Geological Magazine

www.cambridge.org/geo

Failure modes in fluid saturated rocks: deformation processes and mode-switching

Bruce Hobbs[1,2] and Alison Ord[1]

[1]Centre for Exploration Targeting, School of Earth Sciences, University of Western Australia, Perth, WA, 6009, Australia and [2]CSIRO, Perth, WA, 6102, Australia

Original Article

Cite this article: Hobbs B and Ord A (2023) Failure modes in fluid saturated rocks: deformation processes and mode-switching. *Geological Magazine* **159**: 2002–2019. https://doi.org/10.1017/S0016756822000516

Received: 4 December 2021
Revised: 6 May 2022
Accepted: 6 May 2022
First published online: 2 August 2022

Keywords:
capped yield surface; solution seams; laminated veins; reactivated faults; fault-valve mechanism; mode-switching mechanism; axial plane veins; breccias

Author for correspondence:
Alison Ord, Email: alison.ord@uwa.edu.au

Abstract

We consider the implications of adding a cap to the yield surface for elastic–plastic and elastic–visco-plastic solids with coupling between deformation, fluid flow and mineral reactions. For a suitable combination of (low) permeability and strain rate, opening-mode veins can form in compression. Such behaviour is enhanced by dissolution and by simultaneous mineral reactions with negative ΔV. These are the opening-mode equivalents of compaction bands in rocks with high permeability. Stylolite parallel veins are considered as forming in this way; such veins are commonly the laminated or ribbon-quartz veins associated with intense gold mineralization. Axial plane veins and melt segregations are in this class also. The addition of a yield surface cap limits the permissible stress states portrayed by a failure-mode diagram and has implications for breccia formation. Failure discontinuities that form at the cap require a *decrease* in fluid pressure to form as opposed to extension joints and veins that require an *increase* in fluid pressure; discontinuities that form at the cap are in orientations that are commonly interpreted as reactivated early discontinuities. The switching between high fluid pressure and low fluid pressure, which we call *mode-switching*, arises from competition between mineral dissolution and deposition. This is an alternative to the fault-valve mechanism and does not require fault reactivation or failure at a 'seal' linked to seismicity or fault reactivation. The capped yield surface concept provides a unifying self-consistent approach for vein/breccia formation and for the kinematics of brittle and visco-plastic rocks.

1. Introduction

This paper is concerned with the various modes of failure that occur in fluid saturated, elastic–plastic and elastic–visco-plastic rocks. The common way of approaching this topic for pressure sensitive plastic materials (Sibson, 1981; Cox, 2010) is based on a Mohr–Coulomb failure criterion with no cap (Borja & Aydin, 2004; see Section 2) on the yield surface. The lack of inclusion of a cap implies the rock can sustain unlimited compressive and shear stresses without yielding. Even if a yield surface cap is included (Issen & Rudnicki, 2000) the yield in compression is considered to be by grain crushing, and compaction bands are the result (Fossen *et al.* 2007). Veveakis & Regenauer-Lieb (2015) and Alevizos *et al.* (2017) have described situations for capped yield surfaces and fluid saturated rocks where opening-mode displacement discontinuities can arise in compression for suitable combinations of strain rate and permeability. The opening-mode behaviour is enhanced by coupling between chemical reactions (especially those with large negative ΔV or involving dissolution) and deformation (Alevizos *et al.* 2017). In this paper we first concentrate on vein systems and extend the framework already established by the above authors to include coupling between deformation, fluid flow and mineral reactions in materials with capped yield surfaces. The result is a new interpretation of some classes of veins that form as opening structures in the compaction field of deformation. These opening-mode structures are exemplified by the two types of veins shown in Figure 1. In Figure 1a solution seams or stylolites are shown *parallel* to the vein margin whereas in Figure 1b stylolites are *normal* to the vein boundary. With the assumption that the teeth on stylolites point towards σ_1 (Toussaint *et al.* 2018), this means that the vein in Figure 1a forms normal to σ_1 whereas that in Figure 1b forms parallel to σ_1. The addition of a cap on the yield surface also has important implications for failure-mode diagrams (Cox, 2010), fault reactivation (Sibson *et al.* 1988), breccias and for kinematic interpretations of vein systems in general. In particular an aseismic alternative to the fault-valve mechanism (Sibson, 2020), which we call *mode-switching*, arises. This paper is concerned with these implications and processes.

Various authors (for instance, Pollard & Segall, 1987; Pollard & Aydin, 1988; Schultz & Fossen, 2008) have developed terminology to describe displacement discontinuities in brittle rocks. We extend the spirit of that terminology to describe the kinematics of veins as shown in Table 1. The term *displacement discontinuity* is used as a general term to include veins, compaction bands, joints, shear zones and faults. In this paper compressive stresses are taken to be positive. The term *yield surface* is used as that surface in stress space that defines the change from

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Table 1. Kinematic framework for displacement discontinuities associated with veins in brittle rocks

	Displacement state	
Displacement mode	Compression	Extension
Opening	Veins parallel to solution seams and stylolites (as in Fig. 1a)	Crack-seal veins (Bons *et al.* 2012) Open space veins Stylolite normal veins (as in Fig. 1b)
Shearing	Closing shear vein (Eichhubl *et al.* 2010)	Opening shear vein (Bons *et al.* 2012)
Closing	Compaction bands (Issen & Rudnicki, 2000)	Closing vein (No known example)

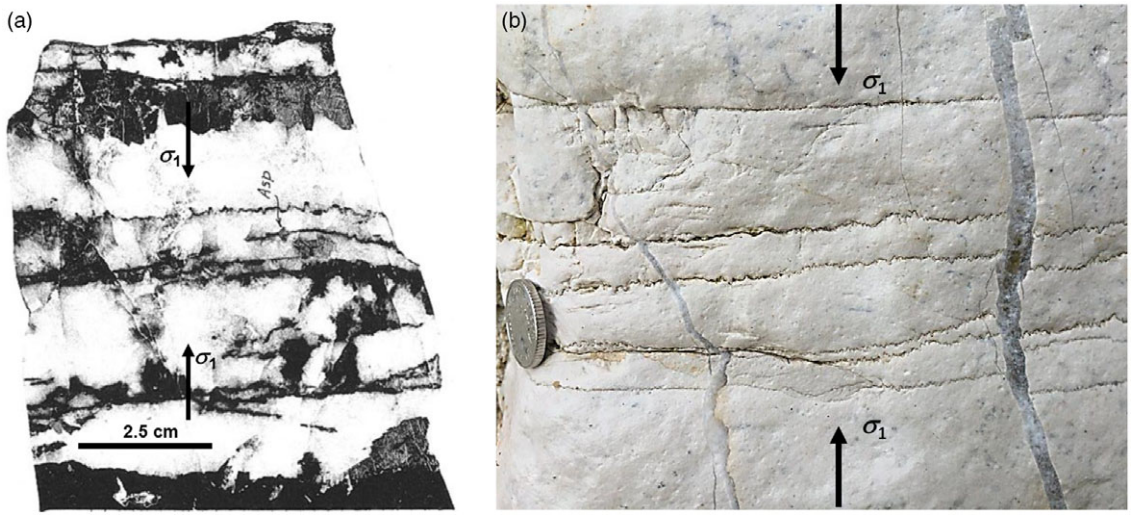


Fig. 1. Two different types of veins. (a) Laminated quartz vein in relatively impermeable metamorphosed pelitic host rock. Solution seams and stylolites are parallel to vein boundary. Bendigo, Australia. Reprinted from Chace (1949). (b) Stylolites normal to vein boundary in limestone.

elastic behaviour, inside the surface, to rate independent plastic or rate dependent visco-plastic behaviour on the surface. The yield surface can change its position, shape and size as deformation evolves. The term *failure surface* refers to the displacement discontinuity that develops upon yielding.

The classical view of the origin of fracturing associated with vein formation (Bons *et al.* 2012) employs the Mohr–Coulomb yield envelope and the concept of effective stress: the material yields when the Mohr stress circle touches the yield surface; this happens when the fluid pressure is increased (Fig. 2a). Although this approach has been and continues to be extremely useful, there are limitations:

- The initial stress difference, $(\sigma_1 - \sigma_3)$, has to be small (less than the cohesion) for the increase in fluid pressure, P_f, to produce a new effective stress state that touches the yield envelope at the tension cut-off so that pure extension fractures can form. Here σ_1 and σ_3 are the maximum and minimum principal stresses with compressive stress taken to be positive.
- If σ_3 = lithostatic load = ρgh, then increasing the fluid pressure to lithostatic may not result in the effective stress circle touching the yield surface at the tension cut-off (Fig. 2a). In general it touches before reaching the tension cut-off resulting in opening-mode shear fracture. Here, ρ is the rock density, g is the acceleration due to gravity and h is the distance below the surface of the Earth.
- The yield surface is not bounded for increasing normal stress. In fact as indicated above, Figure 2a implies that the material can sustain infinite normal and shear stresses without yielding. In order to address the last dot point a 'cap' is added to the yield surface as indicated in Figure 2b.

The structure of this paper is as follows. In Section 2 we discuss capped yield surfaces and the planar failure discontinuities predicted by classical approaches. In Section 3 failure modes in the capped region of the yield surface are considered for deformations coupled with fluid flow and mineral reactions; the concept of *cnoidal waves* is introduced. Section 4 considers the resulting implications for various vein types. Section 5 considers the significance of a capped yield surface for failure-mode diagrams (Cox, 2010) whilst Section 6 considers the orientations of these failure modes and shows that failure at the cap produces discontinuities in orientations otherwise interpreted as reactivated features. Section 7 considers concepts of fault reactivation/fault-valve models and introduces the concept of *mode-switching*. Section 8 discusses axial plane structures such as veins and melt segregations whilst a discussion of breccias is included in Section 9. Section 10 is a general discussion and conclusions are drawn in Section 11.

2. Capped yield surfaces

The classical Mohr–Coulomb failure criterion is open ended in the sense that the material can support an infinite normal or mean stress at an infinite shear stress without yielding as shown in Figure 2a. Real materials are not like this, and both the upper values

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

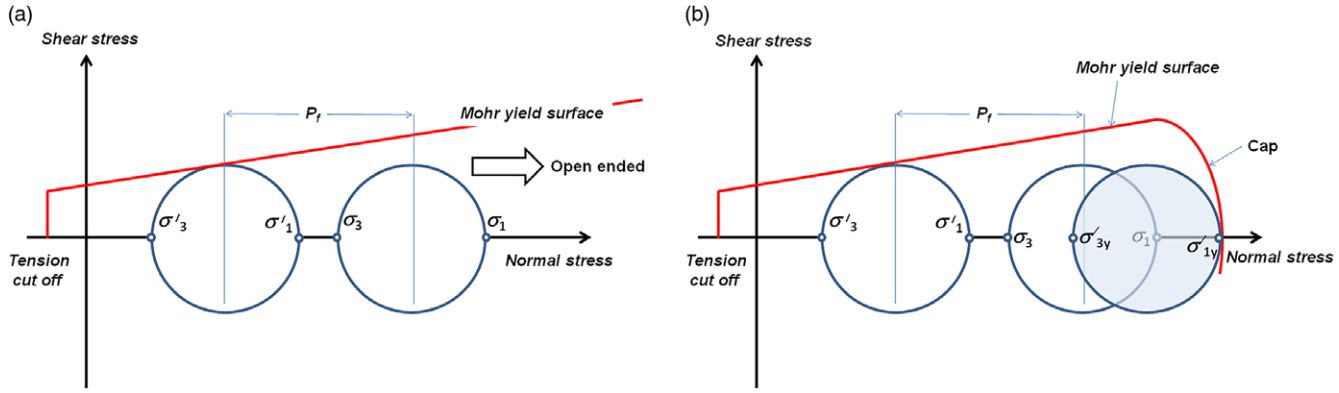


Fig. 2. Uncapped and capped yield surfaces. (a) The classical Mohr–Coulomb failure criterion. Failure occurs when the pore pressure is increased by P_f so that the Mohr stress circle touches the yield surface. The form of the yield surface implies the normal stress can increase indefinitely without resulting in yield. (b) In order to define an upper limit to the normal stress that a material can sustain, a 'cap' is added to the classical yield surface. The material can now yield by *increasing* the fluid pressure and by *decreasing* the fluid pressure. Similar behaviour is discussed by Rudnicki (2000) and Bésuelle & Rudnicki (2004). The material can also yield by keeping the fluid pressure constant and increasing ($\sigma_1 - \sigma_3$) as in Figure 4a. Stress states with primes are effective stresses.

