. Indeed, laboratory studies indicate that active faults are weaker than inactive ones [Rabinowicz, 1951]. At the geological scale, some studies indicate that major faults such as the San Andreas Fault may be weaker than their surroundings [Zoback et al., 1987], although a consensus is still lacking [Scholz, 2000].

A probable microscopic explanation of why active faults are weak involves the evolution of fault gouge [Scholz, 1990]. As fault gouge becomes induced and stronger if stationary, activity on fault, quantified by the fault offset, reduces its coefficient of friction [Rabinowicz, 1951]. Alternatively, the combination of time-dependent healing and strain-dependent weakening produces apparent strain rate weakening [Dieterich, 1978], and fault “activity” may be better described by a sliding velocity rather than a shear strain since activation [Scholz, 1990].

The most successful constitutive laws of friction to date are the rate- and state-dependent friction laws (RSDF) [Scholz, 1990]. Beyond steady state velocity weakening or strengthening effects, RSDF laws produce transient effects that are important for earthquake mechanics [Dieterich, 1992; Marone, 1998] and make the effective stress exponent time-dependent. This is because they involve two rheological variables: the instantaneous sliding velocity and a state variable that evolves either with slip [Ruina, 1983] or time [Dieterich, 1979].

After giving the expressions for the effective stress exponent of a material that obeys the RSDF laws, we will address two physical processes in the granular fault gouge that may be at the origin of the state variable evolution. The first is purely mechanical. The fault gouge dilates as it shears, requiring more work than needed to overcome only the frictional resistance. Therefore the apparent coefficient of friction is higher than actual one [Frank, 1965; Orowan, 1966]. Localization can occur if the dilation rate decreases with shear, so that the apparent coefficient of friction on decreases, although the actual coefficient of friction does not. However, this process needs to be modulated by the effect of fluids within the gouge that, if undrained, resist dilation. The second localization mechanism is thermal. The gouge, and especially its fluid portion, dilates when heated. Hence the heat produced by shearing may weaken the fault [Shaw, 1995] and lead to localization.

3.4. Ductile Mechanisms

At sufficiently high temperatures, rocks flow like viscous fluids through dislocation and diffusion creep. Although experimentally derived rheologies are usually strain rate strengthening [Evans and Kohlstedt, 1995], localized shear zones are a common occurrence in ductile rocks [Ramsey, 1980]. Localization is made possible by the response to a perturbation of strain rate of internal variables other than the strain rate. Two possible localization mechanisms are shear heating and recrystallization, where localization occurs through a feedback in temperature or grain size.
Because rocks are weaker at high temperature and shearing produces heat, temperature and deformation rate can feed back on one another to produce localization in the ductile regime. Shear heating has been proposed to explain ductile shear zones [Brun and Cobbold, 1980; Fleitout and Froidevaux, 1980; Hobbs et al., 1986] and has been studied in relation to plate formation and orogeny [Froidevaux and Schubert, 1975; Melosh, 1976; Bercovici, 1993; Schott et al., 1999]. Heat must accumulate to overcome the direct strain rate strengthening of the creep laws [Poirier, 1980; Hobbs et al., 1986], so that localization though shear heating is favored by near-adiabatic conditions [Bai and Dodd, 1992]. However, adiabatic conditions are an instantaneous approximation of heating, whereas a finite time is required to accumulate sufficient heat in a potential shear zone. Therefore a high heating rate is needed, which is favored by low temperature and high strain rate. In section 7.1 we show that a preexisting heterogeneity such as a brittle fault is needed for shear heating to localize deformation. However, we ignore the additional coupling of grain size evolution [Kameyama et al., 1997] or stored elastic energy [Regenauer-Lieb and Yuen, 1998] to shear heating, which could make localization easier. Regenauer-Lieb and Yuen [1998] show in particular that when elasticity is considered, localization can be so fast that an adiabatic approximation is reasonable.  

A different process leading to localization in the ductile regime is the interplay of grain size and grain-size-sensitive diffusion creep. The transition from grain-size-insensitive dislocation creep to diffusion creep has been observed in natural shear zones, concurrent to grain-size reduction [Handy, 1989; Jaroslaw et al., 1995; Jin et al., 1998]. However, it is important to note that although the combination of diffusion and dislocation creep mechanisms results in a material that is weaker than if only dislocation creep operated, each mechanism is strain rate strengthening, so that localization is not ensured. In fact, as long as the grain size is constrained to follow its recrystallized equilibrium value, localization is not predicted from our analysis (section 7.2.1). As in natural shear zones, stress increases with localization [Jin et al., 1998], localization may indeed be progressive rather than dynamic (section 2.2). However, dynamic localization is possible if the grain size is initially out of equilibrium and evolves toward the recrystallized equilibrium value at a deformation-controlled rate (section 7.2.2). The departure from the equilibrium grain size can result from change of tectonic environment or transitions in microstructure [Tullis et al., 1990; Rutter, 1999] due, for instance, to metamorphism or neocrystallization [White and Knipe, 1978; Beach, 1980; Rubie, 1983; Brodie and Rutter, 1985; Fitz Gerald and Stun滋t, 1993; Brown and Solar, 1998; Newman et al., 1999].

4. Effective Stress Exponents for Brittle Failure

A rock may be loaded elastically up to a yield stress $\sigma_y$ beyond which the total strain $\varepsilon$ is the sum of the elastic strain $\varepsilon^e = \sigma / G$ and a plastic strain $\varepsilon^p$: 

$$\varepsilon = \varepsilon^e + \varepsilon^p.$$  

The elastic strain $\varepsilon^e = \sigma / G$, with $G$ a general elastic modulus as in (10).

Deformation commonly localizes as plastic strain accumulates, especially when plastic flow is accompanied by a decrease of the yield strength from $\sigma_y$ to a final strength $\sigma_f = \sigma_y - \Delta \sigma$ over a critical plastic strain $\varepsilon_c$. The weakening may betray a loss of cohesion in the rock and could be linear

$$\sigma = \sigma_y - \Delta \sigma \varepsilon / \varepsilon_c$$  

which may be expected as (15a) is an approximate version of (15b) for $\sigma \approx \sigma_y$.

The effective stress exponent is negative only if $\varepsilon_c$ is greater than $\Delta \sigma / G$, the elastic strain corresponding to the unloading by $\Delta \sigma$. Therefore elastic unloading must occur faster than plastic deformation for deformation to localize. As the coupling between elastic and plastic strains is internal, elasticity enhances localization (section 2.5).
For $\Delta \sigma = 20$ MPa and $G = 30$ GPa, the minimum $\varepsilon_c$ is of order $0.6 \times 10^{-7}$. Buck and Poliakov [1998] used $\varepsilon_c = 0.03$ to 0.3, or $n_c = -450$ to $-45$, to study localization of deformation at mid-ocean ridge spreading centers and produced realistic looking abyssal hills.

