

Instability of Deformation

Harry W. Green, II

*Department of Earth Sciences and
Institute of Geophysics and Planetary Physics
University of California
Riverside, California 92521*

Chris Marone

*Department of Geosciences
Pennsylvania State University
University Park, Pennsylvania 16802*

INTRODUCTION

This chapter addresses the mechanisms by which instabilities can develop in rocks in the nominally ductile regime under conditions that the authors believe are relevant to Earth. We will not address brittle failure or frictional sliding processes in detail, but much of our discussion will be couched in terms of knowledge gained in studies of these processes. As a consequence, we will begin by discussing why brittle failure and frictional sliding cannot be the mechanism by which rocks become unstable at depth, despite the strong evidence that they are the underlying mechanism of earthquakes at shallow depths (e.g., Scholz 1990, 2002).

Brittle shear failure in the normal sense is fundamentally a tensile process (tensile microcracks must first nucleate and self-organize via interaction of the stress concentrations at their tips). As a consequence, brittle failure and frictional sliding are strongly inhibited by pressure because work must be done against the pressure to open the Mode I cracks. The inhibition is so strong that the stress required to create a fault or initiate sliding on an existing fault becomes greater than the room-temperature flow stress of many rocks at pressures equivalent to only a few tens of kilometers in Earth. Increasing temperature has little effect on brittle processes, but in contrast, the ductile flow stress of rocks falls exponentially with temperature. Because both pressure and temperature increase with depth in Earth, earthquakes by *unassisted* brittle fracture mechanisms or frictional sliding can only occur at depths less than ~30-50 km. Nevertheless, earthquakes occur to depths of almost 700 km in Earth and seismological evidence demonstrates unequivocally that they occur by displacement across a surface or narrow zone (i.e., they occur by faulting). This conundrum can be explained qualitatively by certain specific types of mineral reactions occurring under certain restricted conditions. The specific conditions required for faulting or sliding involve both the nature and kinetics of the mineral reactions and the nature of the way the rock that fails is loaded. That is, in order for a shearing instability to develop, certain rigorous conditions must be met involving the relative properties of the material that is faulting and its surroundings.

In this chapter, we will outline the requirements for a shearing instability to develop, followed by discussion of two different self-organizing processes associated with three different categories of mineral reactions that are currently known to lead to faulting at pressures where conventional brittle fracture is inhibited. In addition, we will discuss the popular theoretical concept of thermal runaway in shear zones as an earthquake mechanism in the context of the nature of loading in Earth and the lack of any obvious self-organizing aspect. We will show that the three known reaction-related, self-

organizing, faulting mechanisms have been characterized to various degrees and probably all have relevance to faulting in Earth. However, none of them has been sufficiently investigated to allow quantitative analysis of the criteria for instability in the same way that brittle deformation has been analyzed.

SHEARING INSTABILITY

As we have seen in the foregoing discussion, the great pressures at depth mean that earthquakes under such conditions cannot represent conventional frictional shear or fracture. Nevertheless, seismic studies show that deep earthquakes have double-couple focal mechanisms and moment tensors that are consistent with shear rupture along a narrow zone (Wiens et al. 1993, 1994). Thus, it is reasonable to pursue parallels between deep earthquake instability and shallow crustal earthquakes, which are comparatively well understood (e.g., Scholz 1990, 2002).

By definition, all earthquakes radiate energy at seismic frequencies, and thus rupture must propagate at velocities approaching the elastic wave speed of the surrounding rock. Dynamic rupture velocities and stress drops of earthquakes at depth are interpreted to be in the range of 2 to 4.5 km/s and 10-100 MPa, respectively (Fukao, and Kikuchi 1987, Wiens and McGuire 1995, McGuire et al. 1997). These observations provide two important constraints on the mechanism of intermediate and deep earthquakes. First, they show that elastic strains must exist in the hypocentral region at great depths. Because ongoing plastic deformation during subduction would relieve elastic strains generated at shallow depths, the elastic strains that drive intermediate and deep earthquakes must result from local, active deformation. A second point is that these earthquakes represent a mechanical shear instability with dominantly shear components of motion. Thus, a key question is that of how the earthquakes nucleate and how strain rates increase from background levels of 10^{-15} to 10^{-12} s^{-1} to the values of 10^{-1} s^{-1} or more that are appropriate for seismic slip rates of 1-10 m/s.

Intermediate and deep earthquakes represent mechanical shear instability in the sense that some form of sudden weakening causes strength to drop below the applied stress level, resulting in a local force imbalance and dynamic acceleration. Shear instability is the result of interaction between fault zone rheology and continuum coupling with the surroundings. There are two requirements for instability. First, the rheologic response must result in weakening; a perturbation in strain rate that results in strengthening will not grow to become dynamically unstable. Second, the rate of weakening must satisfy a stability criterion governed by the continuum interactions between the nucleation region and its surroundings. In the most general case of a gradual transition from quasi-static to dynamic slip, a complete model of shearing instability requires an understanding of both the initial weakening event and how its growth leads to degradation of the local elastic stiffness (e.g., Rice and Ruina 1983). Several mechanisms have been proposed for earthquakes at depth and the exact form of initial weakening is under investigation. Nevertheless, for our purposes we can discuss instability in a generic form and examine the two primary effects.

We may start by considering a nucleation zone embedded in a homogeneous medium (Fig. 1). In terms of the mineral reaction mechanisms discussed below, the nucleation zone may be thought of as a single fluid-containing shear crack or a region containing a critical spatial-density of fluid-filled Mode I cracks (or comparable anticrack structures produced by phase transformation). At this stage, we do not consider shear heating or thermal runaway. Although one could envision Figure 1a as a region in which shear heating had reached a critical threshold via localized slip, we focus attention here on the process(es) that can lead to such initial slip localization. We assume a region of

finite thickness W , and for simplicity, we take a circular region of radius r . We require only that the nucleation zone be mechanically weaker than the surroundings, so that its growth leads to weakening of the overall region, with stress concentration at the zone ends. The relationship between slip and weakening can be estimated using elastic dislocation theory:

$$\Delta\tau = C \frac{Gu}{r}, \quad (1)$$

where $\Delta\tau$ is the change in stress associated with slip u on a circular dislocation of radius r , G is the elastic shear modulus, and C is a constant. The effective elastic stiffness K of the region is then

$$K = C \frac{G}{r}. \quad (2)$$

Earthquake instability requires that fault zone strength fall faster than the unloading stiffness K of the surrounding region. Figure 1b is drawn in a way that emphasizes slip weakening, however equivalent plots could be constructed for strain-rate weakening, slip velocity weakening, or phase-transformation induced weakening. The energy available to drive shear rupture (including that to overcome shear resistance, create surface area, radiate seismic energy, and cause heating) is given by the shaded region in Figure 1b.

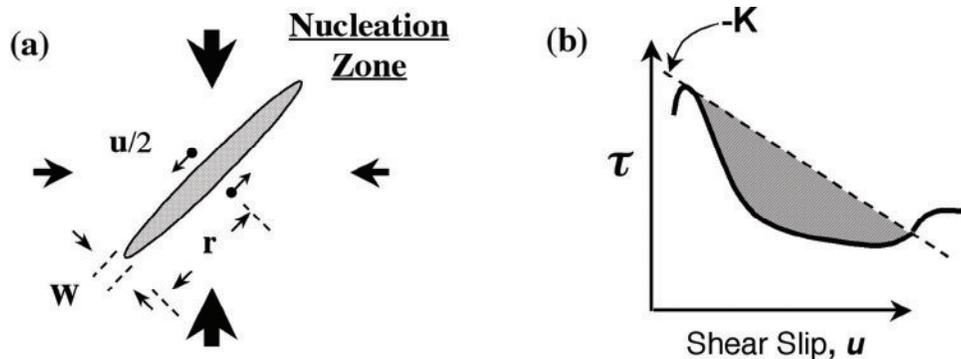


Figure 1. (a) Schematic view of the nucleation zone for a generic model of deep earthquake instability. Stippled area shows weakened region of radius r and width W . Internal shear strain is given by $g = u/W$, where u is slip measured from one side to the other. (b) Stress-displacement plot for the nucleation zone and surroundings. Dashed line of slope K shows the unloading stiffness (see Eqn. 2) for the region surrounding the nucleation zone. Heavy line shows hypothetical strength curve for the nucleation zone. Note that initial weakening causes strength to fall below the applied stress, leading to a dynamic force imbalance. Shaded region shows energy available to drive seismic rupture.