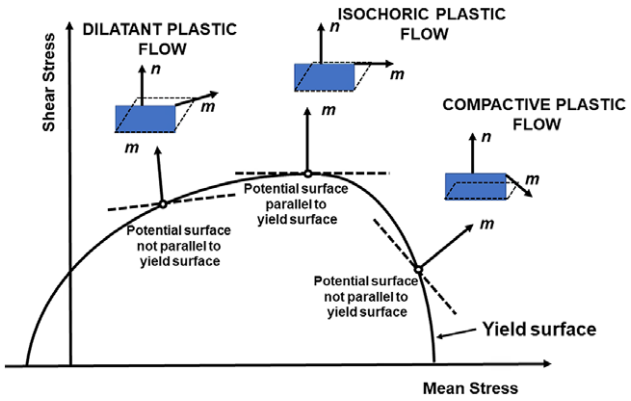


Fig. 3. Model of a capped yield surface (after Aydin *et al.* 2006). Capped yield surface in mean stress – shear stress space with deformation modes shown. ***n*** is the normal to the failure surface in the material; ***m*** is the incremental strain rate vector. The orientation of the potential surface, the slope of which is $\tan^{-1}$ (dilation angle) is marked in each sector of the yield surface. The dilation angle is the angle ***m*** makes with the shearing plane.

of shear stress and normal stress that can be supported without yielding are *capped* (Schofield & Wroth, 1968; Dimaggio & Sandler, 1971; Fossum *et al.* 1995; Wong *et al.* 1997; Olsson, 1999; Fossum & Fredrich, 2000). A model for such capped yield surfaces is shown in Figure 2b. The various modes of yielding are shown in Figure 4b. An alternative way of portraying the yield surface, applicable both to pressure sensitive brittle materials and to rate sensitive visco-plastic materials is Figure 3 in mean stress, shear stress space. This latter way is the dominant way of portraying stress states at yield in the mechanics literature, but we restrict ourselves initially to the use of Mohr-space except to point out in the Discussion that Cox-failure-mode diagrams (Cox, 2010) are a special case of Figure 3.

3. Failure modes in the capped region

Failure in the capped region of the yield surface is commonly considered as comprising closing-mode, planar compaction bands normal to σ_1. Compaction bands were first discussed from a mechanics point of view in Olsson (1999) and Issen & Rudnicki (2000), although hints were made in Vardoulakis & Sulem (1995); the literature until recently has considered a *decrease* in porosity within a localized layer as the only form of failure in the so called 'compactional regime' (see Figs 3, 4b). The idea that compaction bands exist has co-existed with a large number of field studies (see Fossen *et al.* 2007 for a review) where compaction bands have been documented in initially high porosity sandstones. However, recently, it has been suggested (Veveakis & Regenauer-Lieb, 2015; Regenauer-Lieb *et al.* 2016; Alevizos *et al.* 2017) that such behaviour is not the only form of yield and the inverse is also possible where the yielding layer constitutes an *increase* in porosity. Thus, the deformation in the unstable layer is *dilatant* rather than *compactive* so that dilatant instabilities form as layers normal to the maximum compressive stress. If the deformation is coupled to fluid flow and dissolution/mineral reactions then opening-mode veins can form normal to compression (Alevizos *et al.* 2017).

To give some physical insight into this counter-intuitive behaviour we note that natural examples of opening-mode veins formed in compression (Fig. 1a) occur typically within altered chlorite-rich rocks derived through hydrothermal alteration of mafic or sedimentary rocks. A common mineral reaction involved here is:

$$3\,\textit{tremolite} + 2\,\textit{clinozoisite} + 10\,CO_2 + 8\,H_2O + 10\,O_2 \rightarrow 3\,\textit{chlorite} + 10\,\textit{calcite} + 6\,\textit{quartz}$$

This reaction is exothermic and has $\Delta V = +4.45$ %, but if quartz and calcite are removed in solution (to be deposited in the vein) the effective decrease in volume is −41.9 %. Thus, if reactions (including dissolution) such as this proceed during deformation with stress states at the cap part of the yield surface, opening-mode discontinuities are easily accommodated by the decrease in volume generated by the alteration/dissolution reactions. A selection of these mineral reactions is given by Haack & Zimmermann (1996, table 1) where large ΔV's are reported for reactions involving no removal of quartz.

The formation of these opening-mode discontinuities in compression arises from competition between the fluid diffusivity and the deformation diffusivity in the compression direction. If the fluid diffusivity exceeds the deformation diffusivity then opening-mode discontinuities form; otherwise closing-mode compaction bands form. Veveakis & Regenauer-Lieb (2015) showed that the development of opening-mode instabilities normal to

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

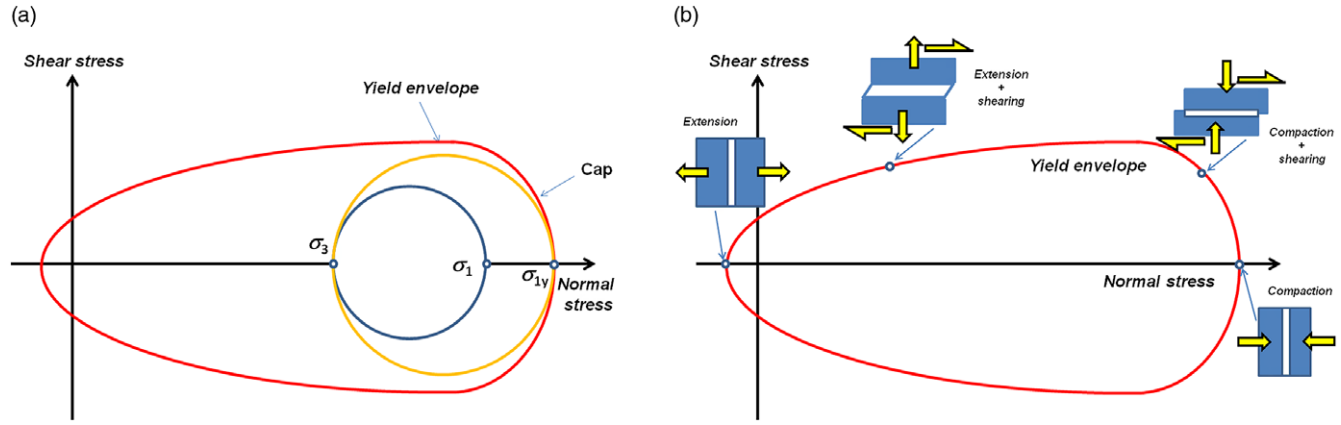


Fig. 4. A realistic yield surface with curved boundaries. (a) Yield by *increasing* $(\sigma_1 - \sigma_3)$. The material yields at $\sigma_1 = \sigma_{1y}$ (the orange circle). (b) The various modes of yielding: pure extension fracture, extension plus shearing, compaction plus shearing and pure compaction.

compression is described by a reaction diffusion equation involving the diffusion of the fluid pressure, the volumetric plastic strain rate and the mass balance arising from mineral reactions. The solution to this reaction diffusion equation involves elliptical functions, *sn* and *cn*, where *sn* and *cn* refer to the elliptical *sn*- and *cn*-functions, which are the elliptical equivalents of the circular *sine*- and *cosine*-functions (Schwalm, 2015). The patterning of opening discontinuities in compression is expressed as *cnoidal* waves (Veveakis & Regenauer-Lieb, 2015). A detailed elaboration of the theory behind failure at the capped region is Cier *et al.* (in press) where the acoustic tensor is used to define the conditions for bifurcation from the homogeneous to the localized state.

The condition required to form dilatant bands in compression is that a critical value of a parameter, λ, is exceeded:

$$\lambda = \frac{mechanical\ diffusivity}{fluid\ diffusivity} = \frac{\dot{\varepsilon}_n \mu}{k p_n'} H^2 \lambda_{crit}$$

where $\dot{\varepsilon}_n$ is the strain rate in the loading direction, μ is the fluid viscosity, k is the permeability, p_n' is the volumetric effective mean stress and H is a length scale in the compression direction. As an example, if $\dot{\varepsilon}_n = 10^{-10}\ s^{-1}$, $\mu = 10^{-3}$ Pa s, $k = 10^{-18}\ m^2$, $p_n' = 50$ MPa and $H = 100$ m, then $\lambda = 20$. In the theory developed by Alevizos *et al.* (2017), $\lambda_{crit} = 13$ (Fig. 5) and the number of discontinuities is $0.26\sqrt{\lambda} = 1.2$ so that quite widely spaced (relative to H) opening-mode discontinuities are expected. We note that Cier *et al.* (in press) have modifications to this result. Thus, in a given system, dilatant compressive layers are favoured by (1) high strain rates, (2) low permeability and (3) low volumetric effective mean stresses, and are therefore expected in relatively impermeable pelites whereas compaction compressive bands are expected in permeable arenites. The development of dilatant compressive layers is a play-off between the rate of deformation, and/or the rate of volumetric strain arising from mineral reactions, and the rate at which fluid can be supplied to the dilating site. The spacing between dilatant compressional bands is inversely related to λ (Fig. 5a–d). In Figure 5a–d there is no coupling between deformation/fluid flow and mineral reactions or dissolution so that the discontinuities have zero thickness.

Although the spacing between cnoidal discontinuities is described by λ, the thickness is controlled by the coupling between fluid flow and chemical reactions including dissolution (Stefanou & Sulem, 2014; Alevizos *et al.* 2017). Now a parameter η, defined by $\eta = \frac{chemical\ diffusivity}{fluid\ diffusivity}$ is the controlling factor. η is temperature and fluid pressure dependent so that the behaviour of the system depends on the heat balance involving the relative rates of exothermic deformation and exothermic/endothermic mineral reactions such as mineral precipitation/dissolution or melt crystallization/production. An example is shown in Figure 5e, f where there is coupling between deformation/fluid flow and mineral reactions or dissolution and the discontinuities are wide.

In summary,

- In fluid saturated rocks with a cap on the yield surface, some veins form at sites where compressive dilatant bands develop whilst fluids deposit quartz and carbonates in these dilating layers. Solution seams within these layers remove material from the system. Such veins are *normal* to the principal compression (Fig. 1a) as opposed to extensional veins that form *parallel* to the principal compression (Fig. 1b). The space occupied by the opening-mode vein is offset by mineral reactions with negative ΔV and/or by dissolution in the matrix between veins.
- According to Alevizos *et al.* (2017), the spacing between these discontinuities is defined by λ whereas the thickness of the resulting layers is defined by η (Fig. 5e, f).
- Opening-mode compressional veins form in relatively *impermeable* lithologies whereas closing-mode compaction bands form in relatively *permeable* lithologies.

4. Laminated veins and crack-seal veins

The best candidates for dilatant compressional veins are the laminated veins characteristic of orogenic gold deposits (Fig. 1a). These almost invariably have stylolites and/or solution seams parallel to the vein margins confirming that σ_1 was normal to the veins at the time of formation of the stylolites. Stylolite parallel laminated quartz veins are known from around the world. Some examples are given by Ferguson & Gannett (1932; crinkly veins, California, USA), Chace (1949; Bendigo, Australia), Hough *et al.* (2010; Walhalla, Australia) and Cheong *et al.* (2000; Nevada, USA). The best candidates for classic crack-seal extension veins are those commonly observed in massive sandstones and limestones (Fig. 1b) and their metamorphosed equivalents (Bons *et al.* 2012).

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

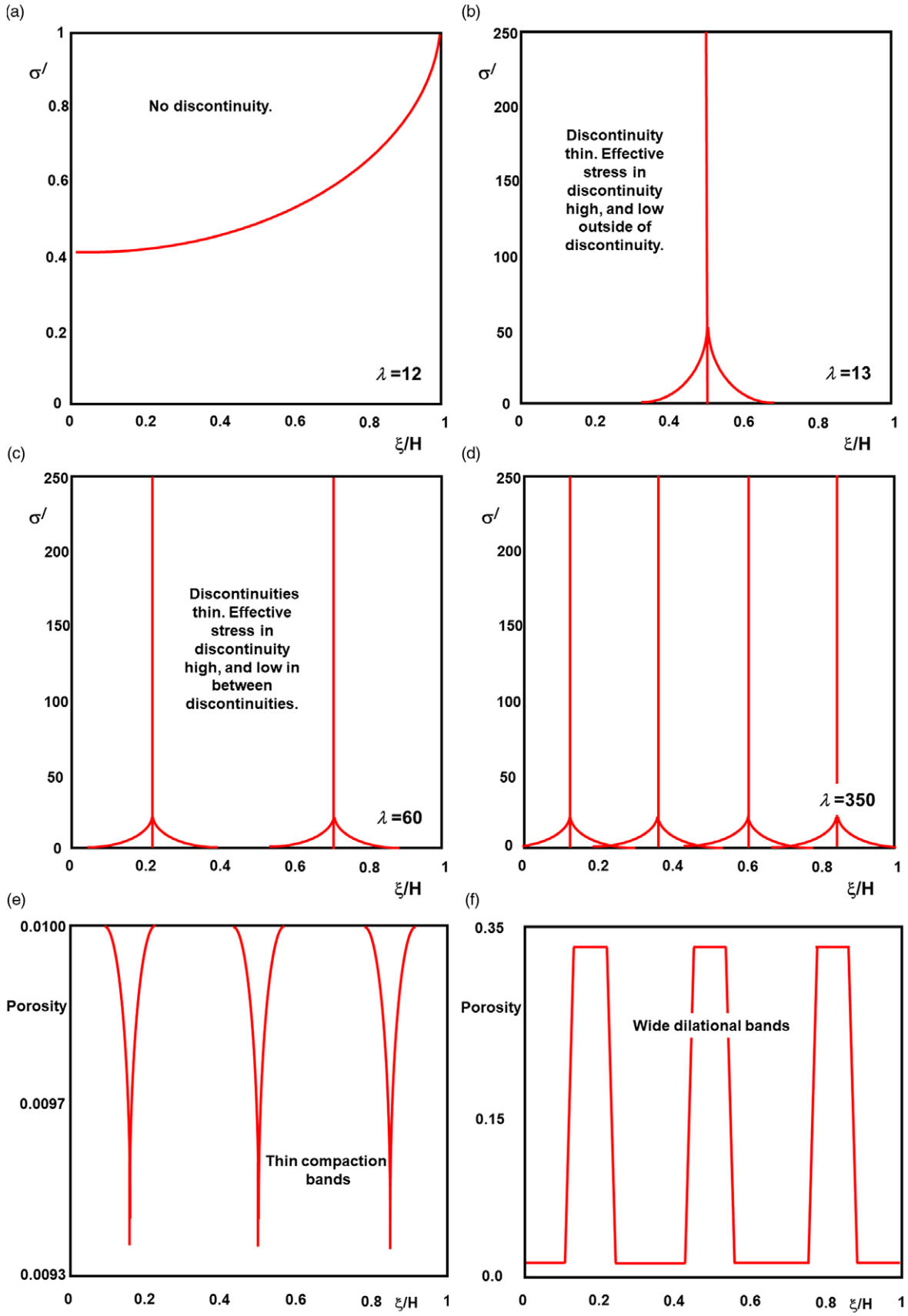


Fig. 5. Spacing of dilatant compression bands as a function of λ. (a–d) The critical value of λ is between 12 and 13 in these examples. The figures are plots of the effective stress against the normalized spacing, ξ/H. After Veveakis & Regenauer-Lieb (2015). (e, f) Plots of porosity against ξ/H after Alevizos *et al*. (2017). In (e) chemical reactions do not contribute to a volume change whereas in (f) they do.