5. Effective Stress Exponent for Elastoplastic Materials

5.1. Flow Theory of Plasticity

Before addressing localization for elastoplastic materials, we find it useful to review the basics of the flow theory of plasticity in two dimensions [Hill, 1950; Lubliner, 1990]. First, we define the stress and strain vectors

$$\sigma = \begin{pmatrix} \sigma_{xx} \\ \sigma_{xy} \\ \sigma_{yx} \\ \sigma_{yy} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{xy} \\ \varepsilon_{yx} \\ \varepsilon_{yy} \end{pmatrix}. \quad (17)$$

The strain in (17) is the total strain, the sum of an elastic strain, which is related to the instantaneous stress by the stiffness matrix, $D$, and a plastic strain. The plastic strain accumulates when a yield criterion is verified. Here we assume that yielding follows a Mohr-Coulomb failure law:

$$f = \sigma_{II} - \sigma_1 \sin \varphi - c \cos \varphi = 0, \quad (18)$$

where $\sigma_1$ and $\sigma_{II}$ are the first and second invariants of the stress tensor, $\varphi$ is the internal friction angle, and $c$ is the cohesion (Figure 4) [Vermeer and de Borst, 1984]. The more familiar criterion $f = \sigma_1 - \sigma_{II} \tan \varphi - c$, with $\sigma_1$ and $\sigma_{II}$ the shear and tangential stresses on a failure plane, is equivalent to (18), assuming that there exists a fabric of planes in the material with their normal oriented at $\theta_0 = \pi/4 - \varphi/2$ from the direction of the least compressive stress, the most favorable orientation for failure. If the fabric is pervasive, the material is considered a continuum. We do not include explicit hardening or weakening in the failure criterion.

The direction of plastic strain increments is determined by the flow potential $g$; plastic strain increments are proportional to $\nabla g$, where $\nabla$ denotes the gradient taken in an orthonormal base describing the space of stress vectors. It is customary to take

$$g = \sigma_{II} - \sigma_1 \sin \varphi, \quad (19)$$

where $\varphi$ is the dilation angle [Vermeer and de Borst, 1984; Mandl, 1988]. The strain field resolved on a plane includes generally shear and dilatant motion. Hence the plastic strain increment is coaxial to the stress increment only if the plane is at an angle $\theta = \pi/4 - \varphi/2$ from the least compressive stress (Figure 4a). If plasticity is nonassociated ($\varphi \neq \psi$), that plane is different from the failure plane. Therefore the strain and stress increments are incompatible, which leads to localization [Rudnicki and Rice, 1975; Vermeer and de Borst, 1984].

The amplitude of the elastoplastic strain increment is such that the material stays at yield. Hence, increments of stress and total strain are related by $d\sigma = M \cdot de$ (Appendix A1) [Vermeer and de Borst, 1984], with

$$M = D - \left[1 - \frac{\nabla g}{\varphi} \cdot \nabla g \right] \cdot D \cdot \nabla g, \quad (20)$$

5.2. Localization in a Continuous Elastoplastic Medium

We first address localization in a continuous medium loaded elastically to yield. Potential discontinuities will be considered in section 5.3. Loading is achieved by cumulative strain increments $de = E \cdot de$, where $E$ is a shape vector and $de$ is the magnitude of each increment. The second strain invariant is used as $d\sigma$, and $E$ is parameterized by $\nu$, the ratio of volume change to $de$, and by the angle $\theta$ between the reference axes and the principal directions of strain:

$$E = \frac{1}{2} \begin{pmatrix} \nu + \cos 20 \\ \sin 20 \\ \nu - \cos 20 \\ \sin 20 \end{pmatrix}. \quad (21)$$
Loading in pure shear corresponds to \( v = 0 \) and uniaxial loading corresponds to \( v = 1 \) (one principal strain increment is null). Using the elastic stiffness matrix \( \mathbf{D} \) and a possible prestress \( \sigma_0 \), the stress at yield is \( \sigma_0 + n \mathbf{D} \cdot \mathbf{E} \mathbf{d} \), with \( n \) the number of increments needed to reach the yield point. Any value of \( v \) is admissible if the cohesion is nonzero (Figure 4b). The Lamé parameters \( \lambda \) and \( G \) and their ratio \( r = \lambda / G \) characterize \( \mathbf{D} \) (Appendix A1). No localization is possible during the elastic loading stage \( (n = 1) \).

At yield, a new strain increment brings a stress increment \( d\sigma = \mathbf{M} \cdot \mathbf{E} \mathbf{d} \), with \( \mathbf{M} \) defined in (20). As \( \sigma \) is a tensor, we must define a measure \( s \) against which we test for localization. Hence the effective stress exponent \( n_s \) measures the possible weakening of \( s \cdot \sigma \) that arises from a change of \( \mathbf{d} \). Using (2) and (20), we obtain

\[
\frac{1}{n_s} = \frac{s \cdot \mathbf{M} \cdot \mathbf{E}}{\delta_{xx} + s \cdot \mathbf{D} \cdot \mathbf{E}}.
\]

(respectively). Each effective stress exponent addresses the localization of the external strain imposed on an elastoplastic material, in relation to a given internal measure of stress. Hence it does not concern the behavior of a potential shear zone inside the material but rather how the material that includes such a zone is seen from a larger scale.

Localization of strain and \( \sigma_{xy} \) occurs when \( \sin \psi < v \), regardless of whether the material is associated or not; localization demands that the volume changes associated with shearing on the plane are different for the loading system (expressed by \( v \)) and the elastoplastic material (expressed by \( \sin \psi \)). In agreement with our result, Vermeer and de Borst [1984] determined that \( \psi \) must be negative (compaction) for localization to occur in a pure shear strain field \( (v = 0) \). For a Poisson solid \( (r = 1) \) with \( \varphi = 30^\circ \) and \( \psi = 0^\circ \), as is typical or rocks, and \( v \) between 0 and 1, \( 1/n_{xy} \) is between 0 and \(-1/2\).

Although localization using \( \sigma_{xy} \) does not depend on the orientation of shearing planes, localization using \( \sigma_{xx} \) does. Hence the apparent weakening of the material of the material is anisotropic. Figure 5 shows how \( n_{xy} \) varies with \( \theta \) for different values of \( v \). Axial splitting \( (\theta = 0) \) and compaction bands \( (\theta = \pi/2) \) are favored over shear failure, as \( n_{xy} \) passes through extrema at \( \theta = j\pi/2 \) with \( j \) an integer. Compaction bands are further unlikely as they require \( v < -1/(1 + r) \). On the other hand, axial splitting is a real possibility as it needs \( 1/(1 + r) > v \cdot \sin \psi \). However, the effect of the prestress remains to be determined: In experiments, axial splitting is prevented by a small confining pressure.

It may be argued that localization requires \( n_{xy} = n_{xx} < 0 \) as, in that case, the strength lost during localization is tensor coaxial with the stress increments during the loading. However, \( n_{xx} = n_{xy} < 0 \) is verified only if \( \sin \psi < \varphi < 1 + r < 1 \), which requires that plastic flow be compressive \( (\psi < 0) \).

5.3. Localization in an Elastoplastic Medium

With a Planar Discontinuity

We now consider the case of an elastoplastic material with a potential planar discontinuity, as in the bifurcation analysis or

\[
\begin{align*}
\sigma_{xx} &= D_{xx} + \sigma_{xx}^0, \\
\sigma_{yy} &= D_{yy} + \sigma_{yy}^0, \\
\sigma_{xy} &= D_{xy} + \sigma_{xy}^0,
\end{align*}
\]

Figure 5. Effective stress exponent of an elastoplastic material without discontinuity for localization of strain and either (left) shear stress or (right) normal stress. Stress is resolved on a plane with its normal at an angle \( \theta \) to the least compressive stress; \( v \) is the ratio of volumetric strain over the second invariant of strain. Solid line is \( v = 1 \); dashed line is \( v = 0.2 \); dotted line is \( v = 0 \); and dash-dotted line is \( v = -1 \). Prepared for \( \varphi = 30^\circ, \psi = 6^\circ, r = 1 \), and no prestress.