The instability condition is determined by the relative rates of weakening and elastic stress release. This can be quantified with reference to a generic slip-weakening model, in which we assume linear weakening over a distance D (Fig. 2). Taking τ^y as the yield strength and τ^f as the failure strength, we can write the instability condition as:

$$K < K_c, \quad (3)$$

where K_c is a critical stiffness given by:

$$K_c = \frac{(\tau^y - \tau^f)}{D}. \quad (4)$$

Because the elastic stiffness is always positive, we see that weakening is a necessary condition for instability. We may define the following cases.

$$\text{Stable:} \quad \tau^y < \tau^f, \quad K_c < 0 \quad (5)$$

$$\text{Conditionally Stable} \quad \tau^y > \tau^f, \quad K > K_c \quad (6)$$

$$\text{Unstable} \quad \tau^y > \tau^f, \quad K < K_c \quad (7)$$

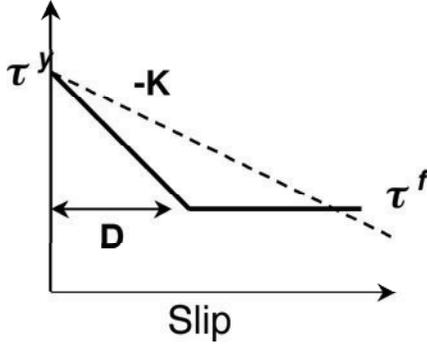


Figure 2. Generic slip weakening model for earthquake instability. Fault strength (solid line) weakens from τ^y to τ^f over slip distance D or shear strain $g_c = D/W$. Applied stress (dashed line) follows elastic unloading stiffness. Instability and seismic rupture are produced by the force imbalance between applied stress and fault strength.

In the first case, deformation is inherently stable because possible nucleation events, including slip or strain rate perturbations lead to strengthening. This is the standard expectation for steady-state plastic rheologies or viscous deformation in which strength increases with strain rate. The necessary condition for instability is met for the conditionally stable case, but the rate of weakening is insufficient to satisfy the instability criterion. In this case, the fault zone weakens but shear is stable and does not generate seismic radiation. In the final case (Eqn. 7), both the necessary and sufficient conditions are met and perturbations of any size could grow to become unstable. The dynamic force imbalance in this case would lead to rapid fault acceleration, shear heating and seismic radiation during slip.

In the context of a stability analysis for earthquake nucleation at high pressure, a key question is that of how rapidly a fault weakens during initial slip. For the analysis presented here we use only slip weakening, hence we are assuming that weakening is a constitutive property of a fault zone, independent of strain rate, slip or other variables. To the extent that this approximation holds, the critical weakening rate for instability can be quantified as a rheologic stiffness, as done in Equation (4).

By combining the relations for effective elastic stiffness (2) and the critical rheologic stiffness (4) we may write the stability criterion as:

$$C \frac{G}{r} < \frac{(\tau^y - \tau^f)}{D}. \quad (8)$$

This relation brings out an important point regarding conditionally stable shear (Fig. 3). In this case, the rate of weakening with slip is slower than the rate of stress reduction by unloading and thus shear is stable. However, growth of the nucleation region will reduce the local stiffness, which may eventually lead to instability. In the context of this approach, a critical nucleation patch size r_c can be defined:

$$r_c = C \frac{GD}{(\tau^y - \tau^f)}. \quad (9)$$

Figure 3 shows that as the nucleation dimension grows from r_1 to r_3 , the effective stiffness falls below the critical stiffness and the instability condition is satisfied.

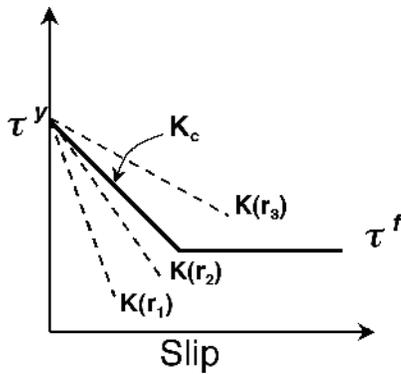


Figure 3. Slip weakening model of fault strength (solid line) together with the effective stiffness for three fault nucleation sizes ($r_1 < r_2 < r_3$). Stiffness decreases with the nucleation dimension. K_c denotes critical stiffness defined by fault strength and rheology. A critical nucleation size for instability r_c can be defined in terms of K_c .

In summary, earthquake instability requires that fault zone strength fall faster during weakening than the unloading stiffness of the surrounding region. In the context of a simplified, generic model of fault rheology, a critical stiffness for instability can be defined. A key question is that of how rapidly a fault weakens during initial slip. Clearly, much of the above analysis is highly oversimplified. More complex approaches would account for the effect of perturbation size on yield strength and failure conditions and incorporate specific mechanisms of plastic instability into relations for the critical rheologic stiffness. Also, the effect of latent heat of transformation, shear heating, and complex nucleation geometry would be included in the rheologic and continuum models. Quantitative treatment of these important effects is beyond the scope of this chapter and requires additional information about potential fault nucleation

EXPERIMENTAL HIGH-PRESSURE FAULTING MECHANISMS

Dehydration-induced embrittlement

Breakdown of hydrous minerals. Raleigh and Paterson (1965; see also Raleigh 1967) conducted deformation experiments on serpentinite under confining pressures of a few hundred MPa. They found that at elevated temperatures the rock was ductile. However, under conditions where the serpentine was breaking down to olivine + talc + H₂O, the rock failed catastrophically by faulting. This effect has been interpreted classically as a result of reduction of the effective normal stress on a fault plane or potential fault plane, resulting in the following relationship:

$$\tau = \tau_0 + \mu(\sigma_n - p_f), \quad (10)$$

where τ is the shear stress, τ_0 is a constant, μ is the coefficient of friction, σ_n is the normal stress on the fault, and p_f is the pressure in the pore fluid. Experimentation under conditions where the pore fluid pressure can be manipulated independently of the stress applied to the solid rock has verified this *effective stress* relationship for both brittle failure of intact rock and frictional sliding under conditions where $p_f < \sigma_3$. However, the qualitative description of the fluid as partially supporting the normal stress is unsatisfactory in a mechanistic sense because if $p_f < \sigma_3$, then the fluid can only exist in pores for which the strength of the solid matrix shelters the pores. If that is the case, how does the fluid partially reduce the normal stress on a fault? Worse, how does it partially reduce the normal stress on a fault that isn't there yet? Clearly, this relationship may provide a satisfactory continuum description of the phenomenon, but not a mechanistic explanation.

The answer comes from modern understanding that brittle shear failure is at its base

tensile failure. Experimentally, it has been known for 40 years that small shear cracks cannot be made to propagate in their own plane. Application of a stress field that would be expected to make such a crack propagate causes “wing” cracks to originate at sites of greatest tensile stress concentration at the tips of the crack and propagate along the trajectory of σ_1 , coming to rest when they are sufficiently far from the stress concentration that initiated them for the stress at the crack-tip to fall below the local tensile fracture strength. Thus, macroscopic shear failure initiates by a gradual process of local tensile failure at points of stress concentration (pre-existing cracks, grain boundaries, pores, grains of contrasting elastic properties, etc.), culminating in a critical population of tensile microcracks that then self-organize to initiate a shear fracture (a fault) that propagates by repetition of the “microcracks-first” process along the path of the growing fault surface.

Although perhaps counterintuitive and an unwelcome complication, this insight provides a built-in understanding of how a pressurized pore fluid facilitates faulting. It does so by enhancing the local tensile stress concentrations that lead to formation of the tiny tensile microcracks and by holding them open after they form (Fig. 4). Thus, the greater the pore pressure, the lower the applied stress necessary to bring local stress concentrations to the failure stress, and the lower the overall stress must climb to in order to generate proliferation of microcracks to the point of self-organization into an incipient fault zone that can then grow by the same pore-pressure-induced process. If the pore pressure should reach σ_3 , the expected failure mode would be runaway propagation of a tensile crack, leading to sample splitting in the laboratory or a fluid-filled vein (or melt-filled dike) in Earth. In the laboratory, this is easily demonstrated to be the case. However, in a natural situation with a slowly evolving pore pressure, the more usual case should be generation of shear failure before the pore pressure reaches σ_3 .

