1. Introduction

Fault zones control a wide range of crustal processes. Although fault zones occupy only a small volume of the crust, they have a controlling influence on the crust's mechanical and fluid flow properties. Much recent work has concentrated on describing and understanding the importance of the structure, mechanics and fluid flow properties of fault zones. This has involved field observations, laboratory experiments, seismology, hydrogeology, and analytical and numerical modelling.

Brittle fault zones are lithologically heterogeneous, anisotropic and discontinuous. Faults are complex zones composed of linked fault segments, one or more high strain slip surfaces nested within regions of high and low strain (often called fault core and damage zone), Riedel shears, splay faults, dilational and contractual jogs, and relay ramps (Rawling et al., 2001; Shipton and Cowie, 2001; Faulkner et al., 2003; Childs et al., 2009). Individual fault zones commonly show significant variation in complexity along strike or down dip, even over relatively short distances (Schulz and Evans, 2000; Shipton and Cowie, 2001; Kirkpatrick et al., 2008; Lunn et al., 2008). Fault zone structure, mechanics and permeability can vary strongly both over geological time (e.g. Eichhubl et al., 2009) and at timescales relevant to a variety of industrial applications.

The strength of the lithosphere varies with depth, temperature and mineralogy (Kohlstedt et al., 1995) but a major load-bearing region is likely present at the base of the seismogenic zone at ~15 to 20 km depth for normal continental crust. The mechanical properties of faults at this depth are thus inferred to control to some extent the strength of the entire crust. This inference is supported by observations of crustal stress magnitudes at shallower crustal levels that appear to be limited by the typical frictional strength of faults (Townend and Zoback, 2000). A related goal to characterizing the mechanical properties of faults at depth is to understand the earthquake process from nucleation and propagation to arrest.

Faults play an important role in controlling the migration of crustal fluids. One example of this is hydrocarbon migration, accumulation and leakage in sedimentary basins. At the basin scale,
faults and fault-related folds control subsidence patterns and hence the distribution of thermally mature zones (e.g. Brister et al., 2002). Faults are also a key component of many hydrocarbon plays; they also may control discrete subsurface pressure cells (e.g. Grauls et al., 2002). At a smaller scale, a better understanding of the role of faults in compartmentalizing fields will yield better estimates of hydrocarbon production (e.g. Manzocchi et al., 1999). Increasingly, the recognition of high transient permeability along faults induced by hydrocarbon production (Losh and Haney, 2006) has focussed interest on the temporal variation of fault-zone permeability. Ore deposits are also commonly related to fault zones due to episodic, localized hydrothermal flows that occur during and immediately after periods of fault movement (e.g. Cox et al., 2001; Sibson, 2001).

Characterizing the fluid flow properties of the crust is necessary to facilitate the development of deep-waste storage repositories (e.g. Ferrill et al., 1999; Douglas et al., 2000), to allow sequestration of industrially-produced greenhouse gases (Streit and Hills, 2004; Dockrill and Shipton, 2010) and to realize the potential of geothermal energy in appropriate locations (Rowland and Sibson, 2004; Fairley, 2009). The physical characteristics and properties of faults will play an important role in regional crustal fluid flow that might affect such applications.

This review concentrates first on the structure of fault zones and fault systems, including scaling relationships that can help in deciphering how fault zones develop and grow. Second, we discuss the mechanics of fault zones; this involves topics ranging from laboratory experimentation to modelling of earthquake rupture. Last, we address the topic of fluid-flow properties of faults, involving studies from in situ observations of fluid flow, laboratory experiments and modelling. We conclude by highlighting some key outstanding questions.