Rudnicki and Rice [1975]. The material is loaded elastically and uniformly to yield, at which point the discontinuity may separate a portion of the material that behaves elastoplastically from another that remains elastic [Vermeer, 1990]. Elastoplastic stress increments are given by (20). No explicit weakening or hardening is included in the flow law, but the activation of the discontinuity may reduce the overall strength of the material in which it is embedded; as in section 5.2, we address only the apparent behavior of the material as seen from a larger scale.

The loading is now defined by stress increments \( d\sigma = S d\sigma_{th} \), with \( d\sigma_{th} \) the increment of the second invariant of the stress tensor and \( S \) a shape vector parameterized by \( r \), the ratio between the variation of first and second stress invariants, and \( \theta \) the angle between the normal of the plane and the direction of least compressive stress (Figure 4b):

\[
S = \frac{1}{2} \begin{pmatrix}
\rho - \cos 2\theta \\
\sin 2\theta
\end{pmatrix}.
\]

We define a reduced plastic increment matrix \( \mathbf{M}' \) that relates the free components of strain increments within the band, \( d\varepsilon_{xx} \) and \( d\varepsilon_{xy} \), to \( d\varepsilon_{th} \). The other component of the strain increment tensor, \( d\varepsilon_{xy} \), and the components \( d\varepsilon_{xx} \) and \( d\varepsilon_{xy} \) of the stress increment tensor are transmitted across the potential discontinuity (Appendix A2) [Rudnicki and Rice, 1975; Vermeer, 1990]. We then test for the localization of \( d\varepsilon_{xx} \) and \( d\varepsilon_{xy} \) for which we define the effective stress exponents \( n_{xx} \) and \( n_{xy} \) respectively.

Assuming no prestress \( (\varepsilon = D^{-1}, S d\sigma_{th}) \) and using \( d\varepsilon = \mathbf{M}' \cdot d\varepsilon_{th} \) in (2), we obtain (Appendix A2 and Figure 6)

\[
\begin{align*}
n_{xx} &= \left[ \cos^2 2\theta - (\sin \varphi + \sin \psi) \cos 2\theta + \sin \varphi \sin \psi \right] \\
&\quad + \frac{2 + \rho}{1 + r} \left[ (\cos 2\theta - \sin \varphi) (\cos 2\theta - \sin \psi) \right]^{-1}.
\end{align*}
\]
Inverse effective stress exponent $1/n_{xy}$

![Inverse effective stress exponent $1/n_{xy}$](image)

Figure 6. Contours of inverse effective stress exponent for localization of (a) shear strain and (b) normal strain for an elastoplastic material with a discontinuity at angle $\theta$ from the principal directions of loading. The loading parameter $\rho$ is ratio of first to second invariant of the stress perturbation. The thick lines mark changes of sign of $n_{xy}$, with contours at every 0.01 to a maximum of 2. The region of negative stress exponent is shaded. darkest shading shows where both $n_{xy}$ and $n_{xx}$ are negative. It is assumed that $r = 1$, $\varphi = 30^\circ$, and $\psi = 6^\circ$ and no prestress.

\[ n_{xx} = \{(1 + \rho) \cos^3 \theta + 2 (1 + \rho) \sin \varphi \sin \psi + \rho \sin \varphi - \sin \psi \cos 20 \]
\[ + 2 (1 + \rho) \sin \varphi \sin \psi + \rho (\sin \varphi - \sin \psi) \cos 20 \]
\[ + (\cos 20 - \sin \theta) (\cos 20 - \sin \psi) (r - (1 + \rho) \cos 20) \}^{-1}. \]

Discontinuity angle $\theta$ is most efficient at $\theta_0 = \pi/4 - (\varphi + \psi)/4$ where $1/n_{xx}$ is minimum [Vermeer, 1990]. The localization of $n_{xx}$ is strongest localization where $1/n_{xx} \to -\infty$. However, this drastic localization may not occur spontaneously because $n_{xy}$ is positive for these values (not all components of the deformation field tend to localize) and, as it is not related to the condition for spontaneous discontinuity formation, $|M'| = 0$ [Rudnicki and Rice, 1975], the required discontinuity may be unavailable in the material. If there is a preexisting discontinuity at the angle for which $1/n_{xy} \to -\infty$, the drastic localization of the normal strains may happen. This might explain the compaction bands documented by Antonellini et al. [1994] and Olsson [1999], as $1/n_{xy} \to -\infty$ at $\theta = 90^\circ$ for $\rho = 1$ and $r = 0$. However, this explanation is unlikely because compaction bands are observed in rocks where $\rho \neq 0$ and because the divergence of $n_{xy}$ is suppressed with confining pressure, whereas compaction bands are observed only with significant confining pressure [Olsson, 1999]. Compaction bands are more likely to originate at the vertex between two yield envelopes [Issen and Rudnicki, 2001; Wong et al., 2001], a feature of the yield criterion that we do not consider herein.

In Figure 7 we present a section through the maps of effective stress exponent of Figure 6 at $\theta = \theta_0$, where both $n_{xx}$ and $n_{xy}$ are negative and $1/n_{xy}$ is minimum. For realistic rock parameters ($r = 1$, $\theta = 30^\circ$, $\psi = 6^\circ$), $1/n_{xx}$ increases with $\rho$ from roughly $-0.0525$ to $-0.0175$, while $1/n_{xy}$ decreases to $-0.02$. A good value where both effective exponents are similar is $n_{xx} \sim -50$, and $\rho \sim 0.75$.

6. Effective Stress Exponents for Localization in Frictional Materials

6.1. Static and Dynamic Friction

[60] A frictional material is one that deforms by sliding on many faults, so that it may be viewed as a continuum with a strength obeying the friction laws. Its deformation field localizes as a particular fault becomes more active than the others. This requires that active faults be weaker than less active or inactive ones, possibly because the dynamic coefficient of friction, $\mu_d$, is lower than the static coefficient of friction, $\mu_s$ [Rabinowicz, 1951] proposed that the transition from $\mu_s$ to $\mu_d$ occurs over a critical sliding distance $D_C$. If the coefficient of friction decreases linearly with sliding distance $d_s$, we obtain

\[ n_s = 1 - \frac{\mu_s D_C}{\mu_s - \mu_d} d_s, \quad d < D_C, \]

with $d$ the imposed displacement, different from $d_s$, because of elastic deformation. This formula is similar to (14a) for localization upon failure (Figure 3) with the loss of strength interpreted as a decrease of coefficient of friction rather than as a loss of cohesion. In either case, external coupling with a stiff loading system can prevent localization (section 2.5), as is done experimentally to study the details of the weakening occurring upon failure or fault activation [Cook, 1981]. From (12), stabilization occurs when $k$, the rigidity of the loading apparatus, exceeds $(\mu_s - \mu_d)/D_C$, which is the rate at which $\mu$ decreases with displacement [Schoetz, 1990, equation 2–24].

[61] In order to model the healing faults as their activity ceases it may be better to associate the weakening of active faults with their sliding velocity $V$. Dieterich [1972, 1978] proposed that the coefficient of friction $\mu$ during sliding depends on the shearing velocity $V$ as

\[ \mu = \mu_0 - c \cdot \ln(V/V_0) \]

with $\mu_0$ and $V_0$ the reference coefficient of friction and velocity and $c$ a material constant. Using (2) with $V$ as $\chi_0$ and $\mu$ as $\sigma$, the effective stress exponent is
where \( D \) is a critical distance related to the macroscopic structure of the fault gouge and possibly to \( D_c \) encountered in section 6.1. The parameters \( V_0 \) and \( \mu_0 \) are reference values of sliding velocity and friction coefficient, and \( a \) and \( b \) are material parameters. In steady state, \( \theta = D/V \) in either case, and (29) becomes (27), with \( c = b-a \).