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

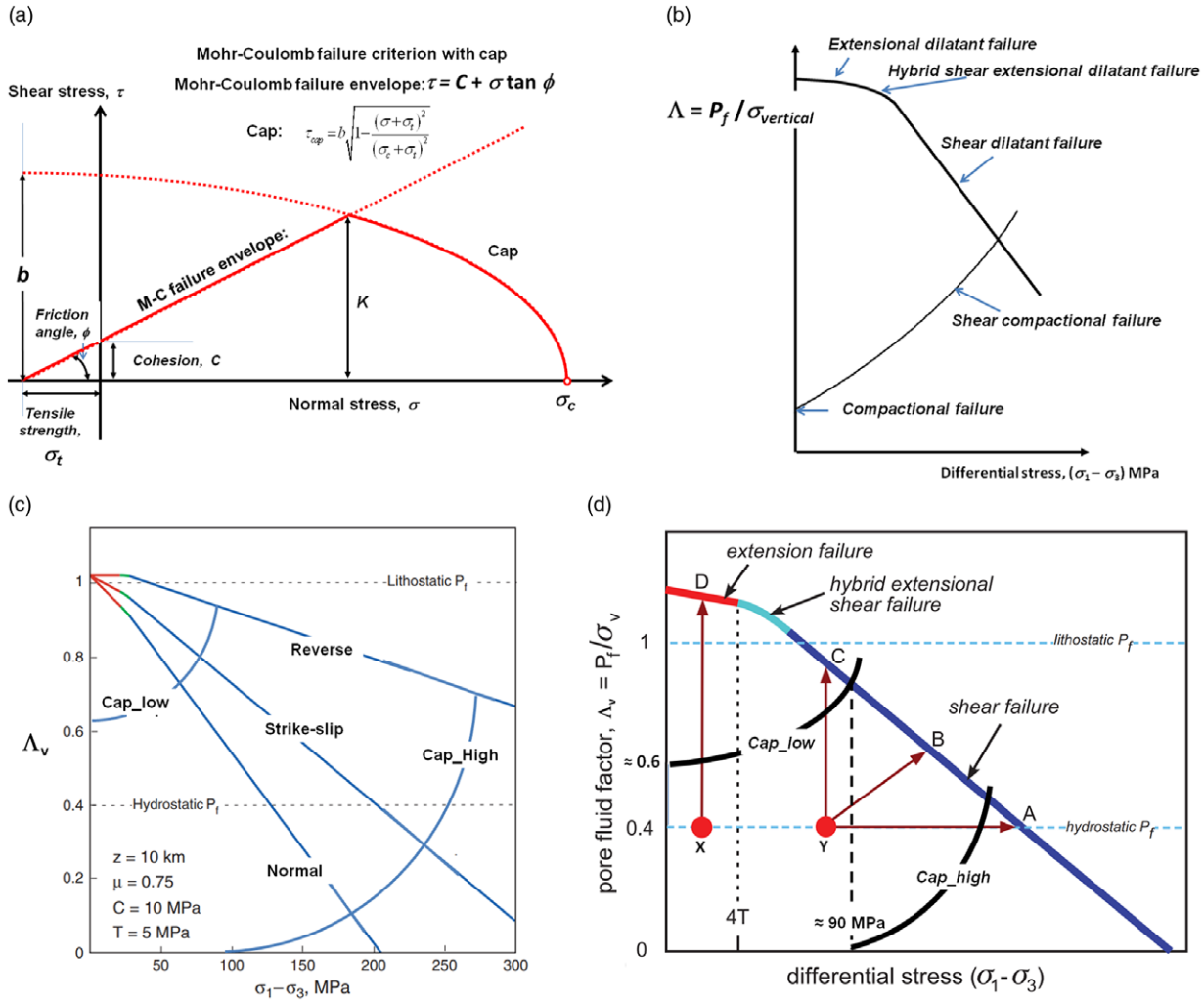


Fig. 6. Mohr–Coulomb failure criteria with elliptical cap. (a) Construction following Reiweger *et al.* (2015). (b) The standard Cox-failure-mode diagram with the additional failure modes arising from a cap. Calculations to define the cap are in the Appendix. (c) The Cox-failure-mode diagram with failure modes for reverse, strike-slip and normal faults indicated. The addition of a cap can severely limit the extent of the failure modes. Cap_low corresponds to $\sigma_c = 100$ MPa and $K = 75$ MPa; Cap_high corresponds to $\sigma_c = 200$ MPa and $K = 175$ MPa. (d) The generic Cox-failure-diagram with the caps in (c) added using the constitutive parameter values of Cox (2010). Initial stress states marked as X, Y and 'reactivated stress states', A, B, can be outside the yield surface and not possible for the assumed constitutive parameters and the low_cap position.

The notion of dilatant compressional veins is quite new and much needs to be understood. At present the theory is one-dimensional and so only planar layers are predicted at periodic spacings. A three-dimensional theory with permeability a function of position would produce non-periodic spacings, more than one set and perhaps non-planar layers. Finally, much more field and microstructural work is needed in parallel with theoretical work whilst shedding the blinkering associated with the classical concepts in Figure 2a. Independently of the theoretical development of this subject we rely on the worldwide observation, mainly from hydrothermal gold deposits, of laminated veins parallel to solution seams and stylolites to support the notion that opening-mode veins can form at high angles to compression.

5. Implications of a cap for failure-mode diagrams

In order to explore the implications of a yield surface cap for the failure-mode diagrams proposed by Cox (2010) we follow Reiweger *et al.* (2015) who developed a Mohr–Coulomb yield surface with an elliptical cap for snow. Elliptical caps in sandstone (Fig. 6a) are supported by the experimental work of Wong *et al.* (1997). The equation for the cap is written:

$$\tau_{cap} = b\sqrt{1 - \frac{(\sigma + \sigma_t)^2}{(\sigma_c + \sigma_t)^2}} \text{ where } b = K\sqrt{\frac{(\sigma_c + \sigma_t)^2}{(\sigma_c + \sigma_t)^2 - \left(\frac{K}{\tan\phi}\right)^2}}$$

This corresponds to an ellipse with centre at $(-\sigma_t, 0)$, major axis, $|\sigma_c + \sigma_t|$, and minor axis, b, as shown in Figure 6a. The maximum normal stress the material can sustain without collapsing is $\sigma = \sigma_c$ and the maximum shear stress the material can sustain without yielding is $\tau_{cap} = K$. Both σ_c and K are material parameters that are independent of each other, except that since σ_t is generally small, $\sigma_c > K$; having set their values, along with σ_t, b is defined. In addition, $(\sigma_c + \sigma_t) > \frac{K}{\tan\phi}$ must always be true for b to be real. The position of the cap for various values of σ_c and K is shown in Figure 7a using the parameters given by Cox (2010). The suggested upper and lower bounds for the cap in Figure 6c, d are prompted by the experimental ranges of σ_c and K reported by Wong *et al.* (1997). These are ≈200 to ≈400 MPa for σ_c and ≈150 to ≈250 MPa for K.

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

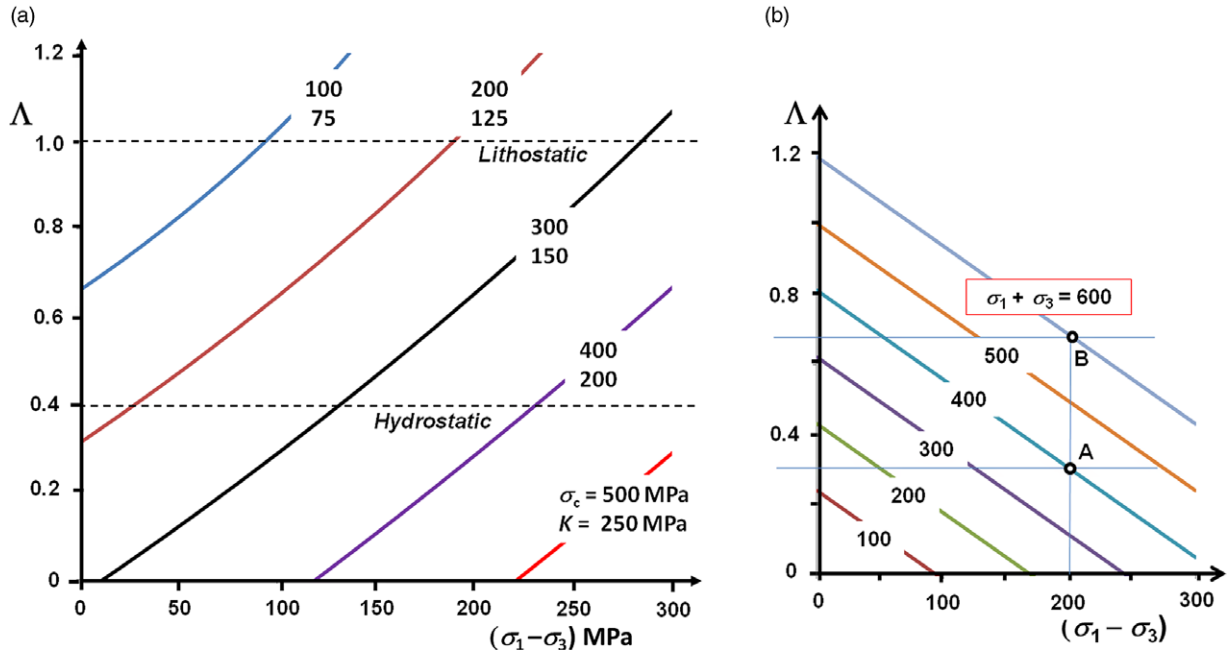


Fig. 7. (a) The cap plotted on a Cox-failure-mode diagram for various values of σ_c and K. Comparison with the diagrams proposed in Cox (2010) shows that the permissible region on the failure-mode diagram where stress states can be below yield can be restricted for some values of the cap parameters σ_c and K. (b) The maximum value of Λ that results in failure for a given value of $(\sigma_1 - \sigma_3)$ and of $(\sigma_1 + \sigma_3)$. For the same value of $(\sigma_1 - \sigma_3)$, at A and B, but different values of $(\sigma_1 + \sigma_3)$, different values of P_f and hence Λ result in yield.

A limitation of the Cox-failure-mode diagram is that a stress state with coordinates $((\sigma_1 - \sigma_3), \lambda)$ gives no information on the value of $(\sigma_1 + \sigma_3)$ unless some assumption is made regarding the value of σ_3. This means that if $(\sigma_1 - \sigma_3)$, the diameter of the Mohr circle, is given then there is no information of the position of the Mohr circle on the normal stress axis in normal stress – shear stress space. Thus, the fluid pressure required to move the Mohr circle so that it just touches the yield surface is not defined by the Cox-failure-mode diagram. This pressure is given by

$$P_f^{critical} = \frac{\sigma_1 + \sigma_3}{2} - \frac{\cot\phi}{2}((\sigma_1 - \sigma_3) - 2c)$$

$$\Lambda^{critical} = \frac{P_f^{critical}}{\rho g h} = \frac{1}{\rho g h}\left[\frac{\sigma_1 + \sigma_3}{2} - \frac{\cot\phi}{2}((\sigma_1 - \sigma_3) - 2c)\right]$$

where ρ is the mean rock density, g is the acceleration due to gravity and h is the depth below the surface.

$\Lambda^{critical}$ is plotted on Figure 7b for $h = 10$ km depth. This means that it may not be possible, depending on the value of $(\sigma_1 + \sigma_3)$, to move an arbitrary initial stress state such as X and Y to positions D and C on the yield surface in Figure 6d by the indicated changes in Λ as shown in that figure.

In the Appendix we derive the expression for P_f at failure in terms of $(\sigma_1 - \sigma_3)$ and the optimally oriented failure plane, θ, as

$$P_f = \left[\frac{1}{2}(\sigma_1 - \sigma_3)(1 + \cos 2\theta) + \sigma_3 + \sigma_t)\right] - \sqrt{a - (a - c)\frac{(\sigma_1 - \sigma_3)^2}{4K^2}\sin^2 2\theta}$$

where $a = (\sigma_c + \sigma_t)^2$ and $c = \left(\frac{K}{\tan\phi}\right)^2$.