2. Structure and development of fault zones and fault systems

2.1. Typical fault zone structure

A simple conceptual model for fault zone structure, developed over the past 20 years, involves strain that is localized in a fault core surrounded by a distributed zone of fractures and faulting in the damage zone (Fig. 2, see references in Wibberley et al., 2008). The fault core generally consists of gouge, cataclasite or ultracataclasite (or a combination of these), and the damage zone generally consists of fractures over a wide range of length scales and subsidiary faults. Strain may be homogeneously distributed across the fault core (Rutter et al., 1986; Faulkner et al., 2003) or may be highly localized onto discrete slip surfaces (Chester and Logan, 1986; De Paola et al., 2008). Brittle fault rocks have previously been considered chaotic, but detailed observations show that they are highly structured, commonly containing P foliations, Riedel shears and Y shears (Logan et al., 1979; Chester et al., 1985; Rutter et al., 1986; Jefferies et al., 2006b). Additionally, fault breccias, such as those described by Caine et al. (2010) can occur as part of the fault core in areas where movement involving fault jogs results in dilatation.

Recent research has questioned the general applicability of this simple model. Fault zones may contain a single fault core (sometimes with branching subsidiary faults), or the fault core may branch, anastomose and link, entraining blocks or lenses of fractured protolith between the layers (Fig. 2b; Faulkner et al., 2003). A comparison between the Punchbowl fault (Chester et al., 1993) and the Carboneras fault (Faulkner et al., 2003), two 40 km-displacement strike-slip faults that operated at similar depths, illustrates the effect of the contrast in mechanical properties between fault zone and protolith. The widths of the core zones of these faults (50 cm for the Punchbowl fault in granite and low-porosity sandstones versus multiple strands over a 1 km wide zone for the Carboneras fault in predominantly mica schists) may reflect differences in the mechanical strength contrast between protolith and the deformation zone. The second style of fault zone structure has been observed in the SAFOD borehole, where multiple fault strands, individually up to several metres thick, exist and at least two of these strands of fault gouge are moving simultaneously (Zoback et al., 2010). The schematic diagrams in Fig. 2 purposely have no scale to illustrate that these fault zone structures occur at a variety of scales. The internal structure of fault zones might provide a useful indication of their mechanical properties (Faulkner and Rutter, 2003; Biegel and Sammis, 2004; Faulkner et al., 2008).

Fig. 1. Flow diagram showing inter-relationships among the three main topics of structure, mechanics and fluid flow. Mode of failure refers to whether or not seismic slip occurs.
Localization may be further enhanced where a fault juxtaposes two protoliths of highly contrasting competence, such as the Median Tectonic Line, Japan (e.g. Wibberley and Shimamoto, 2003).

Fault zone structure depends on the depth of formation (e.g. Ishi et al., in press), protolith, tectonic environment (e.g. strike-slip, extension or compressional), magnitude of displacement and fluid flow. For instance, faults in low porosity rocks (Balsamo et al., 2010) generally have a fine-grained fault core surrounded by a fracture-dominated damage zone. In contrast, coarser grained, high porosity rocks commonly develop by the formation and amalgamation of low porosity deformation bands followed by the nucleation and propagation of high permeability slip surfaces (Fossen et al., 2007).

Several authors provide qualitative descriptions of fracture damage zones surrounding a fault core (e.g. McGrath and Davison, 1995; Shipton and Cowie, 2003; Kim et al., 2004; Berg and Skar, 2005; Cembrano et al., 2005; Johansen et al., 2005; Cook et al., 2006; de Joussineau and Aydin, 2007; Fossen et al., 2007). Damage zones contain fractures at a range of different scales from grain-scale microfractures to macrofractures that may accommodate small shear offsets and a small quantity of cataclastite. It can be difficult to distinguish between subsidiary fault structures (which may be viewed as faults in their own right) and fault damage zone fractures. In the tip zones of large displacement faults in particular the complexity of deformation is marked (Kirkpatrick et al., 2008). The orientation of the macroscopic deformation surrounding fault tips may include horsetail geometries, wing cracks and synthetic or antithetic subsidiary faults (e.g. de Joussineau and Aydin, 2007; Moir et al., 2010). In low porosity rocks or under low effective stress conditions, damage zones consist of dilatant fractures (Blenkinsop, 2008), whereas higher porosity rocks often develop structures in their damage zones such as compaction bands in sandstone (analogous to anticracks) or cataclastic deformation bands (e.g. Johansen et al., 2005; Fossen et al., 2007).