The rheological system involved in the RSDF laws is composed of the two variables, \( \theta \) and \( V \). The velocity is the quantity that localizes in out treatment of frictional materials obeying the RSDF laws. As the state variable depends on the integrated history of velocity, we must specify a study case, whereby a perturbation in velocity is sustained for all time \( t > 0 \). Initially, the state variable is at equilibrium. We obtain (Appendix B)

\[
\frac{1}{n_v} = \frac{1}{\mu} \left\{ a - b \frac{D}{V} \left[ 1 - \exp \left( -\frac{V}{D} \right) \right] \right\}.
\]

(31)

An illustrative example evolution of \( n_v \) upon perturbation of sliding velocity is given in Figure 8. Initially, the effective stress exponent is positive because of the stabilizing direct effect \( (a > 0) \). However, if \( b > a \), \( n_v \) changes sign after time \( t_c \) because of the evolution of the state variable, with

\[
t_c \sim -\frac{D}{V} \ln \left( 1 - \frac{a}{b} \right).
\]

(32)

For velocity to localize in a material obeying the RSDF laws the velocity perturbation must be held for at least \( t_c \). In the infinite time limit the effective stress exponent tends toward the value for steady state sliding (equation (28), \( n_v \sim -250 \)). This value is probably the most relevant for large-scale geodynamics, where long timescales are involved. However, earthquakes mechanics are dominated by the transient behavior \([\text{Dietrich}, 1992; \text{Marone}, 1998]\). If the RSDF includes more than one state variable \([\text{Gu et al.}, 1984; \text{Gu and Wong}, 1994; \text{Bian pied et al.}, 1998]\), it may be stable in the steady limit, but there can be a time window when the material is apparently weakening \([\text{Montesi et al.}, 1999]\).

### 6.3 Fault Gouge and Pore Fluid Effects

#### 6.3.1 Apparent coefficient of friction and dilatancy

One physical aspect of fault gouge mechanics that may explain the weakening of active faults is the dilation of gouge upon shearing \([\text{Frank}, 1965; \text{Orowan}, 1966; \text{Marone et al.}, 1990]\). Gouge dilation works against the effective normal stress, \( \sigma_e = \sigma_n - p_f \) where \( \sigma_n \) is the normal stress imposed on the fault and \( p_f \) is the pressure of a pore fluid within the gouge. The energy needed for dilation is provided by the loading, in particular by the shear stress \( \tau \). Therefore, the apparent friction coefficient \( \mu_a = \tau/\sigma_e \) is higher when \( \sigma_e < \sigma_n \) and lower when \( \sigma_e > \sigma_n \).
than the actual friction coefficient $\mu$. On the basis of such thermodynamic considerations, Frank [1965] proposed

$$\mu_u = \mu + \frac{\nu}{\gamma} \frac{dv/d\gamma}{1 - \mu \nu/\gamma},$$

(33a)

with $\nu$ the volume of the gouge and $\gamma$ the shear strain [see also Edmond and Paterson, 1972; Marone et al., 1990]. In contrast, Orowan [1966], taking into account the change of direction of the microscopic friction force between gains as they climb onto one another, derived

$$\mu_u = \frac{\mu + \nu/\gamma}{1 - \mu \nu/\gamma},$$

(33b)

Localization is possible if the dilation rate $dv/d\gamma$ decreases with strain [Frank, 1965; Edmond and Paterson, 1972; Fischer and Paterson, 1989; Marone et al., 1990].

[68] On the other hand, dilation decreases the pore fluid pressure and therefore increases the effective normal stress [Reynolds, 1885; Frank, 1965], thereby stabilizing slip [Rice, 1975; Rudnicki, 1979]. The capacity of the fluid is defined as $K_f = -dp_f/dv$. Physically, the fluid capacity represents not only the compres-
sibility of the fluid, which is very small if the pore fluid is liquid, but also the ability of the fluid to circulate in the gouge and to sustain pressure gradients cannot be sustained and $K_f$ tends toward zero, even if the fluid is incompressible.

[69] This rheological system is composed of the gouge volume, the shear strain, and the pore fluid pressure. We test for the localization of the shear strain, $\gamma$, with respect to the shear stress. In addition, we consider only the initiation of sliding, so that $\gamma = \pi/G$, with $G$ the shear modulus of intact gouge. Hence the effective stress exponents is

$$\frac{1}{n_e} = \frac{1}{G} \left( \frac{d\nu}{d\gamma} \sigma_e - \nu \frac{dp_f}{d\gamma} \right).$$

(34)

Replacing $\nu_0$ by (33a) and (33b), respectively, and using $K_f = -dp_f/dv$, we obtain

$$\frac{1}{n_e} = \frac{\sigma_e}{G} \frac{d^2\nu}{d\gamma^2} + \frac{K_f}{G} \frac{d\nu}{d\gamma} \left( \frac{\mu + \nu/\gamma}{1 - \mu \nu/\gamma} \right),$$

(35a)

for Frank’s [1965] theory, and

$$\frac{1}{n_e} = \frac{\sigma_e}{G} \frac{1 + \mu^2}{(1 - \mu \nu/\gamma)^2} \frac{d^2\nu}{d\gamma^2} + \frac{K_f}{G} \frac{d\nu}{d\gamma} \frac{\mu + \nu/\gamma}{1 - \mu \nu/\gamma},$$

(35b)

for Orowan’s [1966] theory. Figure 9 shows how the effective stress exponent varies in function of the dilation rate for several values of the parameter $d^2\nu/d\gamma^2 \sigma_e/K_f$. The dilation rate is limited to $dv/d\gamma > -\mu$ for Frank’s theory and $1/\mu > dv/d\gamma > -\mu$ for Orowan’s theory by the condition that $\nu_0 > 0$.

[70] If the gouge is fully drained ($K_f = 0$), localization requires that $d^2\nu/d\gamma^2 < 0$, i.e., that the dilation rate decrease with strain [Frank, 1965; Marone et al., 1990]. However, for Frank’s theory, the stabilizing effect of pressure changes brings an upper limit to the dilation rate that can localize. Beyond that, the pressure change dominates and localization is impossible. However, changes of pore fluid pressure can help the system localize if the gouge compacts. Indeed, the pore fluid pressure increases as the gouge compacts, supporting more of the normal stress. Hence localization is possible in a compactiongouge ($dv/d\gamma < 0$) if $d^2\nu/d\gamma^2 \sigma_e/K_f < \mu^2/4$ (Figure 9). Indeed, dynamic weakening of an undrained fault gouge undergoing compaction was observed by Blanpied et al. [1992]. Changes of pore fluid pressure has less effect in Orowan’s theory. In particular, they cannot change the domain of dilation rate that localize or not, although localization is generally less strong when the pore fluid effects are taken into account.

[71] More rigorous treatments of fluid flow in a deforming matrix exist [Rudnicki and Chen, 1988; Segall and Rice, 1995]. For instance, Sleep and Blanpied [1994] showed that compaction, resulting from ductile creep in the fault gouge driven by the overpressurization of the pore fluid, is sufficient to destabilize fault slip. At high temperatures, the matrix is assumed to deform viscously [McKenzie, 1984; Spiegelman, 1993; Bercovici et al., 2001]. In that case, porosity (or damage) and surface tension interact to concentrate a fluid phase to a localized band [Bercovici et al., 2001b]. How the band would affect the strength of the material in which it forms has not yet been discussed.