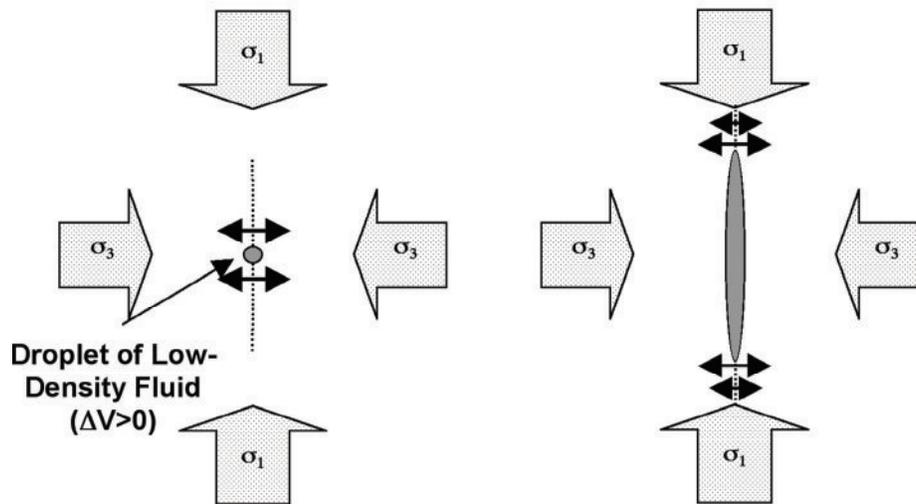


Figure 4. Fluid-assisted opening of “tensile microcracks.” Nucleation of a droplet of low-density fluid (net ΔV of reaction greater than zero) at a stress concentration in a nonhydrostatic stress field induces deviatoric tensile stresses around the bubble, with maximum tangent to the plane normal to σ_1 as shown on left. Such stresses will enhance nucleation of a tensile microcrack that will grow until the stress at the crack tip produced by the combination of the applied stress and the fluid pressure within the crack falls below the local tensile fracture stress.

There is abundant laboratory evidence at moderate pressures that this mechanism can induce brittle failure at greatly reduced stresses compared to those required in the absence of a pore fluid. Importantly, as was the case in the experiments of Raleigh and Paterson

(1965), development of a pore pressure can lead to brittle failure where otherwise (i.e., in the absence of a fluid) rock behavior will be ductile. As a consequence, dehydration embrittlement is an attractive hypothesis to explain earthquakes at pressures where unassisted brittle failure is completely inhibited.

There is a problem, however. Fluids are very much more compressible than solids, hence dehydration reactions, which all produce a positive ΔV of reaction at low pressures (because of the very low density of hydrous fluids), will see the magnitude of ΔV progressively reduced as pressure rises and ΔV will become negative at some point (because the net effect of the more dense solid product phases will come to dominate). For antigorite, the serpentine mineral stable at the highest temperatures and pressures, that point is at $\sim 750^\circ\text{C}$, 2.2 GPa (Ulmer and Trommsdorff 1995), a pressure representing a depth of only ~ 70 km in Earth (Fig. 5). For other hydrous minerals likely to be reasonably abundant in mantle lithologies, the conditions are comparable.

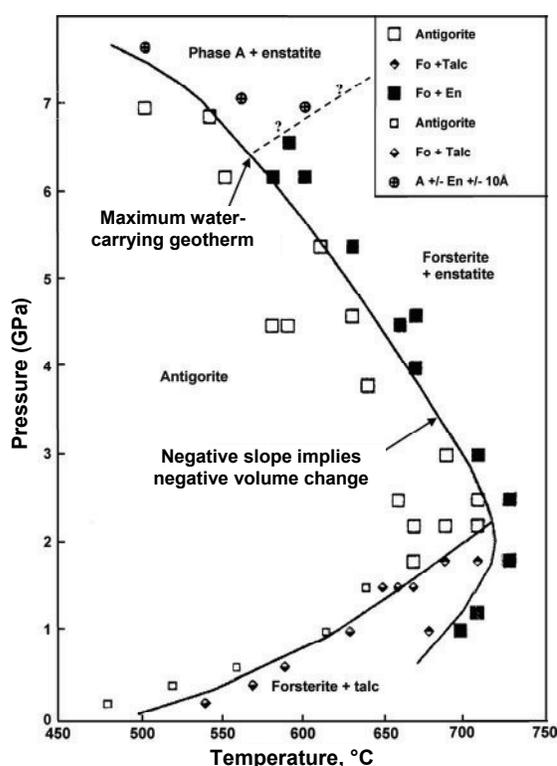


Figure 5. Antigorite phase diagram. Decomposition relations for a natural antigorite (modified after Ulmer and Trommsdorff 1995). Between ~ 2.2 and 6.4 GPa, dehydration yields anhydrous phases and the slope of the dehydration boundary is negative, implying $\Delta V < 0$. Above 6.4 GPa, Phase A, a dense hydrous magnesium silicate, forms rather than olivine and the slope of the dehydration curve becomes more negative.

However, if $\Delta V < 0$, the conventional rock mechanics view is that the pore-pressure effect will disappear (e.g., Wong et al. 1997). That is because when the net ΔV becomes negative, generation of a droplet of fluid results in a low-pressure pore surrounded by compressive hoop stresses, hence any stress concentrations that encourage formation of Mode I microcracks are pushed *away* from local failure rather than toward it. The greater the negative ΔV , the lower is the pore pressure produced and the greater the compressive “capsule” in which it is confined. Both effects push the material away from local tensile failure, hence this has been interpreted as inhibiting bulk shear failure. If the latter interpretation is correct, then dehydration embrittlement of serpentine and other common hydrous alteration minerals of peridotite (e.g., chlorite) will be limited to relatively shallow depths as a mechanism for triggering earthquakes.

This hypothesis has been tested in the laboratory (Reinen et al. 1998) and found that dehydration of antigorite under stress at confining pressures above 3 GPa still leads to

faulting, showing that bulk shear failure can still occur with $\Delta V < 0$. However, in keeping with the predictions of brittle rock mechanics, the microstructures accompanying faulting at those high pressures show no evidence of the Mode I microcracks diagnostic of brittle shear failure and fault gouges (Fig. 6) also do not consist of angular fragments with a fractal size distribution. These observations indicate that the mode of failure is not traditional frictional failure nor embrittlement of that type assisted by dehydration.

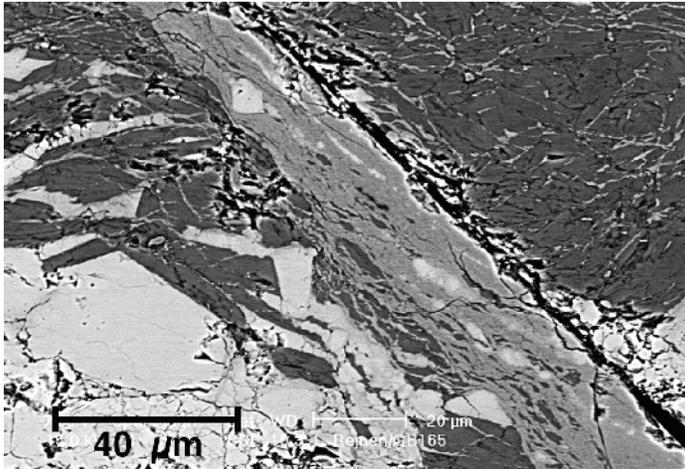


Figure 6. Faulting of antigorite when $\Delta V < 0$. Despite the fact that ΔV for the dehydration reaction becomes negative above 2.2 GPa, dehydration under stress leads to faulting. However, neither Mode I microcracks nor a fault gouge with angular fragments is observed.