Figure 7a shows a plot of $\Lambda_v = \frac{Fluid\ pressure}{Vertical\ stress}$ against $(\sigma_1 - \sigma_3)$ for various values of σ_c and K. One sees that the permissible stress states can be severely restricted compared to failure-mode diagrams with no cap.

6. Orientations of failure surfaces

In uncapped models of yield, the angle, θ, between the normal to the failure surface and the maximum compression axis, σ_1, is given by half the angle between the σ_N-axis and the line connecting the centre of the Mohr circle to the point of intersection of the Mohr circle with the yield surface (Fig. 8a, left side). In this case 2θ is always greater than 90° and so the failure surface is inclined at less than 45° to σ_1. In Figure 8a (left) 2θ is 120° and so the failure surface is inclined at 30° to σ_1 and has a reverse sense of movement. The same general rule holds for a capped yield surface (Fig. 8a, right) except that now 2θ is always less than 90° and the failure surface is inclined at greater than 45° to σ_1. In Figure 8a (right) 2θ is 45° and so the failure surface is inclined at 67.5° to σ_1 and has a reverse sense of movement if σ_1 is horizontal. This orientation and the associated reverse sense of displacement have been interpreted as inconsistent with Andersonian fault mechanics (Anderson, 1905, 1951) and hence necessitates reactivation of pre-existing faults (Sibson, 1985; Sibson *et al.* 1988; Cox, 2010). However, one can see that this geometry is an intrinsic result arising from failure at the cap of the yield surface and does not necessarily imply reactivation.

An additional important result follows from the geometry of the cap. If this cap is elliptical then for a critical radius, r, of the Mohr circle whose radius of curvature matches that of the ellipse and that just touches the cap at σ_c ($\theta = 0°$) there is just one failure surface normal to σ_1. All smaller Mohr circles (Fig. 8b) result in $\theta = 0°$. The critical radius for an elliptical cap is $r = \frac{b^2}{a}$ where $2a$, $2b$ are the major and minor axes of the ellipse. Thus, the angle between

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

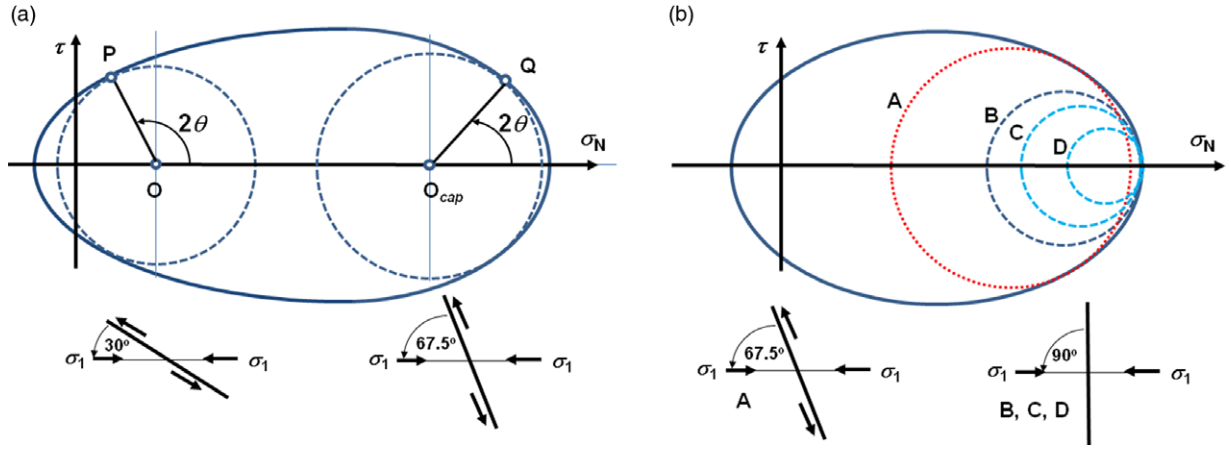


Fig. 8. Orientations of failure surfaces associated with a capped yield surface. θ is the angle between σ_1 and the normal to the failure surface. (a, left) 2θ for failure on the uncapped yield surface. The angle between the failure surface and σ_1 is always less than 45°. (a, right) 2θ for failure on the capped yield surface. The angle between the failure surface and σ_1 is always greater than 45°. (b) Failure on the cap for various values of the Mohr circle diameter. Once the radius of curvature, r, of the Mohr circle matches that of the ellipse, θ drops to zero and remains at zero for all smaller Mohr circles. Note that this same argument follows for the tensile end of the yield surface if it is elliptical.

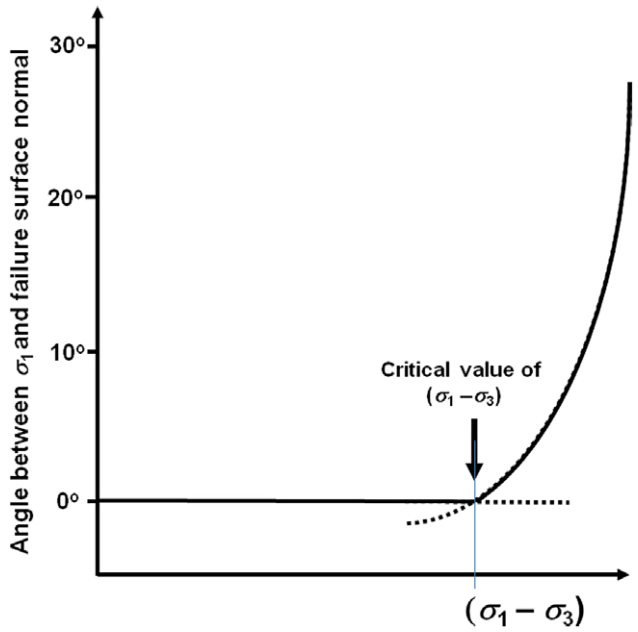


Fig. 9. Angle between the normal to the failure surface and σ_1 for cap failure as the diameter, $(\sigma_1 - \sigma_3)$, of the Mohr circle increases at the compressive end of the yield surface. For a range of diameters, this angle is zero, and at a critical diameter, the angle rapidly increases to ~30°. The value of the critical diameter depends on the shape of the cap and on the constitutive parameters of the material (Rudnicki, 2004).

the failure surface and σ_1 moves rapidly between ~60° and 90° as the diameter of the Mohr circle becomes smaller (Fig. 9). A detailed discussion of this variation in θ is given by Rudnicki (2004). Figure 9 supports a common observation in natural examples that failure surfaces are either normal to compression or at an angle of 60–70° to the compression axis, with angles in the range <90° to >60° rare. The same is true at the tensile end of the yield surface (if the yield surface is rounded) where failure surfaces in natural examples tend to be parallel or at 30° to the compression axis. Rarely, *hybrid* failure surfaces are reported with a small angle between the failure surface and σ_1 (Price & Cosgrove, 1990).

7. Implications for the fault-valve model

Cyclical and coupled changes in stress and fluid pressure states, with associated transitory fluid flow events during seismic cycles, are referred to as *fault-valve* behaviour (Sibson, 2020). Intrinsic to such a concept is the seismic cycle so that the whole process is driven by tectonic forcing and depends on failure defined by open-ended yield surfaces. Thus, in Figure 2a a seismic event is identified with an increase in fluid pressure causing the Mohr circle to touch the yield surface. The concept also involves episodic breaching and sealing of an 'impermeable seal' as fluid pressure increases and decreases. Here we explore the implications of adding a cap to the yield surface. This enables a different model to be developed that is driven by coupled deformation–chemical processes, requires no 'seal' and that is only incidentally related to seismic processes. We label this model the *mode-switching process*; we emphasize, seismicity is not crucial to this process.

Historically (Sibson, 1985), the fault-valve model depends on the geometry of the gold mineralized system at Val d'Or in Canada (Fig. 10, left-hand side). A steeply dipping mineralized zone comprising laminated quartz veins is identified with a reverse fault that is episodically cross-cut by opening-mode crack-seal veins. Recently, Cowan (2020) has pointed out that the mineralized zone is a shear zone occupying the steep dipping limb of a fold. Since the 'fault' does not have the geometry to be expected from Andersonian fault mechanics (Anderson, 1905, 1951; Scholz, 1989), Sibson (1985) proposed that the fault was a reactivated early normal fault. Episodic increases in fluid pressure are then responsible for episodic reactivation and associated seismic events. However, the introduction of a cap on the yield surface prompts a completely different model (Fig. 10, right-hand side), which we explore below.

7.a. The mode-switching model

We consider a vertical two-dimensional section through a region of the upper crust where brittle mechanisms in concert with pressure solution and fluid transport/deposition of dissolved materials are the essential modes of deformation (Wintsch & Yi, 2002; Gratier *et al.* 2013). Fluid is injected into the base of this region at a

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

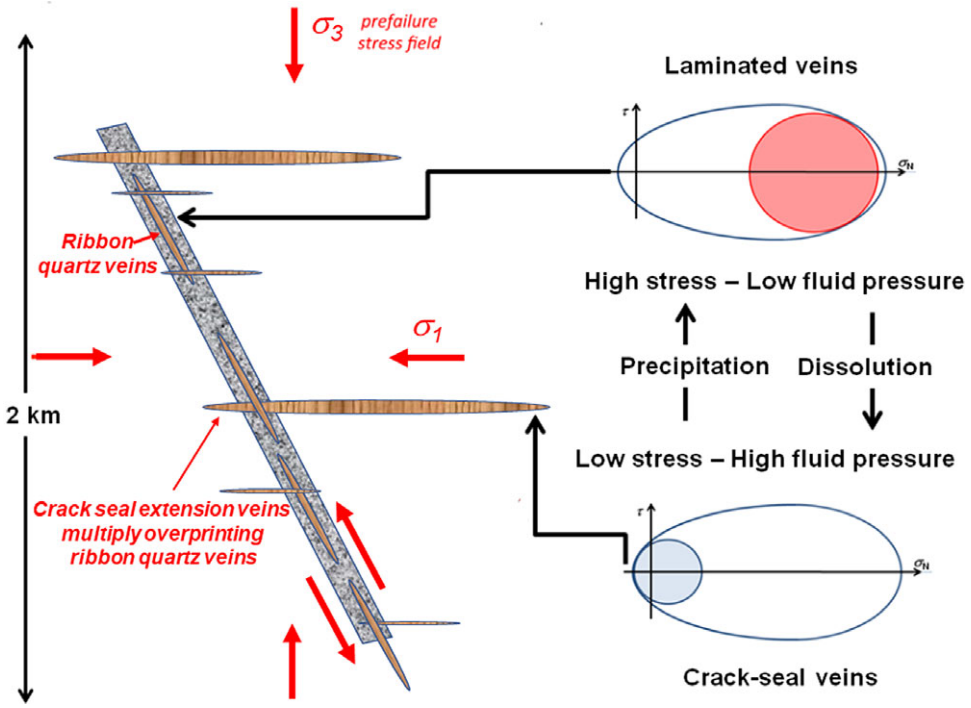


Fig. 10. Interpretation of the Sibson (2020) model of Val d'Or geometry in terms of a mode-switching model. Adapted from Sibson (2020).

constant Darcy velocity, $\hat{V}$, and the fluid pressure or fluid flux is fixed at the top of the region. Then, inverting the normal way of expressing Darcy's law (Phillips, 1991; Zhao *et al.* 2008, pp. 7–15), we write

$$\rho_{fluid}g - \nabla P^{fluid} = \frac{\mu \hat{V}}{K}$$

where ∇P^{fluid} is the gradient in fluid pressure, K is the permeability, μ and ρ_{fluid} are the fluid viscosity and density, and g is the acceleration due to gravity. If a part of the region has permeability, K_1 and fluid pressure gradient, ∇P_1^{fluid}, and the permeability changes, by precipitation or dissolution, to K_2, whilst $\hat{V}$ remains constant in order to satisfy mass continuity, then, for constant μ and g (Zhao *et al.* 2008; Hobbs & Ord, 2015, pp. 386–8), the fluid pressure gradient changes to ∇P_2^{fluid} given by

$$\nabla P_2^{fluid} = \frac{K_1}{K_2}\nabla P_1^{fluid} + \rho_{fluid}g\left(1 - \frac{K_1}{K_2}\right)$$

Hence, if $K_1 = 0.1K_2$, then $\nabla P_2^{fluid} = 0.1\nabla P_1^{fluid} + 0.9\rho^{fluid}g$. Thus, if ∇P_1^{fluid} is lithostatic (1.7×10^4 Pa m^{-1}) then ∇P_2^{fluid} is 1.07×10^4 Pa m^{-1}, which is just above a hydrostatic gradient (10^4 Pa m^{-1}). Here we assume $\rho_{rock} = 2700$ kg m^{-3} and $g = 10$ ms^{-2}. The principle is illustrated in Figure 11a, b where, as an example, we take a block 30 m high.