### 6.3.2. Pore fluid heating

[72] Pore fluids promote localization if their pressure increases with either strain or strain rate. In the previous paragraph, pressure increased if the gouge compacted. Alternatively, pore fluid pressure increases as the fluid dilate if it is heated by the deformation. Shaw [1995], following Sibson [1973] and Lachenbruch [1980], proposed that the normal stress $\sigma_n$ on a fault gouge decreases linearly with the heat $Q$

$$\sigma_n = \sigma_n^0 - \alpha Q,$$

(36)

with $\alpha$ a coefficient and $\sigma_n^0$ the value without deformation. Heat may diffuse away over the timescale $t_h$ but is replenished at a rate proportional to $\nu$, where $\nu$ is the shearing velocity on the fault. The approximations that $t_h$ is small (steady state conductive) or that $t_h$ is large (adiabatic heating) result in weakening with velocity $\nu$ or slip $s$, respectively [Shaw, 1995]. The corresponding expression for $\sigma$ are

$$\sigma = \frac{\mu \sigma_n^0}{1 + t_h \mu_0 \nu},$$

(37a)

$$\sigma = \mu \sigma_n^0 \exp(-\alpha \mu_0 s),$$

(37b)
The effective stress exponent is negative for all finite $\alpha$. As $\alpha$ is not determined, it is difficult to address whether elastic effects would stabilize this mechanism of localization or not.

7. Effective Stress Exponent for Localization During Ductile Creep

7.1. Shear Heating

7.1.1. General analysis. \cite{73} Shearing a rock at the rate $\dot{\varepsilon}$ and stress $\sigma$ (rock strength) releases energy by viscous dissipation, a fraction $\beta$ of which is converted into heat $H$:

$$H = \beta \sigma \dot{\varepsilon}. \tag{39}$$

As the strength of ductile rocks is temperature-activated, the heat anomaly due to a local increase of strain rate can weaken the rock and localize the strain rate. The rheology of ductile rocks is such that

$$\sigma = B \varepsilon^{1/n} \exp \left( 1/n0 \right), \tag{40}$$

where $n$ is the stress exponent (not to be confused with the effective stress exponent $n_e$) and $B$ is a material constant. The rheological temperature $\theta$ is defined as $\theta = TR_G/Q$, with $T$ the absolute temperature, $R_G$ the gas constant, and $Q$ the activation energy. We ignore for now additional dependencies of $B$ on grain size (for diffusion creep), chemical activity, etc. Then, (40) introduced in (2) with $\varepsilon$ as $\chi_0$ gives

$$\frac{1}{n_e} = \frac{\dot{\varepsilon}}{\sigma} \frac{\partial \sigma}{\partial \varepsilon} = \frac{1}{n} \left( 1 + \frac{\partial \theta}{\partial \varepsilon} \right). \tag{41}$$

\cite{74} The temperature is influenced by the strain rate through the heat produced by shearing. Perturbing the heat production by $dH$ changes the temperature by

$$dT = T dH, \tag{42}$$

where the coefficient $T$ is the heat retention factor that may depend on the size of the perturbation and the time $t$ since its onset, as well as on the thermal properties of the rock and the geometry of the perturbation. For instance, if heat retention is adiabatic, we have

$$T = t/\rho C, \tag{43}$$

with $\rho$ the density and $C$ the thermal capacity of the rock. If heat conduction is included, the temperature anomaly depends on distance from the perturbation \cite{CarlslawJaeger1959}. Only the temperature change at the location of the perturbation, $dT$, is important for shear-heating feedback. We define $dT$ as the average temperature anomaly within a thickness $d$, small but finite, of the perturbation origin, which is identified with the perturbation “size”. More specifically, we use the formula for a planar heat source continuously active from time 0 (onset of the perturbation) to $t$ \cite{CarlslawJaeger1959, equation 10.4.9} to compute

$$T = \frac{t}{\rho C} \left[ \text{erf} \left( \frac{d}{2\sqrt{\kappa t}} \right) + \frac{d}{2\sqrt{\kappa t}} \exp \left( -\frac{d^2}{4\kappa t} \right) - \frac{d^2}{2\kappa t} \text{erfc} \left( \frac{d}{2\sqrt{\kappa t}} \right) \right], \tag{44}$$

with $\kappa$ the thermal conductivity of the material.

\cite{75} The linear relation between the anomalies of heat production and temperature (42) is valid when $dH$ is small. It gives

$$\frac{\partial T}{\partial \varepsilon} = T \frac{\partial H}{\partial \varepsilon} = \beta T \sigma \left( 1 + \frac{1}{n} \right). \tag{45}$$

Then the effective stress exponent (41) becomes

$$\frac{1}{n_e} = \frac{1}{n} \left[ 1 - \left( 1 + \frac{1}{n} \right) \frac{\partial H}{\partial \varepsilon} \right]. \tag{46}$$

where $H$ is a nondimensional heat production defined as

$$H = THR_G/Q. \tag{47}$$

\cite{76} In the $\theta$–$H$ space, contours of effective stress exponent follow parabolas (Figure 10). High temperatures are stable because the low strength of the rock generates only a little heat; shear heating is not efficient enough to overcome the direct strengthening effect of increasing the strain rates. As $H$ depends on stress and thus on temperature by (40) $H$ and $\theta$ are not independently controlled.

7.1.2. Application to Earth materials. \cite{77} The importance of localization by shear heating in the lithosphere is evaluated using a set of experimentally derived rock rheologies. We use the flow laws of Gleason and Tullis \cite{GleasonTullis1986} for quartzite, Caristan \cite{Caristan1982} for diabase, Mackwell et al. \cite{Mackwell1998} for ultradry diabase, and Karato et al. \cite{Karato1986} for olivine in the dislocation creep regime. In addition, we use the results of Dimanov et al. \cite{Dimanov1998} for plagioclase deforming in the diffusion creep regime, assuming a fixed grain size of 0.1 mm.

\cite{78} The effective stress exponent for shear heating is shown in Figure 11 as a function of temperature, assuming a strain rate of 10^{-15} s^{-1} and $\tau \sim 10^8$ K W^{-1}. If the heat retention is total (adiabatic conditions), $\tau = t/\rho C \sim 10^8$ K W^{-1} corresponds roughly to a perturbation active during ~1000 years, in a rock of density $\rho \sim 3000$ kg m^{-3} and heat capacity $C \sim 1000$ J kg^{-1} K^{-1}. 

\[\text{Figure 10. Contours of inverse effective stress exponent } 1/n_e \text{ for shear heating with } n = 3, \text{ as a function of the generalized coordinates } \theta \text{ and } H \text{ (see section 7.1.1). Contours are plotted every 0.1 down to -2. Contours are dashed where } n_e > 0 \text{ (no localization). The thick line marks the transition to localization } (1/n_e = 0).\]
Although very strong localization is predicted (1/ne < -1), it requires temperatures below 300°C except for the example of diffusion creep. However, ductile flow is replaced by brittle failure or a high stress deformation mechanism such as Peierl’s creep at these temperatures. If T was higher, the maximum temperature for localization would increase, but heat losses by conduction could be important, as the implied time to hold the perturbation would be longer. Higher strain rates also help localization, but they do not occur spontaneously in the Earth’s lithosphere.