Much remains to be learned about the fundamental physics of failure under these conditions. Nevertheless, at least we know that dehydration of hydrous minerals remains a viable trigger mechanism for earthquakes to depths of at least 250 km and probably much more, regardless of the sign of the ΔV of dehydration (so long as hydrous phases are present and are breaking down to form less-hydrous assemblages).

Exsolution of H₂O from nominally anhydrous minerals. A new variant of dehydration embrittlement recently has been discovered that provides additional insight into fluid-enhanced faulting (Green 2001). In a study of the rheology of eclogite at $P = 3$ GPa, experiments on a “wet” eclogite resulted in faulting when specimens were deformed at temperatures between the wet and dry solidi. The experimental specimens were fabricated from a natural eclogite with no hydrous phases, but with significant H₂O dissolved in both of the major, nominally anhydrous, phases (omphacitic pyroxene and garnet). Above the water-saturated solidus, H₂O exsolved from the silicates and triggered melting at grain boundaries. When this happened under stress, the specimen failed by faulting and was characterized by myriads of glass-filled Mode I cracks as well as a “normal” fault gouge except for the presence of small amounts of glass. In this case, ΔV was clearly greater than zero because the only phases being produced were fluid and melt, both of which are less dense than either omphacite or garnet; the microstructural characteristics of the specimens are as expected for $\Delta V > 0$. Thus, this failure was truly dehydration embrittlement in the original sense.

Change of micromechanism of dehydration-induced failure with change in sign of ΔV . The change of microstructures associated with dehydration-induced failure when the sign of net ΔV changes from positive to negative strongly suggests a change in failure mechanism. We defer discussion of this potential mechanism change until after discussion of transformation-induced faulting.

Transformation-induced faulting

The other experimentally established high-pressure faulting instability is the anti-crack mechanism known to operate under certain restrictive conditions during the olivine \rightarrow wadsleyite and olivine \rightarrow ringwoodite transformations (Green and Burnley 1989,

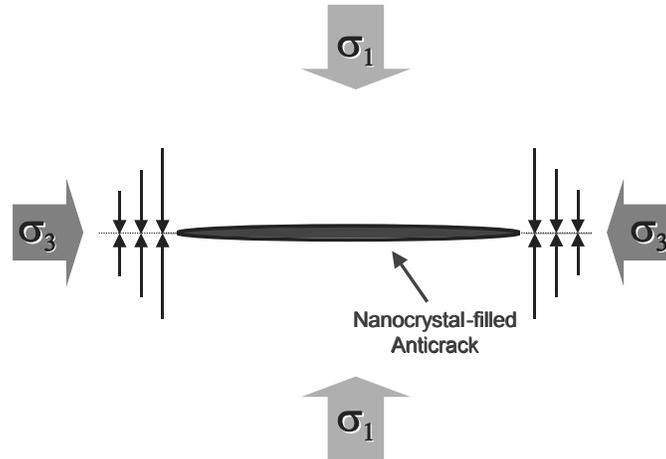


Figure 7. Nucleation and growth of anticracks is analogous to nucleation and growth of fluid-filled cracks during dehydration under stress (compare with Fig. 4). In this case, formation of a nucleus of the dense polymorph induces strong compressive stress concentrations and a small rise in T that, in turn, cause rapid-fire nucleation in the plane normal to s_1 . The result is microanticracks that have high compressive stresses at their tips and are filled with a nanocrystalline aggregate of the denser phase that has fluid-like properties even at seismic slip rates.

Green et al. 1990, Burnley et al. 1991, Green & Houston 1995, Kirby et al. 1996). This is also a self-organizing instability; it requires an exothermic polymorphic phase transformation that leads to production of Mode I microanticracks (Fig. 7) filled with an extremely fine-grained aggregate of the new phase that interact via the compressive stress concentrations at their tips in an analogous way to the interactions between tensile stress concentrations at the tips of open or fluid-filled Mode I microcracks (Green and Burnley 1989).

The reason an exothermic polymorphic transformation is required to support this instability is shown in Figure 8. Cooling the high-temperature phase below the phase boundary causes it to become metastable; as the difference in Gibbs Free Energy (ΔG) between the two phases increases, so does the nucleation rate. However, when undercooling is sufficiently great, the nucleation rate becomes more dependent on the kinetics of the transformation than on the driving force, hence the nucleation rate reaches a maximum and at lower temperatures it declines at an increasingly rapid rate; at temperatures lower than defined by the left-hand branch of the curve, the reaction rate is zero. Along the near-vertical left-hand branch of the nucleation-rate curve, formation of a nucleus of the stable phase will yield a small heat release that will result in a small increase in local temperature, leading to increase in the nucleation rate. At the same time, if ΔV of the reaction is negative, the nucleus immediately will be surrounded by compressive stress concentrations that will increase the driving force for nucleation. If the material is under stress, the maximum compressive stress concentration around the initial nucleus will be located adjacent to the nucleus on planes oriented normal to σ_1 , hence additional nuclei will preferentially form adjacent to the first nucleus and lie in the plane normal to σ_1 . This anisotropic contraction of the transformed region increases the stress concentrations lying in the plane normal to σ_1 . Therefore, if the thermal conductivity of the metastable host is low and the kinetics of plastic relaxation of the stress concentrations are sufficiently slow, the combination of these two effects can lead to runaway nucleation in the plane normal to σ_1 – yielding a Mode I anticrack filled with a nanocrystalline solid with properties approximating that of a low-

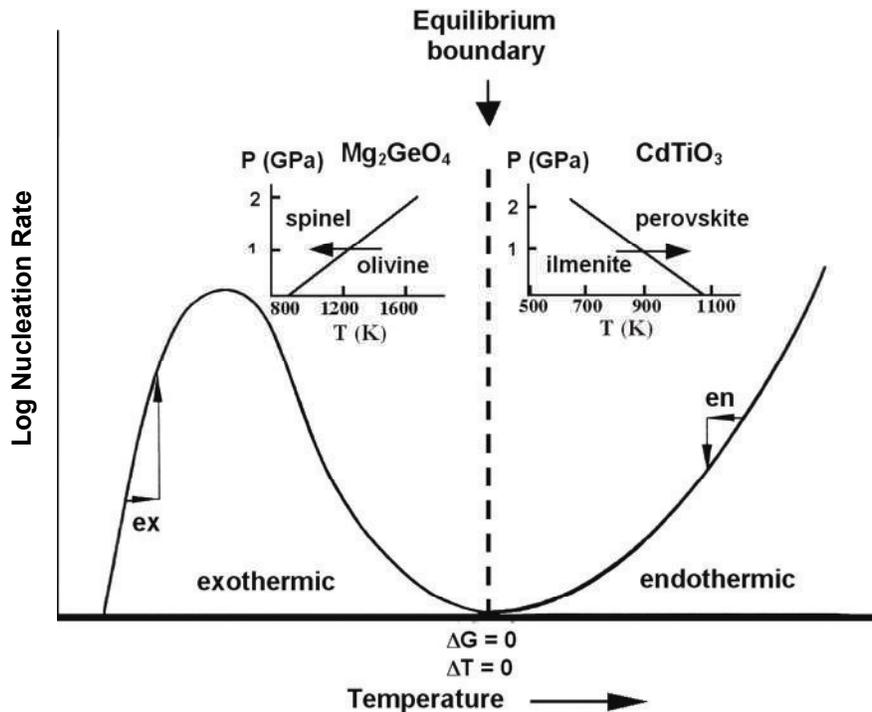


Figure 8. Thermodynamics of nucleation during polymorphic phase transformations. The topology of nucleation kinetics for transformation at constant pressure is described for exothermic reactions to the left of the vertical dashed line and for endothermic reactions to the right of that line. Examples are shown for cases where the high-pressure phase is the low-entropy phase (left) and the high-entropy phase (right). See text for discussion (modified after Green and Zhou 1996).

viscosity fluid.

As in the case with Mode I cracks, the greater the aspect ratio of the anticrack, the greater the stress concentration at the tip. Therefore, microanticracks interact with each other in a stressed material in an analogous way to microcracks, culminating in initiation of a fault and catastrophic shear failure.