The essence of the mode-switching model is shown in Figure 11c, d. An antiform is shown composed of material with permeability, K_5. This is embedded in material with permeability, K_4, with $K_5 < K_4$. In Figure 11c a fault with permeability K_3 cuts the hinge of the antiform and $K_3 > K_4 > K_5$. Fluid is injected at the base of the model with fixed Darcy velocity, $\hat{V}$, and leaves at the top with the same velocity. The fluid stream lines are shown in yellow and the fluid pressure contours in black. In Figure 11c the stream lines are focused into the fault and the fluid pressure gradient in the gap in the low permeability layer is decreased by dissolution relative to the surrounding material. In Figure 11d the permeability in the gap and the fault below the antiform is decreased by precipitation. The stream lines now are deflected around the position of the fault and the pressure gradient in the gap is now increased relative to the surroundings. This increase in fluid pressure results in dissolution by pressure solution and the permeability increases returning the situation to Figure 11c. The cycle then repeats driven by the imposed deformation. Although this model is not fully coupled, in that the dissolution/precipitation processes, and hence permeability changes, are not explicitly modelled, the model captures the essence of a fully coupled model. A fully coupled model is the subject of another paper in preparation.

The process known as *pressure solution* consists of dissolution of material at places where the Helmholtz energy arising from strain is high and transfer of that material is into a nearby fluid. The thermodynamics of the process was first discussed by Gibbs (1876) and elaborated upon by Kamb (1961), Cahn (1989) and Sekerka & Cahn (2004). Gibbs pointed out that the solution becomes supersaturated and, given the opportunity, precipitates on some other kinetically favourable surface (Frolov & Mishin, 2010). The process of transport of material from strained sites to deposition sites is known as *solution transfer*. The driving force for the dissolution process is the increased strain energy of the material expressed as the Helmholtz energy density, ψ. There have been various attempts to derive a constitutive law for materials undergoing deformation by pressure solution (see Gratier *et al.* 2013, table 2.3 for examples) assuming various geometrical and physical models for the process but, independently of the precise physics, dimensional analysis suggests that such a constitutive law for one dimension and isothermal small elastic strains is of the form

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

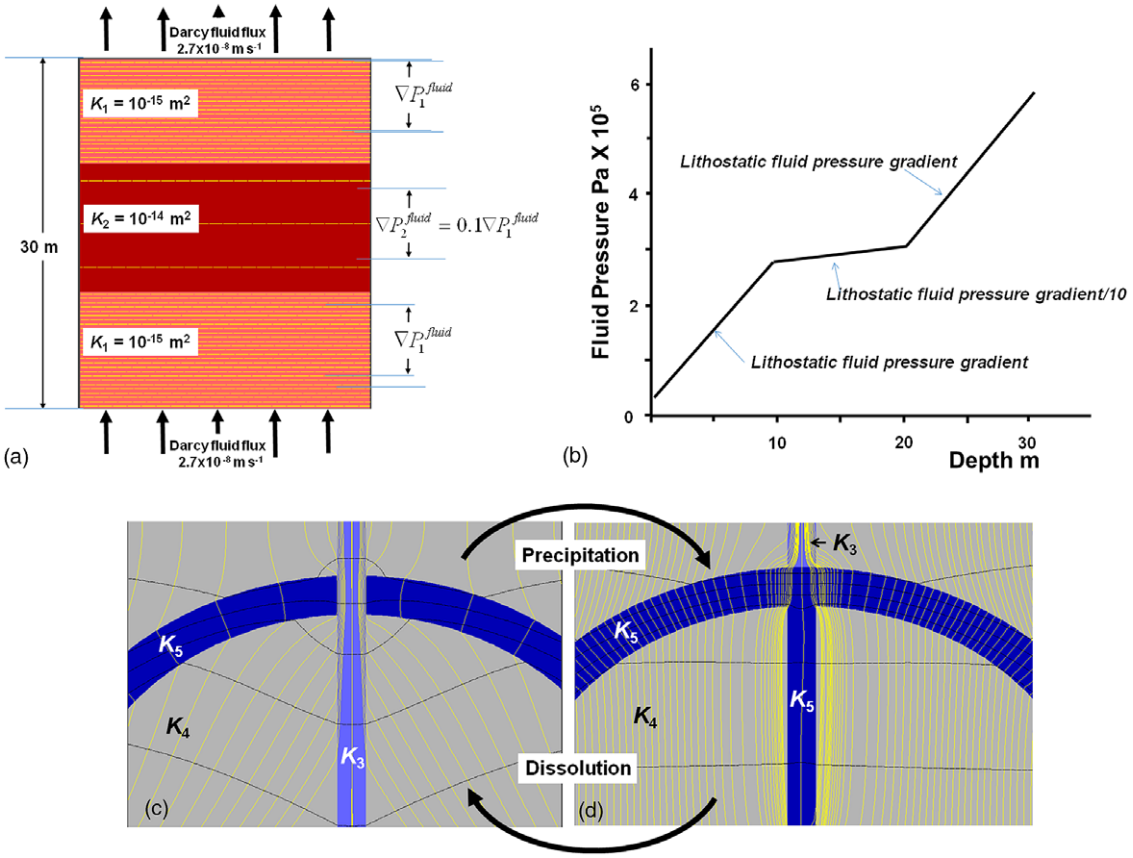


Fig. 11. Fluid flow models with changes in permeability. (a) A block of material with layered permeability. Fluid flux fixed at base and top at 2.7×10^{-8} m s^{-1}. Permeability (K_1) in bottom and top layers is 10^{-15} m^2. The middle layer has permeability $K_2 = 10^{-14}$ m^2. This means the fluid pressure gradient in the bottom and top layers is lithostatic. For mass continuity, the fluid pressure gradient in the middle layer is (lithostatic/10). The yellow lines are fluid pressure contours at 1×10^5 Pa spacing. (b) Fluid pressure gradients in (a). (c, d) The mode-switching model, see text for description. Yellow lines are fluid streamlines. Black lines are fluid pressure contours. These models were constructed using the finite difference code FLAC (ITASCA, 2008). The fluid viscosity is taken to be 10^{-3} Pa s.

$$\text{Strain Rate } \dot{\varepsilon} = \text{Geometrical Factor } A \times \text{Helmholtz Energy density } \psi \times \text{1/Shear modulus } 1/G \times \text{Dissolution rate constant } k$$

or,

$$\dot{\varepsilon} = A(t)\frac{\psi(t)}{G}k(t) = A(t)\frac{\sigma(t)\varepsilon}{G}k(t) \qquad (1)$$

Here we have written the Helmholtz energy density as the product of stress and strain. An expression in terms of the elastic moduli for a general stress state is given by Houlsby & Puzrin (2006, p. 78). Equation (1) is of the same form as derived by Rutter (1976), Paterson (1995) and Shimizu (1997, 1995) and others (Gratier *et al.* 2013, table 2.3) except that those authors assumed that only one component of the stress contributes to the energy driving dissolution and not the total strain energy. The distinction is important since neglecting the strain dependence introduces an error of order $\approx 10^{-3}$, the maximum value of the elastic strain. Dependence on grain size and grain boundary structure is included in the dimensionless geometrical factor, A. However if the stress depends on grain size or shape then an exponential dependence of strain rate on stress can be introduced (Gratier *et al.* 2013). The elastic shear modulus, G, is included to ensure the equation is dimensionally correct; neglect of a factor such as G introduces an error of $\approx 10^{11}$. Equation (1), in the absence of a grain size/shape stress dependence, predicts a linear dependence of strain rate on stress and upon dissolution rate as do Paterson and Shimizu and many other authors. Note that Equation (1) specifies a dependence of A, ψ and k on time arising from the evolution in geometry, stress and/or dissolution mechanisms. This introduces the possibility of Equation (1) being nonlinear in time as discussed by Gratier *et al.* (2013, section 3.4.4).

Thus, for a constant strain rate, the stress is inversely proportional to the dissolution reaction rate, which in turn is a function of the fluid pressure, pH, fluid chemical composition and temperature (Fournier & Potter, 1982; Dove & Rimstidt, 1994; Manning, 1994). Here we consider only the dependence on fluid pressure. The dependence of dissolution rate on fluid pressure arises from expressions such as equation (21) of Dove & Rimstidt (1994) where the dissolution rate constant for quartz is proportional to the square of the fugacity of H_2O. Thus, a decrease in fluid pressure results in decreased solubility of quartz and hence precipitation of quartz leading to decreased permeability. Importantly, decreased pressure also results in *decreased* dissolution rate that, through Equation (1), results in an *increase* in effective stress. The decrease in fluid pressure drives the Mohr circle to the cap end of the failure surface, and this coupled with high stress results in failure at the cap. The inverse is also true so that an increase in fluid pressure results in increased solubility and hence increased

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

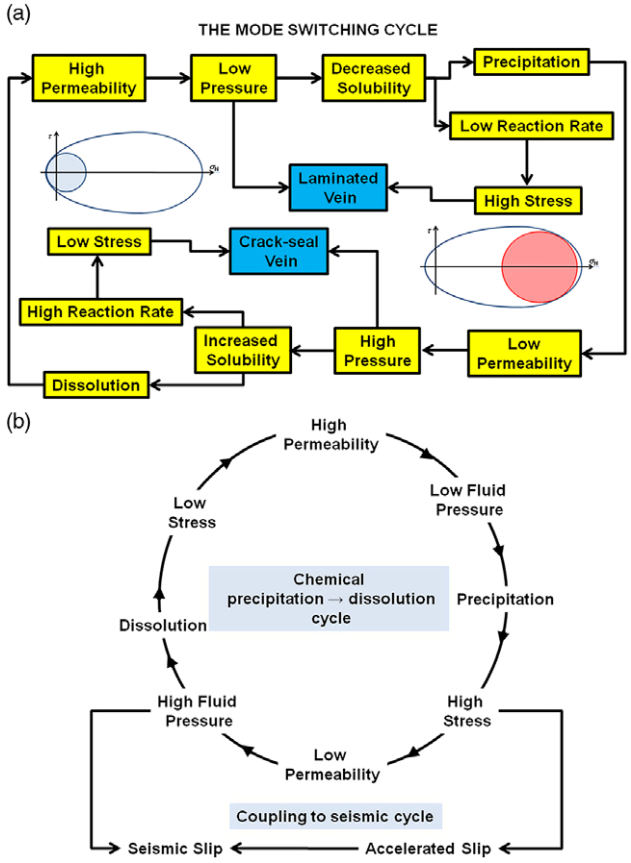


Fig. 12. The mode-switching cycle. (a) The complete cycle. High permeability leads to low fluid pressure and hence low equilibrium solubility and hence precipitation. The low fluid pressure also leads to a low reaction rate and hence, from Equation (1), high effective stress. These conditions of low fluid pressure and high stress lead to opening-mode discontinuities forming normal or at a high angle to compression. Precipitation produces low permeability and hence high fluid pressure and high equilibrium solubility and dissolution. The high fluid pressure also leads to a high reaction rate and hence, from Equation (1), low stress. These conditions of high fluid pressure and low stress initiate opening-mode discontinuities forming parallel or at a low angle to compression. The dissolution produces high permeability and the cycle repeats. (b) The mode-switching cycle coupled to the seismic cycle where high stress leads to accelerated slip and high fluid pressure leads to failure at the extensile end of the yield surface.

permeability. Increases in fluid pressure also lead to a *decrease* in the stress. The coupling between decreased stress and increased fluid pressure leads to failure at the extension end of the yield surface. The sequence of processes accompanying a decreased↔increased fluid pressure cycle is captured in Figure 12a.

Thus, the mode-switching process is a chemical precipitation↔dissolution cycle driven by an imposed strain rate coupled to fluid flow. It can operate independently of any seismic activity but can also be coupled to seismicity since high fluid pressures can, in principle, nucleate sliding on fault surfaces. Such failure occurs at the extension end of the yield surface whereas high stresses at low fluid pressures at the cap end of the yield surface result in accelerated slip on suitably oriented faults (Fig. 12b). The mode-switching mechanism represents a switching between failure modes at the cap end of the yield surface (non-Andersonian structures at high angles to compression) and failure modes at the tensile end (Andersonian structures at low angles to compression).

8. Axial plane structures: quartz and melt veins

The enigmatic occurrence of veins and melt accumulations parallel to the axial planes of folds is well documented. For reviews see Vernon & Paterson (2001), Weinberg *et al.* (2015) and Druguet (2019). Some examples are given in Figure 13. Explanations for these structures vary widely and include the influences of initial fabric anisotropy and fabric heterogeneity, reorientation of older structures, relaxation or switching of stresses and preferential melting along planes normal to compression. The array of explanations reflects the difficulties inherent in concepts based on uncapped yield surfaces. Weinberg *et al.* (2015) and Veveakis *et al.* (2015) presented models for axial plane leucosomes in terms of compaction-driven instabilities based on a capped yield surface. Their model is based on Veveakis & Regenauer-Lieb (2015), as is this paper.

The important point elaborated upon by Alevizos *et al.* (2017) is that the development of both axial plane veins and melt accumulations involve mineral reactions (dissolution and deposition, melting and crystallization) that are either endothermic or exothermic. Such processes are fundamental in controlling the widths of opening-mode compaction instabilities at a high angle to compression as discussed in Section 3.

The progressive development of axial plane veins is envisaged as follows: axial plane veins are opening-mode discontinuities formed in compression under relatively low fluid (including melt) pressure – high stress conditions, perhaps as part of the mode-switching cycle. Even though the fluid pressure is low in the discontinuity it is higher than in the regions outside the discontinuity. The low pressure enables precipitation of quartz in the vein. If the fluid pressure increases in the discontinuity then dissolution may be initiated as is common in a laminated vein.