[79] A more interesting question is what heat retention factor τ is needed for localization, given a flow law, the deformation rate and the ambient temperature. Once τ is known, we invert (43) or (44) to obtain the minimum time (or, equivalently, the critical strain) during which the perturbation has to be active to allow localization (Figure 12). Because localization is favored at low temperature, we use the temperature at the brittle-ductile transition for a given flow law, calculated following Brace and Kohlstedt [1980] for a geotherm of 30 K km⁻¹ and assuming horizontal shortening. Therefore, the temperature increases with strain rate and varies with rock type.

[80] Under the adiabatic assumption, localization may occur after 3–6% strain, with little dependence on the strain rate (Figure 12). Heat loss can only increase the time during which a perturbation must be held before it triggers localization. To include heat losses in the analysis, we assume a planar perturbation (equation (44)), κ = 1 mm² s⁻¹, and follow one of two assumptions regarding the perturbation size d:

1. The perturbation width does not depend on strain rate: d = 1 km. The adiabatic assumption is valid for large strain rates (Figure 12), but for δ < 10⁻¹⁴ s⁻¹, the time for which the perturbation must be sustained before localization occurs is prohibitively long.
2. The perturbation width is inversely proportional to the strain rate so that the velocity across the perturbed region is a fraction f = 1% of a total velocity v = 30 mm yr⁻¹ typical of plate tectonics:

\[ d = fv/δ. \]  

The adiabatic limit occurs at small strain rate, where the perturbation width is large. The time for localization is always longer than 1 Myr.

[81] Localization is improbable at low strain rates. The adiabatic limit requires perturbation widths in excess of 100 km at 10⁻¹⁶ s⁻¹. Even if the plane approximation made to derive (44) remained valid for shear zones so wide, holding such a perturbation for at least 1 Myr, the minimum time required for localization, would probably indicate a preexisting material heterogeneity. Hence the localization should be regarded as inherited rather than dynamic. On the other hand, high strain rates permit localization of thin perturbations after a time short compared to tectonic timescales. A kilometer-thick shear zone can be considered adiabatic at strain rates in excess of ~10⁻¹² s⁻¹. Such strain rates occur in the continuation of brittle faults. Then localization requires perturbations held for only 1000 years, which is of the same order as the earthquake cycle.

[82] We conclude that shear heating does not arise spontaneously in the lithosphere but requires preexisting heterogeneities. In particular, it is possible beneath seismogenic faults. The relation of ductile shear zones and brittle faults has been pointed previously from field evidence [Ramsay, 1980; Sibson, 1980; Wintlinger et al., 1998]. Certainly, some additional feedback processes may increase the likelihood of localization. In particular, Regenauer-Lieb and Yuen [1998] included elasticity and showed that a thermal crack can develop after <1 Myr, fed by the release of stored elastic strain energy. However, other feedback processes that we also neglected here can stabilize localization.

7.2. Onset of Diffusion Creep

[83] Several deformation mechanisms can coexist in ductile rocks. For instance, dislocation creep and diffusion creep involve the motion of two different types of microscopic defects [Ranalli, 1995]. They probably add to one another, although one often dominates the total strain rate. As diffusion creep depends on grain size, and the grain size may evolve due to dislocation motions, it is possible to devise a feedback mechanism by which ductile rocks soften. But do they weaken as well? Is the effective stress exponent negative? In what follows, we first assume that the grain size adjusts instantaneously to its stress-dependent recrystallized equi-
librium value and find that dynamic localization is impossible. When the equilibrium assumption is relaxed, localization through grain-size feedback is allowed.

7.2.1. Grain size at equilibrium. [84] The total strain rate of a rock undergoing simultaneously dislocation and diffusion creep, \( \dot{\varepsilon} \), is the sum of a dislocation creep strain rate

\[
\dot{\varepsilon}_D = A_D \sigma^n,
\]

and a diffusion creep strain rate

\[
\dot{\varepsilon}_G = A_G d^{-m} \sigma^p,
\]

where \( A_D, A_G, n, m, \) and \( p \) are material parameters and \( d \) is the grain size of the material. Additional dependencies on temperature, chemical composition, water and oxygen fugacity, etc., are ignored.

[85] Stress-dependent dislocation motion induces recrystallization of grains that competes with kinetic grain growth, so that the grain size evolves toward an equilibrium value \( D \):

\[
D = D_0 \sigma^{-r}.
\]

If grain size evolution is rapid, \( d = D \). The effective stress exponent is then derived from (2) with \( \dot{\varepsilon} = \dot{\varepsilon}_D + \dot{\varepsilon}_G \) as the localizing variable:

\[
n_e = \sigma \left( \frac{\partial \dot{\varepsilon}_D}{\partial \sigma} + \frac{\partial \dot{\varepsilon}_G}{\partial \sigma} \right) \frac{n + (p + mr)R}{1 + R}.
\]

with \( R \) the ratio of strain rate accommodated by diffusion creep over the strain rate accommodated by dislocation creep:

\[
R = \frac{\dot{\varepsilon}_G}{\dot{\varepsilon}_D} = \frac{A_G d^{-m}}{A_D} \sigma^{p+m-r-n}.
\]

[86] The effective stress exponent is always positive, regardless of the stress or the temperature; the latter enters in the coefficients \( A_D \) and \( A_G \). In fact, \( n_e \) is bounded by \( n \) at low stress (dislocation-dominated limit) and \( p + mr \) at high stress (diffusion-dominated limit) (Figure 13). Dislocation creep dominates at low stress if \( p + mr > n \) because the large equilibrium grain size makes diffusion creep inefficient. The alternative assumption that the grain size is fixed at an initial value gives (52) as well, with \( r = 0 \). However, diffusion creep dominates at small strains if \( d \) is fixed [Karato et al., 1986; Handy, 1989; Jin et al., 1998]. Weakening \( (n_e < 0) \) never occurs because neither the dislocation creep law nor the diffusion creep law weaken individually.

[87] Although dynamic localization is not possible, the effective stress exponent is always more than 1 and increases with stress. Hence progressive localization is possible. We recall from section 2.2 that progressive localization requires that a perturbed material is softer than in its initial configuration. Hence a further overall increase of stress activates preferentially the perturbed location. However, the initial perturbation is not able to start a runaway process in that material. A localized shear zone may develop but only if externally applied stress increases. Indeed, Jin et al. [1998] documented that the stress in a natural shear zone was higher than in its surroundings; localization by grain-size feedback may be progressive. In that case, localized shear zones are little more than flow markers and do not influence significantly large-scale tectonics unless another localization process is simultaneously active.

7.2.2. Nonequilibrium grain size. [88] In section 7.2.1 the grain size was fixed at its equilibrium value \( D \) (equation (51)). However, this assumption may not be valid if the stress or the temperature vary rapidly due to tectonic motions [Kameyama et al., 1997; Braun et al., 1999] or if grain nucleation is important. We now suppose instead that the grain size of the material is initially at \( d_0 \neq D \) and evolves toward equilibrium. A perturbation in strain rate changes both the equilibrium grain size and the rate of grain-size evolution, so that a perturbed location may appear weaker after a given time.

[89] Following Kameyama et al. [1997] and Braun et al. [1999], we assume that the grain size evolves toward the equilibrium value \( D \) according to

\[
\frac{d}{dt} = -\frac{\dot{\varepsilon}_D}{\varepsilon_T} (d - D),
\]

where \( \varepsilon_T \) is a critical strain. Unlike previous authors, we include only the strain rate accommodated by dislocation creep in (54):

\[
d = d_0 - \int_0^t \frac{\dot{\varepsilon}_D}{\varepsilon_T} (d - D) dt \approx d_0 - \frac{\dot{\varepsilon}_{Dt}}{\varepsilon_T} (d_0 - D).
\]

The approximation is valid for \( A_D \sigma^n / \varepsilon_T \ll 1 \). By making this assumption we underestimate the time required to reach a certain \( d \) starting from \( d_0 \). If larger values of \( t \) are of interest, the system of equations must be solved numerically [Braun et al., 1999].