In contrast, endothermic transformations have entirely different characteristics (right side of Fig. 8). As the low-temperature polymorph is progressively heated above the transformation temperature, both the driving force and the kinetics for the transformation increase monotonically. As a consequence, there is no kinetically controlled branch of the nucleation-rate curve and any thermal effects of nucleation would actually produce a small reduction of local temperature. Thus, the characteristics that can lead to positive feedback are absent; when the temperature is sufficiently high for the reaction to run, simple nucleation and growth of the high-temperature phase occurs, leading to growth of normal, blocky, crystals; no anticracks form and no instability is generated.

This argument strongly suggests that the sign of ΔS ($\geq \Delta H/T$) is critical to this instability; if the reaction is not exothermic, there will be no instability. However, nothing in the logic presented here requires that the algebraic sign of ΔV should matter. The magnitude of ΔV controls the magnitude of the stress concentration at the tips of microlenses, but there is no reason that the sign of ΔV should matter. This implication was tested in the CdTiO_3 system (Green and Zhou 1996); in the up-temperature (endothermic) direction, no instability developed. Under all conditions, transformation from the phase with ilmenite structure to that with perovskite structure ($\Delta V < 0$; $\Delta H > 0$) yielded blocky crystals of the daughter (high-temperature) phase growing on grain

boundaries of the parent phase. In contrast, as described in the previous paragraph, when tested in the reverse direction ($\Delta V > 0$; $\Delta H < 0$), the reaction produced fine-grained lenses (“microcracks”) of the ilmenite phase parallel to σ_1 and shear fracture followed. The logic for runaway growth of nanocrystalline lenses when $\Delta V > 0$ is analogous to that given above for the case $\Delta V < 0$; in this case the first nucleus is surrounded by extensile hoop stresses that prejudice the location of succeeding nuclei and the small amount of heat released leads to runaway nucleation in a plane parallel to σ_1 .

Clearly, then, a critical requirement for failure is an exothermic reaction. But is that a sufficient cause? The reaction ringwoodite \rightarrow perovskite + magnesiowüstite that defines the base of the upper mantle is endothermic but if olivine were to be carried metastably completely through the transition zone and then decompose into the lower-mantle assemblage, the reaction would be exothermic. It is necessary, therefore, to know whether such a disproportionation reaction could support the instability if the reaction is exothermic and ΔV is large. The reason to question this possibility is that a diffusive step is required for phase separation during this type of reaction. Is the time necessary to separate the chemical components into two phases sufficiently long that it abrogates the localized runaway nucleation of the daughter assemblage to form anticracks and lead to instability? This hypothesis was tested using the reaction



because it is both more exothermic and has a larger volume change than olivine \rightarrow spinel in Mg_2GeO_4 , the system in which we discovered transformation-induced faulting (Green and Burnley 1989), or in the olivine \rightarrow wadsleyite transformation in olivine of mantle composition (Green et al. 1990). The hypothesis was confirmed; decomposition of albite under stress yielded symplectites consisting of blocky crystals of jadeite filled with wormy intergrowths of coesite; no anticracks were observed and faulting did not occur (Gleason & Green 1996).

Table 1 summarizes the systems in which the anticrack mechanism has been investigated, verifying that it can operate during a variety of exothermic polymorphic transformations but not during endothermic transformations nor during disproportionation reactions, even when ΔV is large and the heat evolved is large and positive.

Table 1. Systems tested for transformation-induced faulting.

<u>Reaction</u>	<u>System</u>	<u>Faulting?</u>	<u>Pressure (Gpa)</u>	<u>Microstructure</u>	<u>Reaction type</u>
$\alpha \rightarrow \gamma$	Mg_2GeO_4	Yes	1-2	anticracks	Exo. poly. trans
$\alpha \rightarrow \beta$	$(\text{Mg,Fe})_2\text{SiO}_4$	Yes	14-15	“	“ “ “
$\alpha \rightarrow \beta$	Mn_2GeO_4	Yes	4-4.5	“	“ “ “
$\text{pv} \rightarrow \text{il}$	CdTiO_3	Yes	0.2	“cracks”	“ “ “
$\text{lce1h} \rightarrow \text{ll}^*\text{H}_2\text{O}$		Yes	0.2-0.5	“anticracks”	“ “ “
$\text{il} \rightarrow \text{pv}$	CdTiO_3	No	0.2	Blocky xtls	Endo. “ “
$\text{ab} \rightarrow \text{jd} + \text{coes}$		No	3-3.5	Symplectite	Exo. disproportion.

~~~~~  
*Predictions for the top of the lower mantle:*

|                                            |    |    |             |                      |
|--------------------------------------------|----|----|-------------|----------------------|
| $\gamma \rightarrow \text{pv} + \text{mw}$ | No | 25 | Symplectite | Endo. disproportion. |
| $\alpha \rightarrow \text{pv} + \text{mw}$ | No | 25 | Symplectite | Exo. disproportion.  |

~~~~~  
 *Kirby et al. (1991); compression features similar to anticracks were observed.

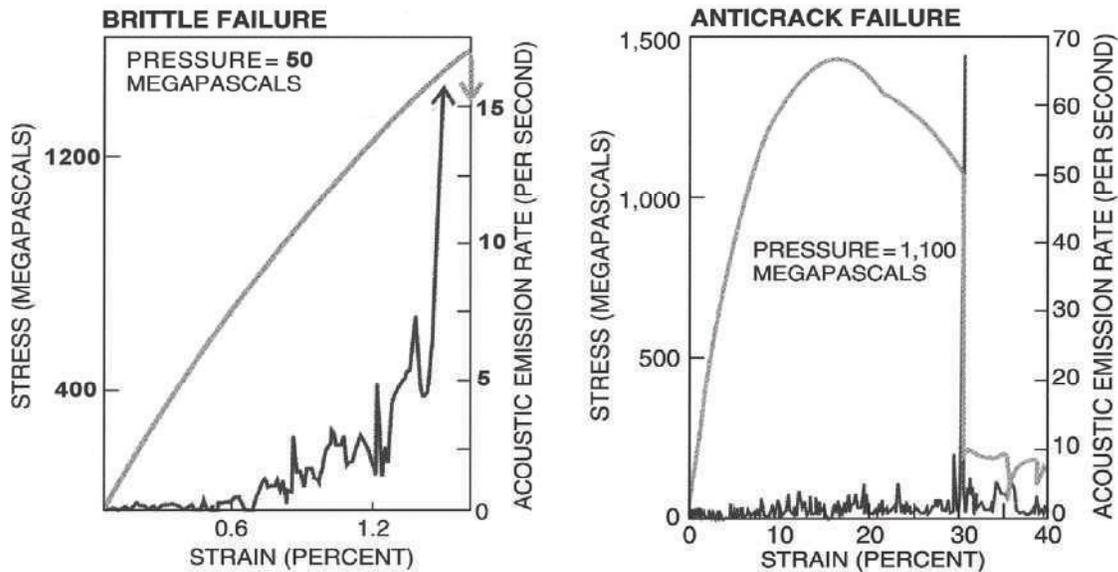


Figure 9. Comparison between low-pressure brittle failure of Westerly Granite (left) and high-pressure anticrack failure of Mg_2GeO_4 undergoing the olivine \rightarrow spinel transformation. Note that brittle failure produces an exponentially increasing number of acoustic emissions that rises very high before failure, whereas anticrack failure produces significant acoustic emissions only during macroscopic failure (modified after Green, 1994).

Like normal brittle failure, anticrack failure in the laboratory generates acoustic emissions (Green et al. 1992) and thus is potentially an earthquake mechanism (Fig. 9).

“Brittle” versus “plastic” shear failure

Transformation-induced failure clearly operates by a different physical process than conventional brittle failure, yet it mimics many aspects. In particular, it achieves self-organization via Mode I nanocrystalline lenses (either parallel or perpendicular to σ_1 , depending on the sign of ΔV). Nevertheless, it produces acoustic emissions only during shear failure; the anticrack growth phase before bulk failure is silent (Fig. 9). This observation plus the current interpretation of the underlying physics of anticrack growth as a runaway nucleation phenomenon indicates that this failure mechanism is perhaps better classified as a plastic instability rather than brittle failure. For dehydration-induced faulting, the change in microstructures associated with change in sign of ΔV also suggests a change in failure mechanism. In particular, change from Mode I microcracks and angular, fractal, fault gouge to microstructures indicative of mineral reactions suggests that dehydration with negative ΔV also may trigger a plastic instability. The details of this proposed switch of mechanisms remains unclear.