Quantitative studies of damage zones, commonly involve determining the density of fractures (usually from line counting) as a function of distance from the fault core. For low porosity rocks, macrofractures (mesoscale features that may be readily identified in the field) and microfractures (measured from orientated thin-sections) commonly show an exponential decrease with distance from the fault core (Vermilye and Scholz, 1998; Wilson et al., 2003; Mitchell and Faulkner, 2009) (Fig. 3). This relationship has been linked to the decay of stress away from a fault tip predicted from fracture mechanics models. Microfractures in deformation band-dominated damage zones in high porosity sandstones show no observable change of microfracture density surrounding faults (Anders and Wiltschko, 1994; Shipton and Cowie, 2001) due to the different micromechanics of deformation-band faulting and

Fig. 2. Typical fault zone structures. (a) Shows a single high-strain core surrounded by a fractured damage zone (after Chester and Logan, 1986) and (b) shows multiple cores model, where many strands of high-strain material enclose lenses of fractured protolith (after Faulkner et al., 2003).

Fig. 3. An example of (a) microfracturing and (b) macrofracturing surrounding three strike-slip fault zones in northern Chile. The fault zones are in low porosity crystalline rocks and the Caleta Coloso fault has ~5 km of displacement, the Cristales fault has 220 m of displacement and the Blanca fault 35 m of displacement. From Mitchell and Faulkner (2009).
fracture-dominated faulting (see Fig. 4). Nevertheless, where damage zones are dominated by deformation bands, these may show an exponentially decreasing density from faults in high-porosity sandstone (Berg and Skar, 2005; de Joussineau and Aydin, 2007), although in at least some cases, fault localization occurs once the damage zone has already developed (Saillet and Wibberley, 2010).

A maximum microfracture density is often attained immediately adjacent to the fault core that is dependent on rock type but independent of the fault displacement (Fig. 3). Although only three faults are shown in Fig. 3, three additional faults with smaller displacements (of 13 cm to 2 m) showed a similar maximum fracture density (of ~20 per mm) (Mitchell and Faulkner, 2009). This is suggestive of a critical amount of fracture damage that the rock can accumulate at the boundary between the fault core and the damage zone. The same type of fracture damage distribution occurs at the small scale (i.e. mm) in experiments (Janssen et al., 2001). A recently recognized type of fracture damage surrounding the surface trace of large active fault zones involves the in situ fragmentation of the protolith around the fault core, such as inferred from observations at several localities of the same fault (Reches and Dewers, 2005; Dor et al., 2006). The original structure and textures of the protolith are preserved, but pervasive microfracturing of the protolith has lead to complete incohesion. They have been termed ‘pulverized’ rocks.

2.2. Fault zone scaling and development

As summarized in the previous sub-section, fault zone structure depends on a variety of factors. Hence caution must be used in the scaling relations among fault zone characteristics such as fault thickness or damage zone width. This problem can be particularly acute when datasets from different areas are compared. The way in which different authors have defined damage zone width (mean width at one site, single measurement along a scanline or maximum damage zone width envelope) makes comparing studies very difficult (Shipton et al., 2006). Another issue is that the data may be biased by differences in the methods used to measure the fault “width” (e.g. should the “width” of the San Andreas fault zone as measured from a geological map or satellite photo include all the low-strain rock in between the various strands of the fault system?) (Schulz and Evans, 2000). Where faults with a range of displacements or lengths formed in the same protolith, at approximately the same depth and environmental conditions and under a similar stress field, then useful scaling relationships may be determined (e.g. Childs et al., 2009; Mitchell and Faulkner, 2009).