The situation is somewhat different for axial plane leucosomes: The rock is assumed to be at or just below the liquidus. The pressure in the axial plane discontinuity is higher than outside the discontinuity, so melting will take place preferentially within the discontinuity (Thompson, 1988). This relationship was suggested in a slightly different context for shear zones with high dilatancy by Ord (1990). If the melt pressure increases within the discontinuity then increased melting is favoured. The width of the leucosome is a function of the parameter η (Alevizos *et al.* 2017). The larger is η the wider the leucosome.

It is of some interest to note that processes resulting in compaction melt structures *parallel* to compression have been developed by Rabinowicz & Vigneresse (2004). These in the anatectic domain are the conceptual equivalent of crack-seal veins in the dissolution–precipitation domain. Thus, during a mode-switching cycle it is possible that switching between the development of leucosomes at high angles and low angles to compression can occur resulting in the common observation of layered leucosomes with other leucosomes at high angles to the layering. Examples are figures A2, B45, D24, D30 and E3 in Sawyer (2008).

9. Brecciation

Wong *et al.* (1997) showed experimentally that cataclasis is associated with hardening of an elliptical cap on the yield surface (Fig. 14a). The mechanics and thermodynamics of breakage have been discussed by Einav (2007*a*,*b*) and Nguyen & Einav (2009). The measure of the degree of breakage adopted by Einav (2007*a*) is the ratio of the area under the current fragment size distribution to that under the final fragment size

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

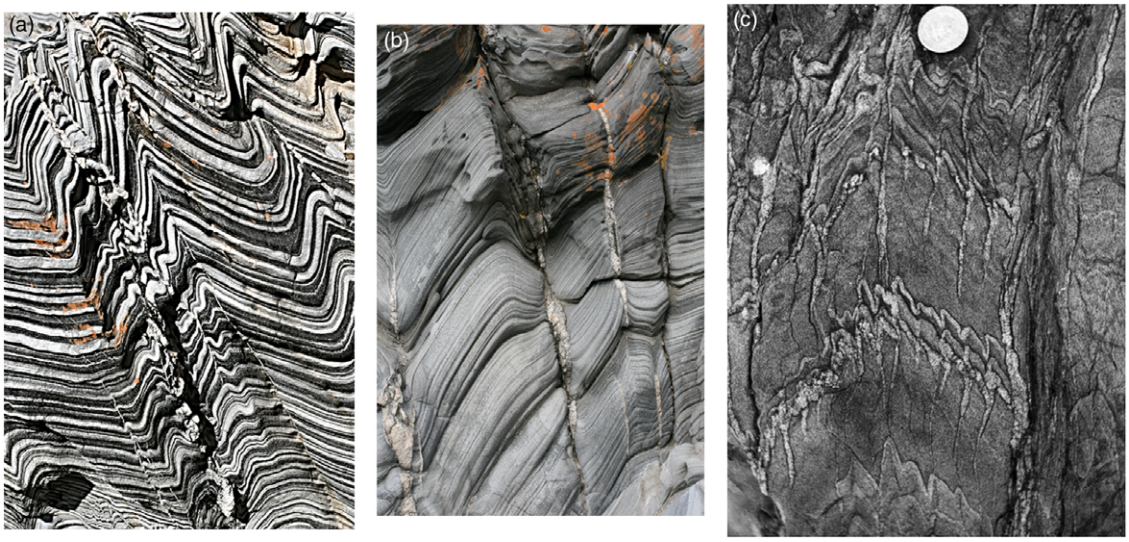


Fig. 13. Axial plane quartz and melt veins. (a, b) Axial plane veins, Harvey's Retreat, Kangaroo Island, Australia. (a) is ≈1 m wide and (b) is ≈2 m wide. (c) Axial plane melt veins. Reprinted from Vernon & Paterson (2001), Copyright (2001), with permission from Elsevier.

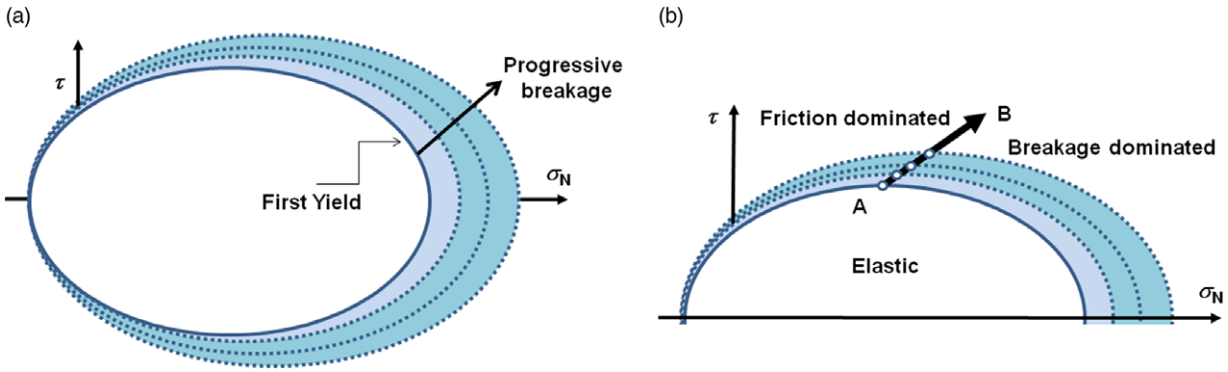


Fig. 14. Hardening of the cap arising from breakage. (a) Expansion of the cap arising from fragmentation. The evolution of the cap is controlled by the degree of breakage (Nguyen & Einav, 2009). (b) Modes of failure. The line AB marks the boundary between friction dominated and breakage dominated failure modes resulting in different brecciation modes.

distribution. Figure 14 shows the hardening of the cap associated with breakage.

On the basis of the work of Wong *et al.* (1997) and Nguyen & Einav (2009) we recognize four end-member classes of breccia (Fig. 15) that arise from the mode-switching cycle. The first corresponds to low stress and high fluid pressure where the stress circle exceeds the tensile end of the yield surface. Failure is essentially by opening-mode veining (A in Fig. 15). The second corresponds to high stress and high fluid pressure where the stress circle exceeds the tensile-shear part of the yield surface. This corresponds to considerable dilation and mineral precipitation with the formation of clast unsupported breccias (B in Fig. 15). The third corresponds to high stress and low fluid pressure where the stress circle exceeds the cap part of the yield surface. The result is a closing-mode breccia with solution seams and mineral precipitation (C in Fig. 15). The fourth corresponds to low stress and low fluid pressure where the stress circle exceeds the cap tip part of the yield surface. This corresponds to collapse fracturing and little mineral precipitation (D in Fig. 15). Each of these classes has clearly defined fabrics associated with their development (Fig. 15).

10. Discussion

This paper discusses the implications of adding a cap to the classical open-ended Mohr–Coulomb yield surface along with the added effects of coupling to fluid flow and chemical reactions that result in negative ΔV or in dissolution, fluid transport and precipitation. Behaviour at the cap is controlled by competition between fluid flow diffusivities and deformation diffusivities. For a constant strain rate, low permeability results in opening-mode discontinuities normal to compression (but only if the deformation is coupled to mineral reactions that lead to a decrease in volume in the matrix of the rock) whereas high permeability results in compaction bands. The result is nonlinear behaviour where the inverse relationship between stress and dissolution/precipitation reaction rates drives oscillatory behaviour between high fluid pressure – low differential stress states and low fluid pressure – high differential stress states. This cyclic behaviour is accompanied by oscillations between classic extension veins parallel to compression (typically crack-seal veins) and stylolite parallel veins normal to compression (typically laminated or quartz-ribbon veins). The deformation–chemical-driven aseismic cyclic behaviour, labelled

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

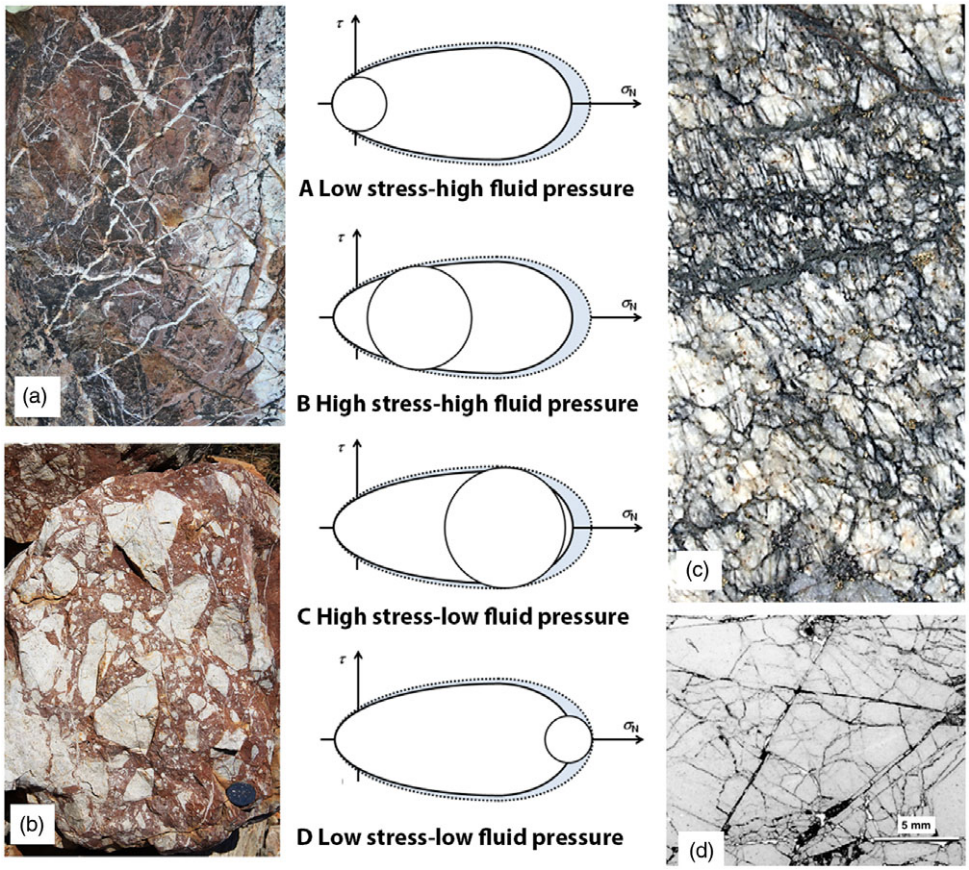


Fig. 15. The four end-member classes of breccia. (a) Low stress – high fluid pressure. Stress circle exceeds the extension end of the yield surface. Failure is essentially by veining. (b) High stress – high fluid pressure. Stress circle exceeds the extension–shear part of yield surface with considerable dilation and mineral precipitation. (c) High stress – low fluid pressure. Stress circle exceeds the cap part of yield surface with solution seams and mineral precipitation. (d) Low stress – low fluid pressure. Stress circle exceeds the cap tip part of yield surface with collapse fracturing and little mineral precipitation.

mode-switching, is an alternative to the seismically driven *fault-valve* process described by Sibson (2020). The mode-switching process does not involve mechanical breaching of a hydraulic 'seal' nor reactivation of faults at high angles to compression and need not be coupled to seismic activity. Axial plane veins (commonly parallel to axial plane solution seams) and melt segregations belong to the class of discontinuities formed by failure at the cap. Brecciation is the result of failure associated with a hardening cap.

The treatment in this paper largely neglects thermal effects such as heat produced by frictional sliding, breakage and exothermal mineral reactions (such as mineral dissolution) or heat absorbed by endothermic mineral reactions (such as mineral deposition). Quartz deposition is particularly important here. We assume that ΔH^0 for the quartz deposition reaction is approximately equal to the activation energy, E_a, determined by Rimstidt & Barnes (1980, their table 4). Thus, for the endothermic reaction

$$H_4SiO_4 \rightarrow SiO_{2(solid)} + 2H_2O$$

ΔH^0 is approximately 71 kJ mol^{-1}. This means that the deposition of 1 cubic metre of quartz requires $\approx 3.127 \times 10^6$ kJ of heat. If this deposition occurs over 10^5 years then the heat absorption rate is a massive 10^{-3} W m^{-3}, which is far in excess of the average radiogenic crustal heat production rate. Shorter time frames increase the heat production rates. The behaviour of the coupled deformation–chemical reaction system we have described in this paper depends on the competition between exothermic processes such as frictional deformation, breakage and the production of new surfaces, mineral dissolution and endothermic processes such as mineral deposition. Typically such systems are episodic in behaviour as heat and mass supply compete with heat and mass consumption (Lesueur *et al.* 2020), and a full understanding of the mode-switching model depends on incorporation of such thermal effects.

A more realistic treatment would also include the influence of anisotropy on the shape of the yield surface, and this is probably needed to explain why one only observes single orientations of failure surfaces in the field rather than ubiquitous conjugate pairs.

Although the treatment here has concentrated on brittle behaviour, Wintsch & Yi (2002) have emphasized that dissolution and replacement creep can be an important mechanism in the deep crust, and so the oscillatory switching between dissolution and deposition may be an important process even at high metamorphic grades where the Mohr–Coulomb yield surface is replaced by a capped elastic–plastic-viscous yield surface. Wintsch & Yeh (2013) have further suggested that oscillations between reaction weakening and reaction strengthening can result in switching between ductile and brittle behaviour. All of these processes have the potential to produce strong feedback relationships so that not only does the fluid pressure oscillate as the crust deforms but also the yield surface can change shape and size with mineral reactions producing oscillations in fluid pressure, dissolution, precipitation, veining and brecciation.