Thermal runaway due to shear heating

In addition to dehydration-induced faulting and transformation-induced faulting, shear failure mechanisms that have been established in the laboratory, it has been proposed a number of times on theoretical grounds that a shearing instability can develop based on thermal feedback between shear heating and thermally-induced viscosity reduction under conditions approximating constant stress (e.g., Grunfest 1963, Griggs and Baker 1969, Hobbs and Ord 1988, Kanamori et al. 1998, Karato et al. 2001). These treatments leave little doubt that such a runaway could lead to failure. What is not clear is whether boundary conditions in Earth are such that the runaway can be supported. For

example, under constant strain-rate or constant displacement conditions, the thermally-induced viscosity resulting from shear heating will lead to stress relaxation rather than runaway heating, and any thermal anomaly will dissipate rather than leading to failure.

As described above, the anticrack mechanism also appears to be a thermal runaway phenomenon. Why is this mechanism self-organizing and shear heating not? Like in brittle failure, the anticrack instability is in the development of the primary (Mode I) failure mechanism. The thermal runaway in anticrack development works off of latent heat release and therefore is proportional to reaction progress and depends on the relative rates of thermal conductivity and reaction. No energy need be supplied externally during the heating process and the volume of material transformed before failure can be extremely small. In contrast, shear heating is proportional to the work done during straining and therefore is a function of stress, strain, viscosity, thermal conductivity, and geometrical boundary conditions (the system must be pre-organized into an appropriately-oriented shear zone). Instability can be obtained only if the boundary conditions ensure that the viscosity reduction induced by increasing temperature results in increase of strain rate rather than relaxation of stress.

APPLICATION TO EARTHQUAKE MECHANISMS

Earthquake distribution with depth

Figure 10 shows the distribution of earthquakes in Earth. Virtually all earthquakes deeper than 20-30 km are generated at the interface between subducting lithosphere and the over-riding plate, or within the subducting lithosphere itself, hence this distribution shows that the generation of earthquakes with depth is generally spread out over the entire depth of subduction to the base of the mantle transition zone, where they stop abruptly. Thus, it is to be expected that a successful candidate for an earthquake trigger mechanism will naturally produce earthquakes continuously with depth, rather than bursts at particular set(s) of conditions.

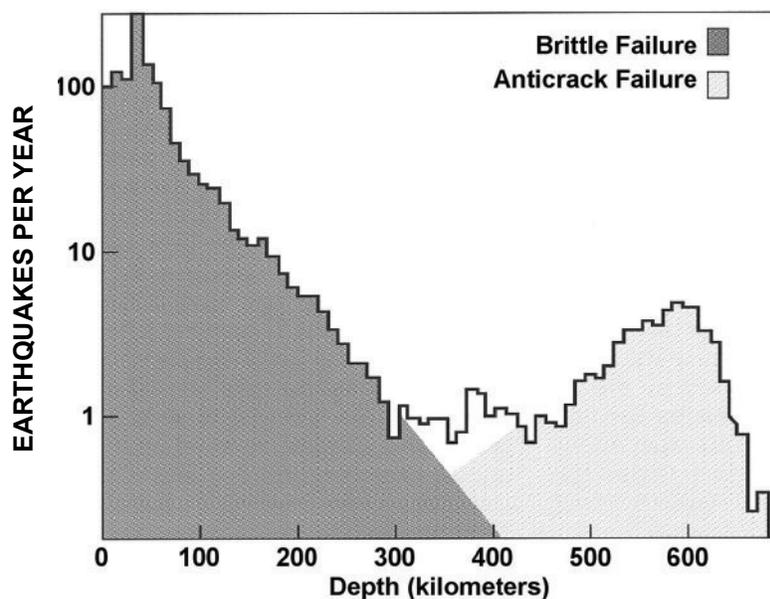


Figure 10. Semi-log plot of earthquake frequency with depth. Note exponential decline from 100 to 300 km and secondary population in mantle transition zone reaching a maximum at ~600 km before total cessation at 680 km (modified after Frohlich 1989).

Mineral reactions available to trigger earthquakes

We know from dredge hauls and heat flow that the upper portion of the oceanic lithosphere undergoes hydrothermal alteration shortly after its formation at oceanic ridges. Seismic velocities suggest that such alteration extends into the mantle beneath the crust and therefore that serpentine partially replaces olivine (and pyroxene), perhaps up to a depth of 10-12 km (Meade and Jeanloz 1991). The rare evidence for hydrous alteration in shallow mantle xenoliths from oceanic islands strongly supports this view. Thus, it has been generally assumed (e.g., Green and Houston 1995) that dehydration-induced faulting is an excellent candidate for triggering these earthquakes, but that this mechanism is probably not a candidate to explain the common occurrence of a second seismic zone deep within the lithosphere at intermediate depths (e.g., 40-200 km below Japan, Igarashi et al. 2001). However, Silver et al. (1995; see also Jiao et al. 2000) suggested that deeper hydration of the oceanic lithosphere might occur along fault zones created by great earthquakes outboard of trenches and others have pointed out the similar pattern of maximum antigorite stability in subduction zones (e.g., Ulmer and Trommsdorff 1995) and the distribution of earthquakes in intermediate-depth double seismic zones (e.g., Peacock 2001). More recently, a somewhat different mechanism has been proposed that suggests extensive serpentinization may occur at trenches (Phipps-Morgan 2001, Phipps-Morgan et al. 2002).

If it is true that the subducting lithosphere is hydrated at least sporadically to a depth of 40 km, dehydration-induced failure is a potential candidate for explanation of all earthquakes at depth. Here we will briefly examine this possibility in three depth intervals: 0-200 km, 200-300 km, 300-700 km.

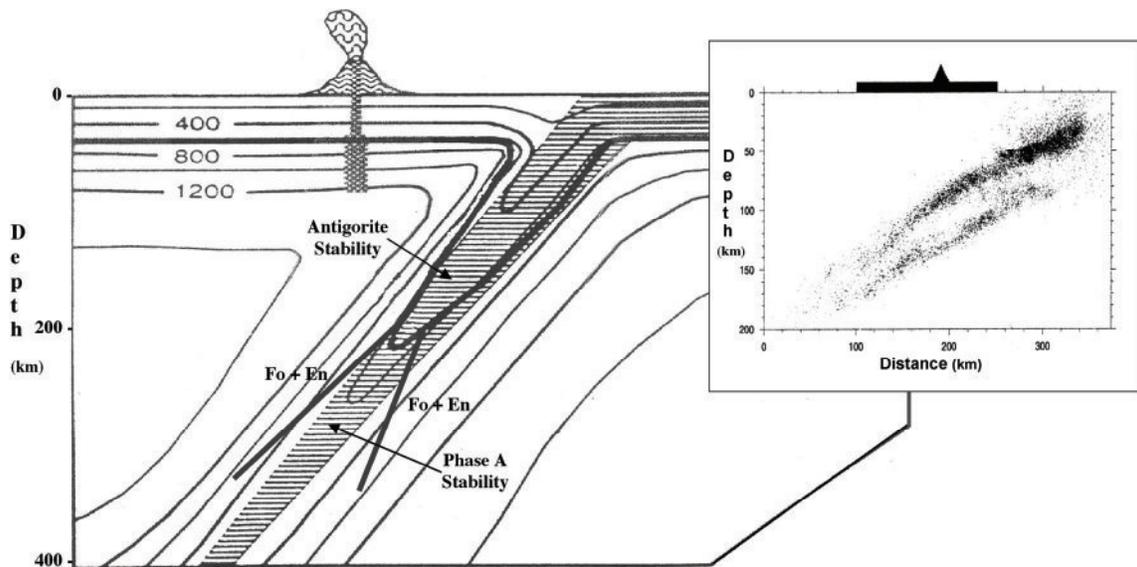


Figure 11. Stability of antigorite mapped onto a thermal model of a subduction zone. The particular model is such that antigorite stability reaches that of Phase A (~200 km depth). Hydrous phases existing in the cold core of the slab could access the antigorite→Phase A + en reaction and pass H₂O into the deeper slab. Otherwise, all H₂O will be released and could flux arc melting (heavy arrow) or trigger earthquakes, but could not be carried to great depths in significant quantities. Note similarity between limit of antigorite stability (heavy line) and earthquake distribution beneath northern Japan (inset). Hatched region represents cold lithosphere. (Modified after Ulmer and Trommsdorff 1995 and Igarashi et al. 2001).