[90] Inserting (49), (50), and (55) into (2) the effective stress exponent becomes

\[
n_e = n + R_0 p - A_D \sigma^n m L t / \varepsilon_T,
\]

with \( R_0 \) the value or \( R \) for grain size \( d_0 \):

\[
R = A_G d_0^{-m} \sigma^p / A_D \sigma^n.
\]

and \( L \) a localization parameter given by

\[
L = \frac{D_0}{d_0} \sigma^{-r} \left( p + n - r - R_0 (n + R_0 p) \right) - p - n + \frac{R_0 (n + R_0 p)}{1 + R_0}.
\]
In deriving (56)–(58) we used several times the approximation that \( ADs \approx eT/C \). At the onset of the stress perturbation, the effective stress exponent is positive, as demanded by the dislocation and diffusion creep laws, but the evolution of grain size toward equilibrium, being enhanced in regions of high \( \varepsilon \), results in a weaker rock after a critical time \( t_C \) given by

\[
t_C = eT/n + Rp(1/ADs)^{n+1}mL
\]

Localization through grain-size feedback is possible only when \( L > 0 \). This excludes a region near the equilibrium grain size; not only the perturbation must be held for a time longer than \( t_C \), but the initial grain size must be sufficiently far from equilibrium. Immediately outside of the range of initial grain sizes for which localization does not occur, the approximation of small time made in (55) is not valid, which makes the time needed for localization even longer than shown. Consequently, the domain where localization is not possible is larger than for strictly \( L > 0 \). The small time approximation corresponds to \( L_m/n + Rp/C \), but if the initial grain size is larger than its equilibrium value, localization may be possible only if the initial grain size is at least 1 order of magnitude smaller than its equilibrium value.

Localization through grain-size feedback is possible only when \( L > 0 \). This excludes a region near the equilibrium grain size; not only the perturbation must be held for a time longer than \( t_C \), but the initial grain size must be sufficiently far from equilibrium. Immediately outside of the range of initial grain sizes for which localization does not occur, the approximation of small time made in (55) is not valid, which makes the time needed for localization even longer than shown. Hence the domain where localization is not possible is larger than for strictly \( L > 0 \). The small time approximation corresponds to \( L_m/n + Rp/C \), but if the initial grain size is larger than its equilibrium value, localization may be possible only if the initial grain size is at least 1 order of magnitude smaller than its equilibrium value.

Localization through grain-size feedback is possible only when \( L > 0 \). This excludes a region near the equilibrium grain size; not only the perturbation must be held for a time longer than \( t_C \), but the initial grain size must be sufficiently far from equilibrium. Immediately outside of the range of initial grain sizes for which localization does not occur, the approximation of small time made in (55) is not valid, which makes the time needed for localization even longer than shown. Hence the domain where localization is not possible is larger than for strictly \( L > 0 \). The small time approximation corresponds to \( L_m/n + Rp/C \), but if the initial grain size is larger than its equilibrium value, localization may be possible only if the initial grain size is at least 1 order of magnitude smaller than its equilibrium value.

Localization through grain-size feedback is possible only when \( L > 0 \). This excludes a region near the equilibrium grain size; not only the perturbation must be held for a time longer than \( t_C \), but the initial grain size must be sufficiently far from equilibrium. Immediately outside of the range of initial grain sizes for which localization does not occur, the approximation of small time made in (55) is not valid, which makes the time needed for localization even longer than shown. Hence the domain where localization is not possible is larger than for strictly \( L > 0 \). The small time approximation corresponds to \( L_m/n + Rp/C \), but if the initial grain size is larger than its equilibrium value, localization may be possible only if the initial grain size is at least 1 order of magnitude smaller than its equilibrium value.

In deriving (56)–(58) we used several times the approximation that \( ADs \approx eT/C \).

Figure 14. Contours of time \( t/eT \) (in seconds) needed to reach localization for the superimposed dislocation and diffusion creep at 800°C, in function of initial grain size and imposed stress while the grain size evolves toward the equilibrium value (similar set of rheologies as in Figure 13). Localization is impossible near the equilibrium grain size (darker shade). The lighter shade indicates the region where \( ADs > eT/C \), where our small time approximation looses its validity and the time for localization is even larger than shown.

8. Conclusions

[93] The importance of localization of deformation can be addressed within a unified framework with the help of the effective stress exponent, \( n_e \). That quantity measures how the apparent rheology of a system of internal variable, \{\chi_i\}, responds to a perturbation of one of these variables, \( \chi_0 \). To compute the effective stress exponent (equation (2)), the coupling between \( \chi_0 \) and the other internal variables must be assessed, as well as direct response of the strength of the material to a perturbation of \( \chi_0 \). In our examples of sections 4–7, \( \chi_0 \) has been the strain or the strain rate undergone by a rock.

[94] The sign of \( n_e \) is a criterion for localization: Localization requires \( n_e < 0 \), or dynamic weakening of the material. The value of \( n_e \) indicates the efficiency of localization: localization is more efficient for \( 1/n_e \) more negative. The transition from nonlocalizing behavior to localizing behavior occurs at \( 1/n_e = 0 \), called the perfectly plastic limit.

[95] Most instances of localized shear deformation in the Earth’s lithosphere rely on the feedback between the rock strength and an internal variable that is not always explicit when the rheology is described. It could be the elastic strain, combined with plastic deformation, the geometry of a potential discontinuity; a variable describing the state of fault gouge in frictional sliding regime; or temperature or grain size variations in ductile rocks. All these
processes give negative $n_a$, although in some cases (rate-and state-dependent friction laws, shear heating, and grain size feedback), it is necessary for the perturbation to be held for longer than a critical time for the material to be weaker than in the unperturbed state. In Table 2, we compile the expressions for the effective stress exponents derived in sections 4–7. The main aspects of localization by these mechanisms are summarized below.

1. Upon brittle failure, the strength of a rock decreases, either from loss of cohesion (section 4) or through the interaction between plastic and elastic strains (section 5) with $n_a \sim -50$ to $-2$.

2. For friction constitutive behavior, the rate-and state dependent friction laws gives $n_a \sim -250$ (section 6.1). As velocity localizes, both spatial and temporal localization are predicted. Perturbations must be maintained for more than a critical time $t_c$ (section 6.2) to overcome the stabilizing direct velocity effect.

3. For the ductile regime, localization arises through feedbacks between the flow laws and temperature (section 7.1) or grain size (section 7.2). However, localization through shear heating requires a preexisting heterogeneity, such as a shallower brittle fault. (section 7.2). However, localization through shear heating requires a preexisting heterogeneity, such as a shallower brittle fault. Localization by grain size feedback requires large deviations from the equilibrium recrystallized grain size. If it can be initiated, localization in the ductile regime is potentially strong.