Finally, we note that the failure-mode diagrams of Cox (2010) are special examples of the mean stress – shear stress diagrams commonly used in the soil mechanics literature in particular (Schofield & Wroth, 1968), but also in the experimental rock deformation literature. In this case, the fluid pressure is equal to the mean stress for elastic deformations (Hobbs & Ord, 2018) so that the use of a mean stress – shear stress diagram up until yield is

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

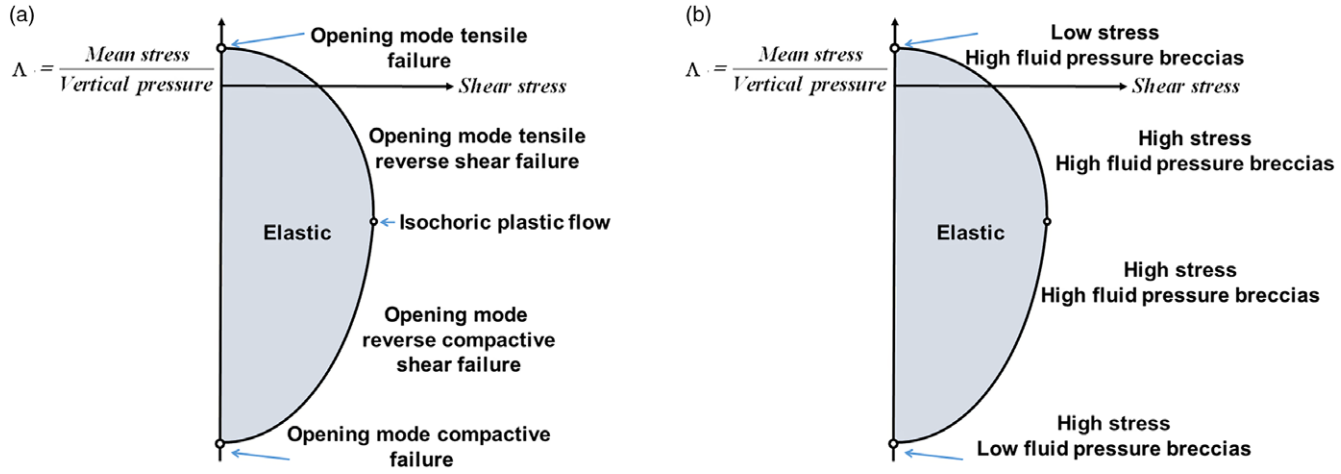


Fig. 16. Cox-failure-mode diagrams expressed as modified mean stress – shear stress diagrams. (a) Plastic failure modes. (b) Classes of breccia.

equivalent to a Cox-failure-mode diagram. The modes of failure on mean stress – shear stress diagrams are shown in Figure 16.

11. Conclusions

We have explored several implications of adding a cap to the yield surface for elastic–plastic and elastic–visco-plastic materials. Such a cap ensures that the necessity, implied by uncapped yield surfaces, for the deforming rock to support infinite shear and normal stresses is overcome. Failure at the cap results in compaction bands for highly permeable rocks but results in opening-mode discontinuities forming normal to compression in low permeability rocks if the deformation is coupled to mineral reactions that lead to a decrease in volume in the matrix of the rock. These opening-mode structures appear as laminated veins with vein parallel stylolites and solution seams and contrast with opening-mode veins that form normal to stylolites at the extension end of the yield surface. The addition of a cap can severely limit the permissible stress states that can be represented on a Cox-failure-mode diagram. Failure at the cap in fluid saturated low permeability rocks also results in opening-shear structures at high angles to compression. These are commonly interpreted as reactivated older normal faults, but the addition of a cap makes such an interpretation unnecessary.

A capped yield surface means that a deforming fluid saturated rock mass automatically oscillates between a low stress – high fluid pressure, low permeability state and a high stress – low fluid pressure, high permeability state controlled by competition between dissolution and precipitation. This process, labelled *mode-switching*, is a coupled mineral reaction – deformation – fluid flow cyclic mechanism that is an alternative to the *fault-valve* process. The mode-switching process involves switching between reverse slip on discontinuities at a high angle to compression and the operation of opening-mode discontinuities at a low angle to compression. Mode-switching does not depend on fluids breaching a permeability 'seal' and is intrinsically aseismic, although it can, in principle, nucleate a seismic cycle process.

The addition of a cap to the yield surface also provides a basis for the development of axial plane veins and melt segregations together with various classes of breccias. The capped yield surface concept provides a unifying self-consistent approach for vein/breccia formation, cyclic yielding and fluid flow and for the kinematics of brittle rocks undergoing solution creep.

Acknowledgements. We thank John Rudnicki and two anonymous reviewers for their helpful and stimulating observations and comments. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. We have no competing interests.

References

Alevizos S, Poulet T, Sari M, Lesueur M, Regenauer-Lieb K and Veveakis M (2017) A framework for fracture network formation in overpressurized impermeable shale: deformability vs diagenesis. *Rock Mechanics and Rock Engineering* **50**, 689–703.

Anderson EM (1905) The dynamics of faulting. *Transactions of the Edinburgh Geological Society* **8**, 340–87.

Anderson EM (1951) *The Dynamics of Faulting*. Edinburgh: Oliver and Boyd.

Aydin A, Borja RI and Eichhubl P (2006) Geological and mathematical framework for failure modes in granular rock. *Journal of Structural Geology* **28**, 83–98.

Bésuelle P and Rudnicki JW (2004) Localization: shear bands and compaction bands. In *Mechanics of Fluid-Saturated Rock* (eds Y Guéguen and M Boutéca), pp. 219–321. International Geophysics vol. 89. London: Academic Press.

Bons PD, Elburg MA and Gomez-Rivas E (2012) A review of the formation of tectonic veins and their microstructures. *Journal of Structural Geology* **43**, 33–62.

Borja RI and Aydin A (2004) Computational modeling of deformation bands in granular media: I. Geological and mathematical framework. *Computer Methods in Applied Mechanics and Engineering* **193**, 2667–98.

Cahn JW (1989) The physical chemistry of stressed solids. *Berichte der Bunsengesellschaft für physikalische Chemie* **93**, 1169–73.

Chace FM (1949) Origin of the Bendigo saddle reefs with comments on the formation of ribbon quartz. *Economic Geology* **44**, 561–97.

Cheong S, Peters SG and Iriondo A (2000) Summary of structural setting and fluid characteristics of low-sulfide Au-quartz veins in northwestern Nevada. In *Geology and Ore Deposits 2000: The Great Basin and Beyond: Geological Society of Nevada Symposium Proceedings* (eds JK Cluer, JG Price, EM Struhsacker, RF Hardyman and CL Morris), pp. 473–506.

Cier RJ, Labanda NA and Calo VM (In press) Compaction band localisation in geomaterials: a mechanically consistent failure criterion. *Journal of Numerical and Analytical Methods in Geomechanics*.

Cowan EJ (2020) Deposit-scale structural architecture of the Sigma-Lamaque gold deposit, Canada—insights from a newly proposed 3D method for assessing structural controls from drill hole data. *Mineralium Deposita* **55**, 217–40.

Cox SF (2010) The application of failure mode diagrams for exploring the roles of fluid pressure and stress states in controlling styles of fracture-controlled permeability enhancement in faults and shear zones. *Geofluids* **10**, 217–33.

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

Dimaggio FL and Sandler IS (1971) Material model for granular soils. *Journal of the Engineering Mechanics Division, American Society of Civil Engineers* **97**, 935–50.

Dove PM and Rimstidt JD (1994) Silica-water interactions. *Reviews in Mineralogy and Geochemistry* **29**, 259–308.

Druguet E (2019) Deciphering the presence of axial-planar veins in tectonites. *Geoscience Frontiers* **10**, 2101–15.

Eichhubl P, Hooker JN and Laubach SE (2010) Pure and shear-enhanced compaction bands in Aztec Sandstone. *Journal of Structural Geology* **32**, 1873–86.

Einav I (2007a) Breakage mechanics – Part I: Theory. *Journal of the Mechanics and Physics of Solids* **55**, 1274–97.

Einav I (2007b) Breakage mechanics – Part II: Modelling granular materials. *Journal of the Mechanics and Physics of Solids* **55**, 1298–320.

Ferguson HG and Gannett RW (1932) Gold Quartz Veins of the Alleghany District, California. United States Geological Survey Professional Paper 172, 232 pp.

Fossen H, Schultz RA, Shipton ZK and Mai K (2007) Deformation bands in sandstone: a review. *Journal of the Geological Society, London* **164**, 1–15.

Fossum AF and Fredrich JT (2000) Cap plasticity models and compactive and dilatant pre- failure deformation. In *Pacific Rocks 2000, Proceedings of the 4th North American Rock Mechanics Symposium* (eds J Girard, M Liebman, C Breeds and T Doe), pp. 1169–76. Brookfield, Vermont: A. A. Balkema.

Fossum AF, Senseny PE, Pfeifle TF and Mellegard KD (1995) Experimental determination of probability distributions for parameters of a Salem limestone cap plasticity model. *Mechanics of Materials* **21**, 119–37.

Fournier RA and Potter RW III (1982) An equation correlating the solubility of quartz in water from 25°C to 900°C at pressures up to 10,000 bars. *Geochimica et Cosmochimica Acta* **46**, 1969–73.

Frolov T and Mishin Y (2010) Effect of nonhydrostatic stresses on fluid-solid equilibrium: I. Bulk thermodynamics. *Physical Review B: Condensed Matter and Materials Physics* **82**, 174113. doi: 10.1103/PhysRevB.82.174113.

Gibbs JW (1876) On the equilibrium of heterogeneous substances. *Transactions of the Connecticut Academy* **3**, 108–248.

Gratier J-P, Dysthe DK and Renard F (2013) The role of pressure solution creep in the ductility of the Earth's upper crust. In *Advances in Geophysics, Vol. 54* (ed. R Dmowska), pp. 47–179. Amsterdam: Elsevier.

Haack UK and Zimmermann HD (1996) Retrograde mineral reactions: a heat source in the continental crust? *Geologische Rundschau* **85**, 130–37.

Hobbs BE and Ord A (2015) *Structural Geology: The Mechanics of Deforming Metamorphic Rocks*. Amsterdam: Elsevier.

Hobbs BE and Ord A (2018) Coupling of fluid flow to permeability development in mid- to upper crustal environments: a tale of three pressures. In *Characterization of Ore-Forming Systems from Geological, Geochemical and Geophysical Studies* (eds K Gessner, TG Blenkinsop and P Sorjonen-Ward), pp. 81–120. Geological Society of London, Special Publication no. 453.

Hough MA, Bierlein FP, Ailleres L and McKnight S (2010) Nature of gold mineralisation in the Walhalla Goldfield, southeast Australia. *Australian Journal of Earth Sciences* **57**, 969–92.

Houlsby GT and Puzrin AM (2006) *Principles of Hyperplasticity: An Approach to Plasticity Theory Based on Thermodynamic Principles*. London: Springer-Verlag.

Issen K and Rudnicki J (2000) Conditions for compaction bands in porous rock. *Journal of Geophysical Research* **105**, 21529–36.

Itasca Consulting Group, Inc (2008) *FLAC — Fast Lagrangian Analysis of Continua, Ver. 6.0*. Minneapolis: Itasca.

Kamb WB (1961) Thermodynamic theory of nonhydrostatically stressed solids. *Journal of Geophysical Research* **66**, 259–71.

Lesueur M, Poulet T and Veveakis M (2020) Three-scale multiphysics finite element framework (FE3) modelling fault reactivation. *Computer Methods in Applied Mechanics and Engineering* **365**, 112988. doi: 10.1016/j.cma.2020.112988.

Manning CE (1994) The solubility of quartz in H_2O in the lower crust and upper mantle. *Geochimica et Cosmochimica Acta* **58**, 4831–9.

Nguyen GD and Einav I (2009) The energetics of cataclasis based on breakage mechanics. *Pure and Applied Geophysics* **166**, 1693–724.

Olsson WA (1999) Theoretical and experimental investigation of compaction bands in porous rock. *Journal of Geophysical Research: Solid Earth* **104**, 7219–28.

Ord A (1990) Mechanical controls on dilatant shear zones. In *Deformation Mechanisms, Rheology and Tectonics* (eds RJ Knipe and EH Rutter), pp. 183–92. Geological Society of London, Special Publication no. 54.

Paterson MS (1995) A theory for granular flow accommodated by material transfer via an intergranular fluid. *Tectonophysics* **245**, 135–51.

Phillips OM (1991) *Flow and Reactions in Permeable Rocks*. Cambridge: Cambridge University Press.

Pollard DD and Aydin A (1988) Progress in understanding jointing over the past century. *Geological Society of America Bulletin* **100**, 1181–204.

Pollard DD and Segall P (1987) Theoretical displacements and stresses near fractures in rock: with applications to faults, joints, dikes, and solution surfaces. In *Fracture Mechanics of Rock* (ed. BK Atkinson), pp. 277–349. New York: Academic Press.

Price NJ and Cosgrove JW (1990) *Analysis of Geological Structures*. Cambridge: Cambridge University Press.

Rabinowicz M and Vigneresse J-L (2004) Melt segregation under compaction and shear channeling: application to granitic magma segregation in a continental crust. *Journal of Geophysical Research: Solid Earth* **109**, B04407. doi: 10.1029/2002JB002372.

Regenauer-Lieb K, Poulet T and Veveakis M (2016) A novel wave-mechanics approach for fluid flow in unconventional resources. *The Leading Edge* **35**, 90–7.