(i) **0 - 200 km.** Depending on the thermal profile of a subduction zone, antigorite can remain stable in the cold interior of the descending slab up to a maximum depth of approximately 200 km. As shown in Figure 11, the shape of the antigorite phase diagram in the thermal model chosen is remarkably similar to that of the earthquake distribution under northern Japan. As a consequence, dehydration of antigorite could explain the entire distribution of earthquakes in the deeper zone, including its approach to and merging with the shallower zone. In addition, antigorite, in combination with a myriad of other hydrous phases potentially present in the hydrated oceanic crust and shallow mantle, can also explain the shallow zone. Of course, the specific depth to which antigorite could carry water and release it to trigger earthquakes is a function of the temperature distribution in the slab. As was shown in Figure 5, the temperature of maximum stability of antigorite falls rapidly at pressures greater than ~ 7 GPa, hence for water to pass to greater depths, it must be passed on to hydrous phases stable to greater depths. In particular, if the minimum temperature in the slab reaches ~ 550 - 600°C before a depth of ~ 200 km is reached, antigorite will completely dehydrate to olivine + enstatite and the slab will be completely dehydrated except for a few rare hydrous phases. However, Figure 5 also shows that if the slab is colder, antigorite will break down into Phase A + H_2O and the door will be open to carry water into the deep interior via a succession of dense hydrous magnesium silicate phases (e.g., Angel et al. 2001). Indeed, this “eye of the needle” has profound importance for mantle dynamics and therefore for past, present, and future evolution of Earth.

(ii) **200-300 km.** Deeper than 200 km, a problem arises for earthquake generation by dehydration. Examination of the data of Angel and Frost (2001) shows that any reasonable particle trajectory within a cold slab (a slab for which the minimum temperature passes below $\sim 600^\circ\text{C}$, 7 GPa) crosses all of the phase boundaries available in directions for which water is *conserved*. That is, each hydrous phase contains less water than the more dense hydrous phase(s) that replace it. Thus, no free fluid is generated by the reactions and dehydration would appear to be unavailable for triggering earthquakes. There are, however, a few hydrous phases that would be expected to be present in minor quantities in hydrous subducting lithosphere that could potentially extend dehydration embrittlement to depths somewhat greater than 200 km. For example, titanian clinohumite is stable to depths exceeding 300 km (Ulmer and Trommsdorff 2001) and could perhaps explain the low frequency of earthquakes at that depth.

(iii) **300-700 km.** The previous paragraph suggests that dehydration cannot serve as a trigger for earthquakes at depths greater than 300-400 km. In contrast, restriction of the anticrack instability to exothermic polymorphic transformations predicts that earthquakes can only be triggered by this mechanism at depths where the transformations of olivine to wadsleyite and ringwoodite occur (i.e., at depths of 300-700 km in subducting slabs). Therefore, this mechanism could explain the rise in frequency of earthquakes at the top of the transition zone and their cessation at its base (Green and Houston 1995). However, application of this mechanism to Earth requires that the central portion of descending slabs is sufficiently cold that olivine does not transform to its denser polymorphs until a critical set of conditions is reached where the kinetics of the reactions are rapid enough to run. It also may seem that the numerous special circumstances necessary for anticrack failure would make the probability small that it would happen routinely in Earth. However, in subduction zones, a given volume of material follows a particle path that is primarily up-pressure, but nevertheless is also continually encountering conditions of higher temperature as heat flows slowly into the slab from the surrounding mantle. As a consequence, if conditions are sufficiently cold in the central parts of a subducting slab for reaction to be inhibited at the top of the transition zone, a wedge of metastable olivine is created, all volumes of which must eventually encounter the conditions critical for

transformation-induced failure unless they pass into the lower mantle and transform there passively to perovskite + magnesiowüstite. Unfortunately, at the present time, neither the exact temperature distribution in slabs nor the exact conditions for metastable olivine preservation are known.

Seismic evidence that such metastable olivine is present in subducting slabs traversing the transition zone has been equivocal (cf. Koper et al. 1998) and some calculations based upon experimental studies of the kinetics of the transformations and thermal models of subduction zones have suggested that metastable olivine is likely only in the coldest slabs (Mosenfelder et al. 2001). Recently, however, evidence has been collected showing slow seismic velocities in a subhorizontal, slab-like, earthquake-generating anomaly beneath Fiji that is most easily explained as a remnant slab of metastable olivine floating in the transition zone (Chen and Brudzinski 2001, Green 2001). If this interpretation is valid, the conflict between observation and prediction may indicate that subduction zones are colder than suggested by current thermal models.

However, it is premature to reach such a conclusion. In addition to the uncertainties in slab temperature distributions and olivine-spinel reaction kinetics, there are specific characteristics of certain earthquakes that are difficult to explain by either of these mechanisms. Some very deep earthquakes, most notably those in the South American slab, are interpreted from their seismic characteristics to have involved seismic slip zones of diameter several 10s of km. For example, both the 1994 Tonga and Bolivia deep earthquakes have been estimated to involve slip on regions with diameters of ~60 km. Therefore, it is difficult to explain the full slip in these earthquakes by any mechanism that is inherently restricted to cold temperatures (i.e., both transformation-induced faulting and dehydration-induced faulting). Therefore, even if slab temperatures are colder than presently envisioned, it remains questionable as to whether there could be sufficiently large volumes of metastable olivine or hydrated mantle available to support such slip. It is important, therefore, to consider possible alternative or cooperative mechanisms. It is this concern that has led to resurgence of thermal runaway models for the deepest earthquakes (e.g., Karato et al. 2001).

CONCLUSIONS AND SPECULATIONS

The discussion presented here shows that dehydration-induced faulting and transformation-induced faulting are complementary; the former appears to be restricted to depths above ~350 km and the latter to depths below ~350 km. Moreover, the waning ability of dehydration to trigger earthquakes below 200 km conveniently falls in the region of lowest earthquake production and the inability of transformation-induced faulting to trigger faulting in endothermic or disproportionation reactions offers a natural explanation for cessation of earthquakes at the base of the transition zone. It is thus possible that earthquake initiation at depth in Earth can be completely explained by a combination of these two mechanisms.

Nevertheless, it is premature to reach this conclusion for three principal reasons:

- (i) It is not clear that double seismic zones at intermediate depths can be assigned to dehydration-induced faulting;
- (ii) very large, very deep earthquakes seem to be inconsistent with expectations about the size and geometry of a metastable wedge of olivine;
- (iii) combination of experimentally-measured kinetics of olivine breakdown reactions and thermal models suggest that subduction zones in general are too warm for metastable olivine to persist to significant depths other than in

Tonga, although if Chen and Brudzinski (2001) have truly observed metastable olivine in an older slab floating beneath Fiji, either the kinetics of transformation, the extrapolation of those kinetics to mantle conditions, or slab models may need revision.

Additional work on this problem needed at this time includes

- (i) laboratory experiments to determine whether the dense hydrous magnesium silicates are capable of generation of faulting via transient release of H₂O during reaction from one to another;
- (ii) quantitative determination of the conditions of failure of olivine of mantle composition as a function of pressure and development of accurate methods to extrapolate those conditions to natural time scales;
- (iii) investigation of mechanisms of hydration of oceanic lithosphere at subduction zones;
- (iv) a thorough analysis of the errors potentially involved in experimental measurement of the kinetics of olivine reaction at high pressures and temperatures, their extrapolation to natural conditions, and the assumptions involved in thermal modeling;
- (v) evaluation of alternative mechanisms in terms of their ability to function as trigger mechanisms for earthquakes under natural boundary conditions.