**Appendix A: Elastoplastic Medium With Discontinuity**

**A1. Elastoplastic Increment Matrix**

[96] The elastic compliance matrix relating the strain and stress vectors defined in (17) is parameterized by the Lamé parameters $\lambda$ and $G$:

$$
D = G \begin{pmatrix}
2 + r & 0 & r \\
0 & 2 & 0 \\
r & 0 & 2 + r \\
0 & 0 & 2
\end{pmatrix},
$$

(A1a)

$$
D^{-1} = \frac{1}{G} \begin{pmatrix}
\frac{2 + r}{2 + r} & 0 & \frac{2 + r}{2 + r} \\
0 & \frac{1}{2} & 0 \\
\frac{2 + r}{2 + r} & 0 & \frac{2 + r}{2 + r} \\
0 & 0 & \frac{1}{2}
\end{pmatrix},
$$

(A1b)

where $r = \lambda/G$.

[97] The elastoplastic strain increment matrix, $M$, has two parts (equation (20)). The first is the elasticity matrix $D$. The second stems from the plastic strain increment. The direction of the stress increment due to plastic flow is

$$
D \cdot \nabla g = \begin{pmatrix}
(1 + r)\sin \psi \cos 20 \\
\sin 20 \\
(1 + r)\sin \psi \sin 20 \\
\sin 20
\end{pmatrix},
$$

(A2)

and its magnitude is $\nabla f^T D d\varepsilon/d$, so that the material stays at yield. Using

$$
D \cdot \nabla f = \begin{pmatrix}
(1 + r)\sin \phi \cos 20 \\
\sin 20 \\
(1 + r)\sin \phi \sin 20 \\
\sin 20
\end{pmatrix},
$$

(A3)

we obtain

$$
d = \nabla f^T \cdot D \cdot \nabla g = 1 + (1 + r) \sin \phi \sin \psi.
$$

(A4)

**A2. Effect of a Discontinuity**

[98] Let us consider an external stress increment as defined in (24) and a planar discontinuity with its normal oriented in the $x$ direction separating a region deforming elastically from a region deforming elastoplastically. The components of the stress increment $d\sigma_{xx}$ and $d\sigma_{yy}$ are transmitted across the discontinuity. Therefore, in the elastoplastic domain, $d\sigma_{xx}$ and $d\sigma_{yy}$ are given by (24) but $d\sigma_{xy}$ is unknown.

[99] In the elastoplastic domain the stress increment verifies $\sigma = M \cdot \varepsilon$, with $M$ defined in (20) and $d\varepsilon$ a strain increment yet be determined. If there is no slip, $d\varepsilon_{xy}$ is transmitted across the discontinuity:

$$
d\sigma_{xy} = \frac{1}{2G(1 + r)}(\rho - (1 + r) \cos 20) \cdot \sin (\psi - \cos 20) d\varepsilon_{xy}.
$$

(A5)

The two remaining components of the strain increment, $d\varepsilon_{xx}$ and $d\varepsilon_{yy}$, are determined from the equations for $d\sigma_{xx}$ and $d\sigma_{yy}$. We obtain a set of two equations with two unknowns:

$$
S_1 d\sigma_{\|} = AG d\varepsilon_{xx} + BG d\varepsilon_{xy},
$$

(A6)

$$
S_2 d\sigma_{\|} = CG d\varepsilon_{xx} + DG d\varepsilon_{xy},
$$

where

$$
S_1 = \rho + \cos 20 \frac{\rho - (1 + r) \cos 20}{2(1 + r)} \cdot \sin \phi \cos 20,
$$

(A7)

$$
S_2 = \sin 20 \cdot \frac{1}{2(1 + r)} \frac{1 + (1 + r) \sin \phi \sin 20}{1 + (1 + r) \sin \phi \sin 20},
$$

(A8a)

$$
A = 2 + r - \frac{(1 + r) \sin \phi \cos 20}{1 + (1 + r) \sin \phi \sin 20},
$$

(A9)

$$
B = -2 \sin 20 \frac{(1 + r) \sin \phi \sin 20}{1 + (1 + r) \sin \phi \sin 20},
$$

(A10)

$$
C = -\sin 20 \frac{(1 + r) \sin \phi \sin 20}{1 + (1 + r) \sin \phi \sin 20},
$$

(A11)

$$
D = 2 - \frac{\sin^2 20}{1 + (1 + r) \sin \phi \sin 20}.
$$

(A12)

The solutions of this system of equations are given by

$$
d\varepsilon_{xx} = \begin{bmatrix}
S_1 & B \\
S_2 & D
\end{bmatrix} d\sigma_{\|},
$$

(A7)

$$
d\varepsilon_{xy} = \begin{bmatrix}
A & S_1 \\
C & S_2
\end{bmatrix} d\sigma_{\|}.
$$

(A8a)

We compute

$$
\begin{bmatrix}
A & B \\
C & D
\end{bmatrix} = \frac{(1 + r)(\cos 20 \sin \phi)(\cos 20 \sin \psi)}{d},
$$

(A8a)
The elastoplastic strain increments are undetermined if $n_{xy} = 1$. This occurs when the discontinuity is oriented at $\theta_0 = \pi/4 - \phi/2$ and $\theta_0 = \pi/4 - \psi/2$. These conditions correspond to the bifurcation analysis of Rudnicki and Rice [1975].

The other determinant can also be zero, indicating drastic localization of the normal strain, $n_{xx} = 0$, but we could not derive a simple expression for the critical angle.

**Appendix B: Rate- and State-Dependent Friction**

[101] The RSDF laws include two rheological variables, $V$ and $\theta$ (equation (29)). Using $V$ as the localization parameter, the effective stress exponent is, in general,
\[
\frac{1}{n_e} = \frac{V}{\mu} \left( \frac{\partial q}{\partial t} + \frac{\partial q}{\partial \theta} \right) = \frac{1}{\mu} \left( a + b \frac{V}{dV} \right), \tag{B1}
\]

As \( \theta \) is an integral function of the velocity history (equations (30a) and (30b)), the effective stress exponent depends on time and on the specific \( \nu \)-\( V \) history. We consider two specific cases. In the first (section 6.1), \( \theta \) remains at equilibrium \( \theta_0 = D/V \), which corresponds to the limit \( t \to \infty \) or \( D \to 0 \). In that case,

\[
\frac{d\theta}{dV} = -\frac{V}{V},
\]

\[
1/n_e = (a - b)/\mu = c/\mu. \tag{B2}
\]

In the other case (section 6.2) the system is initially at equilibrium at velocity \( V \) and \( \theta = D/V \); but for \( t > 0 \), the sliding velocity is \( V + dV \). As the velocity does not depend on time for \( t > 0 \), the state evolution laws (equations (30a) and (30b)) can be integrated and give

Dieterich law \( \theta = \frac{D}{V} + \frac{x}{V} \frac{dD}{dV} \),

Ruina law \( \theta = \frac{D}{V} + \frac{dV}{V} \),

with \( x = \exp(-aV + bDV)/V \); for \( dV < V \) both evolution laws give

\[
\frac{d\theta}{dV} = -\frac{D}{\mu} = -1 + \exp\left(-\frac{V}{D}\right), \tag{B3}
\]

which, substituted in (B1), gives (31).

Acknowledgments. We thank Oded Aharonson, Mark Behn, Muriel Gerbault, Brad Hager, Greg Hirth, Chris Marone, Jim Rice, Jack Wisdom, and Wenlu Zhu for sharing their insight on the various physical processes included in this study and Paul Tackley and Alexei Poliakov for their reviews. We are particularly indebted to David Bercovici, who went well beyond the call of duty to help us clarify the paper. Supported by NASA grant NAG5-4555.

References


L. G. J. Montesi and M. T. Zuber, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. (montesi@mit.edu; zuber@mit.edu)