Reiweger I, Gaume J and Schweizer J (2015) A new mixed-mode failure criterion for weak snowpack layers. *Geophysical Research Letters* **42**, 1427–32.

Rimstidt JD and Barnes HL (1980) The kinetics of silica-water reactions. *Geochimica et Cosmochimica Acta* **44**, 1683–99.

Rudnicki JW (2000) Diffusive instabilities in dilating and compacting geomaterials. In *Multiscale Fracture and Deformation in Materials and Structures – The James R. Rice 60th Anniversary Volume* (eds T-J Chuang and JW Rudnicki), pp. 159–82. Dordrecht: Kluwer Academic Publishers.

Rudnicki JW (2004) Shear and compaction band formation on an elliptic yield cap. *Journal of Geophysical Research: Solid Earth* **109**, B03402. doi: 10.1029/2003JB002633.

Rutter EH (1976) The kinetics of rock deformation by pressure solution. *Philosophical Transactions of the Royal Society of London* **283**, 203–19.

Sawyer EW (2008) *Atlas of Migmatites*. The Canadian Mineralogist, Special Publication no. 9. Ottawa: NRC Research Press.

Schofield AN and Wroth P (1968) *Critical State Soil Mechanics*. New York: McGraw-Hill.

Scholz CH (1989) Mechanics of faulting. *Annual Review of Earth and Planetary Sciences* **17**, 309–34.

Schultz RA and Fossen H (2008) Terminology for structural discontinuities. *American Association of Petroleum Geologists Bulletin* **92**, 853–67.

Schwalm WA (2015) *Lectures on Selected Topics in Mathematical Physics: Elliptic Functions and Elliptic Integrals*. San Rafael, California: Morgan & Claypool Publishers.

Sekerka RF and Cahn JW (2004) Solid-liquid equilibrium for non-hydrostatic stress. *Acta Materialia* **52**, 1663–8.

Shimizu I (1995) Kinetics of pressure solution creep in quartz: theoretical considerations. *Tectonophysics* **245**, 121–34.

Shimizu I (1997) The nonequilibrium thermodynamics of intracrystalline diffusion under nonhydrostatic stress. *Philosophical Magazine A* **75**, 1221–35.

Sibson RH (1981) Fluid flow accompanying faulting: Field evidence and models. In *Earthquake Prediction: An International Review, Volume 4* (eds DW Simpson and PG Richards), pp. 593–603. Washington, DC: American Geophysical Union.

Sibson RH (1985) A note on fault reactivation. *Journal of Structural Geology* **7**, 751–4.

Sibson RH (2020) Preparation zones for large crustal earthquakes consequent on fault-valve action. *Earth, Planets and Space* **72**, 31. doi: 10.1186/s40623-020-01153-x.

Sibson RH, Robert F and Poulsen KH (1988) High-angle reverse faults, fluid pressure cycling, and mesothermal gold-quartz deposits. *Geology* **16**, 551–5.

Stefanou I and Sulem J (2014) Chemically induced compaction bands: triggering conditions and band thickness. *Journal of Geophysical Research: Solid Earth* **119**, 880–99.

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

Thompson AB (1988) Dehydration melting of crustal rocks. *Rendiconti della Societa Italiana di Mineralogia e Petrologia* **43**, 41–60.

Toussaint R, Aharonov E, Koehn D, Gratier P, Ebner M, Baud P, Rolland A and Renard F (2018) Stylolites: a review. *Journal of Structural Geology* **114**, 163–95.

Vardoulakis I and Sulem J (1995) *Bifurcation Analysis in Geomechanics*. London: Blackie Academic and Professional.

Vernon RH and Paterson SR (2001) Axial-surface leucosomes in anatectic migmatites. *Tectonophysics* **335**, 183–92.

Veveakis M and Regenauer-Lieb K (2015) Cnoidal waves in solids. *Journal of the Mechanics and Physics of Solids* **78**, 231–48.

Veveakis E, Regenauer-Lieb K and Weinberg RF (2015) Ductile compaction of partially molten rocks: the effect of non-linear viscous rheology on instability and segregation. *Geophysical Journal International* **200**, 519–23.

Weinberg RF, Veveakis E and Regenauer-Lieb K (2015) Compaction-driven melt segregation in migmatites. *Geology* **43**, 471–4.

Wintsch RP and Yeh M-W (2013) Oscillating brittle and viscous behavior through the earthquake cycle in the Red River Shear Zone: monitoring flips between reaction and textural softening and hardening. *Tectonophysics* **587**, 46–62.

Wintsch RP and Yi K (2002) Dissolution and replacement creep: a significant deformation mechanism in mid-crustal rocks. *Journal Structural Geology* **24**, 1179–93.

Wong T-F, David C and Zhu W (1997) The transition from brittle faulting to cataclastic flow in porous sandstones: mechanical deformation. *Journal of Geophysical Research: Solid Earth* **102**, 3009–25.

Zhao C, Hobbs BE and Ord A (2008) *Convective and Advective Heat Transfer in Geological Systems*. Berlin: Springer-Verlag.

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

Appendix: Equation for an elliptical cap in a Mohr–Coulomb material

We follow Reiweger *et al.* (2015) who developed a Mohr–Coulomb yield surface with a cap for compacted snow. The equation for the cap is written:

$$\tau_{cap} = b\sqrt{1 - \frac{(\sigma+\sigma_t)^2}{(\sigma_c+\sigma_t)^2}} \text{ where } b = K\sqrt{\frac{(\sigma_c+\sigma_t)^2}{(\sigma_c+\sigma_t)^2 - \left(\frac{K}{\tan\phi}\right)^2}}$$

This corresponds to an ellipse with centre at $(-\sigma_t, 0)$, major axis, $|\sigma_c + \sigma_t|$and minor axis, b, as shown in Figure 6a. The maximum normal stress the material can sustain without collapsing is $\sigma = \sigma_c$ and the maximum shear stress the material can sustain without yielding is $\tau_{cap} = K$. Both σ_c and K are material parameters that are independent of each other, except that since σ_t is generally small, $\sigma_c > K$; having set their values, along with σ_t, b is defined. In addition, $(\sigma_c + \sigma_t) > \frac{K}{\tan\phi}$ must always be true for b to be real.

We write the following expressions involving only material parameters:

$$a = (\sigma_c+\sigma_t)^2, \quad c = \left(\frac{K}{\tan\phi}\right)^2$$

Then

$$\tau_{cap} = b\sqrt{1 - \frac{(\sigma+\sigma_t)^2}{a}} \text{ and } b = K\sqrt{\frac{a}{a-c}}$$

The equation for the cap becomes:

$$\tau_{cap} = b\sqrt{1 - \frac{(\sigma+\sigma_t)^2}{a}} = K\sqrt{\frac{a-(\sigma+\sigma_t)^2}{a-c}}$$

We now put $\sigma' = \sigma + P_f$ where σ' is the effective normal stress at yield and P_f is the fluid pressure at yield in the cap region. Notice that one *adds* the fluid pressure rather than *subtracting* as in the classical situation with no cap since here the fluid pressure is *decreased* rather than *increased* to enable yield.

Now the equation for the cap at yield is:

$$\tau_{cap} = K\sqrt{\frac{a-(\sigma' - P_f + \sigma_t)^2}{a-c}}$$

Resulting in: $\tau_{cap} = K\sqrt{\frac{a-(\sigma'-P_f+\sigma_t)^2}{a-c}}$

or, $P_f = (\sigma' + \sigma_t) - \sqrt{a - (a-c)\tau_{cap}^2/K^2}$

We now write the standard relationships (from here on the principal stresses are effective stresses):

$$\sigma' = \frac{\sigma_1+\sigma_3}{2} + \frac{\sigma_1-\sigma_3}{2}\cos 2\theta = \frac{\sigma_1-\sigma_3+2\sigma_3}{2} + \frac{\sigma_1-\sigma_3}{2}\cos 2\theta$$

and $\tau_{cap} = -\frac{\sigma_1-\sigma_3}{2}\sin 2\theta$

or, $\sigma' = \frac{1}{2}(\sigma_1-\sigma_3)(1+\cos 2\theta) + \sigma_3$ and $\tau_{cap} = -\frac{\sigma_1-\sigma_3}{2}\sin 2\theta$

Then the expression for P_f in terms of $(\sigma_1 - \sigma_3)$ becomes

$$P_f = \left[\frac{1}{2}(\sigma_1-\sigma_3)(1+\cos 2\theta) + \sigma_3 + \sigma_t\right] - \sqrt{a - (a-c)\frac{(\sigma_1-\sigma_3)^2}{4K^2}\sin^2 2\theta}$$

The equation for P_f versus $(\sigma_1 - \sigma_3)$ at yield is therefore a parabola. This means the complete failure-mode diagram consisting of the classical Cox function plus a cap is as in Figure 6b.

We now take the values for the various constitutive parameters as in Cox (2010), namely,

$\theta = \frac{1}{2}\tan^{-1}\left(\frac{1}{\mu}\right)$; $\mu = 0.75$; $2\theta = 53.13^\circ$; $\sin 2\theta = 0.8$; $\cos 2\theta = 0.6$;

$\phi = \tan^{-1}\mu = 36.87°$; $\sigma_t = 5$ MPa; $C = 3.75$ MPa; depth, $z = 10$ km so that $\sigma_3 = \rho g z = 265$ MPa.

Then $a = (\sigma_c+\sigma_t)^2 = (\sigma_c+5)^2$; $c = \left(\frac{K}{\tan\phi}\right)^2 = (1.3333K)^2$ so that

$$b = K\sqrt{\frac{a}{a-c}} = K\sqrt{\frac{(\sigma_c+\sigma_t)^2}{(\sigma_c+\sigma_t)^2 - \left(\frac{K}{\tan\phi}\right)^2}} = K\sqrt{\frac{(\sigma_c+5)^2}{(\sigma_c+5)^2 - (1.3333K)^2}}$$

$$P_f = \left[\frac{1}{2}(\sigma_1-\sigma_3)(1+\cos 2\theta) + \sigma_3 + \sigma_t\right] - \sqrt{a - (a-c)\frac{(\sigma_1-\sigma_3)^2}{4K^2}\sin^2 2\theta}$$

$$P_f = [0.8(\sigma_1-\sigma_3) + 270)] - \sqrt{(\sigma_c+5)^2 - 0.16((\sigma_c+5)^2 - (1.3333K)^2)(\sigma_1-\sigma_3)^2/K^2}$$

and

$$\Lambda = \frac{P_f}{\sigma_3} = 0.0038\{[0.8(\sigma_1-\sigma_3) + 270] - \sqrt{(\sigma_c+5)^2 - 0.16((\sigma_c+5)^2 + (1.3333K)^2)(\sigma_1-\sigma_3)^2/K^2}\}$$

Let us choose $\sigma_c = 500$ MPa, and $K = 250$ MPa then

$$b = K\sqrt{\frac{(\sigma_c+5)^2}{(\sigma_c+5)^2 - (1.3333K)^2}} = 250\sqrt{\frac{(505)^2}{(505)^2 - (1.3333\times 250)^2}}$$

or, $b = 332.8$ MPa

The Mohr–Coulomb failure envelope plus cap is shown in Figure A1. The initial stress state shown by the full blue circle can lead to failure by *increasing* the fluid pressure by 210 MPa

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press

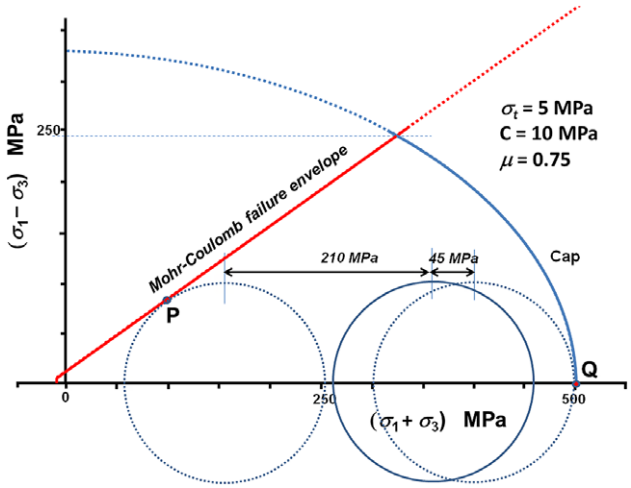


Fig. A1. Failure by increasing and decreasing fluid pressure for a yield surface with a cap using parameters proposed by Cox.

to fail by dilatant shear at P or by *decreasing* the fluid pressure by 45 MPa to fail by compactional failure normal to σ_1 at Q.

If we take the above expression for Λ, namely,

$$\Lambda = 0.0038\{[0.8(\sigma_1 - \sigma_3) + 270]$$

$$-\sqrt{(\sigma_c + 5)^2 - 0.16((\sigma_c + 5)^2 + (1.3333K)^2)(\sigma_1 - \sigma_3)^2/K^2}\}$$

and take $\sigma_c = 100$ MPa and $K = 50$ MPa, then

$$\Lambda = 0.0038\{[0.8(\sigma_1 - \sigma_3) + 270)]$$

$$-\sqrt{(255025 - 111105.56)(\sigma_1 - \sigma_3)^2/250^2}\}$$

Other combinations of σ_c and K are plotted in Figure 7a.

https://doi.org/10.1017/S0016756822000516 Published online by Cambridge University Press