REFERENCES

- Angel RJ, Frost DJ, Ross NL et al. (2001) Stabilities and equations of state of dense hydrous magnesium silicates *Phys Earth Planet Inter* 127:181-196
- Antolik M, Dreger D, Romanowicz B (1996) Finite fault source study of the great 1994 deep Bolivia earthquake. *Geophys Res Lett* 23:1589-1592
- Burnley PC, Green HW II, Prior D (1991) Faulting associated with the olivine to spinel transformation in Mg₂GeO₄ and its implications for deep-focus earthquakes. *J Geophys Res* 96:425-443
- Chen WP, Brudzinski, M.R (2001) Evidence for large-scale, imbricate remnant of subducted lithosphere. *Science* 292:2475-2479
- Frohlich C (1989) The nature of deep-focus earthquakes. *Ann. Rev Earth Planet. Science* 17:227-254
- Fukao Y, Kikuchi M (1987) Source retrieval for mantle earthquakes by iterative deconvolution of long-period P-waves. *Tectonophysics* 144:249-269
- Gleason G, Green HW II (1996) Effect of differential stress on the albite to jadeite + coesite transition at confining pressures of >3 GPa. *EOS, Trans Am Geophys Union* 77:F662
- Green HW II (1994) Solving the paradox of deep earthquakes. *Sci Am* 271:64-71
- Green HW II (2001) Deep Tonga backarc: Graveyard for buoyant slabs bearing metastable olivine? *Science* 292:2445-2446.
- Green HW II (2001) Physical mechanisms for earthquakes at intermediate depths. *EOS Trans Am Geophys Union* 82, Fall Mtg Suppl, Abstr S42D-03
- Green HW II, Burnley PC (1989) A new self-organizing mechanism for deep-focus earthquakes. *Nature* 341:733-737
- Green HW II, Houston H (1995) The mechanics of deep earthquakes. *Ann. Rev Earth and Planet. Sci.* 23:169-213
- Green HW II, Scholz CH, Tingle TN, Young TE, Koczyński T (1992) Acoustic emissions produced by anticrack faulting during the olivine → spinel transformation. *Geophys Res Lett* 19:789-792
- Green HW II, Young TE, Walker D, Scholz CH (1990) Anticrack-associated faulting at very high pressure in natural olivine. *Nature* 348:720-722
- Green HW II, Zhou Y (1996) Transformation-induced faulting requires an exothermic reaction and explains the termination of earthquakes at the base of the mantle transition zone. *In* Avé Lallemant et al. (eds) *Carter Volume. Tectonophysics* 256:39-56
- Griggs DT, Baker DW (1969) The origin of deep-focus earthquakes. *In* *Properties of Matter Under Unusual Conditions*. H Mark, S Fernbach (eds) New York: Wiley Interscience, p 23-42

- Gruntfest IJ (1963) Thermal feedback in liquid flow—plane shear at constant stress. *Trans Soc Rheology* 7:195-207
- Hobbs BE, Ord A (1988) Plastic instabilities: implications for the origin of intermediate and deep focus earthquakes. *J Geophys Res* 93:23-42,10521-10540
- Igarashi I, Matsuzawa T, Umino N, Hasegawa A (2001) Spatial distribution of focal mechanisms for interplate and intraplate earthquakes associated with the subducting Pacific plate beneath the northeastern Japan arc: A triple-planed deep seismic zone. *J Geophys Res* 106:2177-2191
- Jiao W, Silver PG, Fei Y, Prewitt CT (2000) Do deep earthquakes occur on preexisting weak zones? An examination of the Tonga subduction zone. *J Geophys Res* 105:28125-38
- Kanamori H, Anderson DL, Heaton TH (1998) Frictional melting during the rupture of the 1994 Bolivian earthquake. *Science* 279:839-842
- Karato S, Riedel MR, Yuen DA (2001) Rheological structure and deformation of subducted slabs in the mantle transition zone: implications for mantle circulation and deep earthquakes. *Phys Earth Planet Inter* 127:83-108
- Kirby SH, Durham WB, Stern L (1991) Mantle phase changes and deep earthquake faulting in subducting lithosphere. *Science* 252:216-25
- Kirby SH, Stein S, Okal EA, Rubie DC (1996) Metastable mantle phase transformations and deep earthquakes in subducting oceanic lithosphere. *Rev Geophys* 34, 261-306
- Koper KD et al. (1998) Modeling the Tonga Slab: can travel time data resolve a metastable olivine wedge? *J Geophys Res* 103:30079-30100
- McGuire JJ, Wiens DA, Shore PJ, Bevis MG (1997) The March 9, 1994 deep Tonga earthquake: Rupture outside the seismically active slab. *J Geophys Res* 102:15163-15182.
- Meade C, Jeanloz R (1991) Deep-focus earthquakes and recycling of water into the earth's mantle. *Science* 252:68-72
- Mosenfelder JL, Marton FC, Ross CR, et al. (2001) Experimental constraints on the depth of olivine metastability in subducting lithosphere. *Phys. Earth Planet. Inter* 127:165-180
- Ogawa M (1987) Shear instability in a viscoelastic material as the cause of deep focus earthquakes. *J Geophys Res* 92:13,801-13,810
- Peacock SM (2001) Are the lower planes of double seismic zones caused by serpentine dehydration in subducting oceanic mantle? *Geology* 29:299-302
- Phipps-Morgan J (2001) The role of serpentinization and deserpentinization in bending and unbending the subducting slab. EOS, *Trans Am Geophys Union* 82, Fall Mtg Suppl, Abstr T22D-03
- Phipps-Morgan J, Ranero C, Ruepke L (2002) Serpentinization, deserpentinization, and the bending and unbending of subducting slabs (abstr) *Proc Structure and Tectonics of Convergent Plate Margins*. Prague, Czech Republic (in press)
- Raleigh CB (1967) Tectonic implications of serpentinite weakening. *Geophys J Royal Astron Soc* 14: 45-51
- Raleigh CB, Paterson MS (1965) Experimental deformation of serpentinite and its tectonic implications. *J Geophys Res* 70:3965-3985
- Reinen LA, Green HW II, Nielsen SK (1998) Dehydration embrittlement of antigorite serpentinite at high pressures: Implication for intermediate depth earthquakes. EOS, *Trans Am Geophys Union* 79:F853
- Rice JR, Ruina AL (1983) Stability of steady frictional slipping. *J Appl Mech* 50, 343-49
- Scholz CH (1990) *The Mechanics of Earthquakes and Faulting*. Cambridge, UK: Cambridge University Press, 439 p
- Scholz CH (2002) *The Mechanics of Earthquakes and Faulting*, 2nd Edition. Cambridge, UK: Cambridge University Press, 496 p
- Silver PG, Beck SL, Wallace TC, Meade C, Myers S, James D, Kuehnel R (1995) The rupture characteristics of the deep Bolivian earthquake of 1994 and the mechanism of deep-focus earthquakes. *Science* 268:69-73
- Ulmer P, Trommsdorff V (1995) Serpentine stability to mantle depths and subduction related magmatism. *Science* 268:858-861
- Ulmer P, Trommsdorff V (1999) Phase relations of hydrous mantle subducting to 300 km. *In* Fei Y, Bertka CM, Mysen BO (eds) *Mantle Petrology: Field Observations and High Pressure Experiments*. *Geochem Soc Spec Publ* 6:259-281
- Wiens DA, McGuire JJ (1995) The 1994 Bolivia and Tonga events: Fundamentally different types of deep earthquakes? *Geophys Res Lett* 22:2245-2248.
- Wiens DA, McGuire JJ, Shore PJ (1993) Evidence for transformational faulting from a deep double seismic zone in Tonga. *Nature* 364:790-793

- Wiens DA, McGuire JJ, Shore PJ, Bevis MG, Draunidalo K, Prasad G, Helus SP (1994) A deep earthquake aftershock sequence and implications for the rupture mechanism of deep earthquakes. *Nature* 372: 540-543
- Wong T-F, Ko SC, Olgaard DL (1997) Generation and maintenance of pore pressure excess in a dehydrating system, 2. Theoretical analysis. *J Geophys Res* 102:841-852