Faults initiate from the coalescence of a number of growing mode I cracks (Reches and Lockner, 1994; Healy et al., 2006). Commonly, the initiation of faults occurs on pre-existing planes of weaknesses such as cooling joints (Martel et al., 1988; Di Toro and Pennacchioni, 2005), dykes (d’Alessio and Martel, 2005) or tectonic joints (Wilkins et al., 2001; de Joussineau and Aydin, 2007; Moir et al., 2010).

As fault cores develop, wear models have suggested that progressive damage leads to the development of a wider fault core. Broad trends may be seen in fault core thickness — displacement relationships (Scholz, 1987), but the wide scatter in these data may reflect the fact that these compilations are from faults developed in wide range of tectonic regimes and which formed in variable environmental conditions. Although wear models are physically appealing from a mechanical perspective and make intuitive sense, the complexities arising in natural fault zones produce a wide range of resultant fault core internal structures (Evans, 1990; Shipton et al., 2006; Faulkner et al., 2008). In summary, quantitative models for fault core development may not be universally applicable and may, more accurately, provide an upper bound on fault core width.

The growth of fault damage zones and their scaling with fault displacement is an important topic, as the fracture damage zones surrounding faults may provide fluid flow pathways of economic significance and, in addition, act as an energy sink during quasi-static fault growth or dynamic rupture. Vermilye and Scholz (1998) made measurements of small displacement faults and concluded that fault damage zone width (defined by the points where fracture damage falls below background levels on either side of the fault) scales with fault displacement. Mitchell and Faulkner (2009) studied damage zone width defined by macro- and micro-fracture densities for faults developed within the same granodioritic batholith over nearly five orders of magnitude of displacement (Fig. 5). The damage zone width scales with fault displacement, however at higher displacements the rate of growth of the damage zone width with displacement decreases after a few hundred

![Fig. 4. Comparison of microfracture density with distance from the main fault from high porosity rocks with deformation bands (Anders and Wiltshire, 1994; Shipton and Cowie, 2001) with microfracture density from low porosity rocks (Anders and Wiltshire, 1994; Vermilye and Scholz, 1998). The average host rock porosity for each site is given in brackets. Linear regressions show that in low porosity rocks a logarithmic decay of microfracture density occurs with distance from the fault zone. Conversely, in high porosity rocks, microfracture densities drop to background levels at all points outside the fault zone.](image-url)
metres of displacement (Fig. 5). Micarelli et al. (2006) found a similar decrease in the growth of fracture damage zone width at displacements above 1–5 m in high-porosity carbonate sequences in Sicily. However, in this case, displacements of 1–5 m coincide with the onset of cataclastic fault rock generation, which may be controlled by the scale of the bedding in these rocks which is of the same order of magnitude. Savage and Brodsky (submitted for publication) compiled fault width versus displacement data from a range of sources (with inherent limits of comparison) and also suggest that the growth of the macrofracture damage zone width decreases markedly for larger (>100 m) displacement faults.

Many authors have attempted to relate the fracture damage around faults and the width of the damage zone to the mechanics responsible for their formation. Such studies are particularly attractive for, if the underlying physics are understood, predictions of damage zone properties can be made for any faults where the mechanical properties of the protolith are known. Fig. 6 summarizes the primary processes that have been suggested for the production of fracture damage surrounding faults (Wilson et al., 2003; Blenkinsop, 2008; Mitchell and Faulkner, 2009). Many damage zone models have concentrated on the link to process zone mechanics (Anders and Wiltschko, 1994; Vermilye and Scholz, 1998) or the repeated slip on small patches along a single fault (Shipton and Cowie, 2003). The intensity of tensile damage can be related to slip distribution surrounding the fault tip (Savage and Cooke, 2010) or stress-release at extensional relays (Soliva et al., 2010). The various types of models have, in general, predictable fracture populations and orientations. However, many of the predicted fracture orientations are very similar and distinguishing between them is difficult. Given that many of these processes operate at various times on any given fault, the pattern of fracturing may be very difficult to interpret (Fig. 6f). The conclusions from most recent studies have supported the idea that the fracture damage surrounding faults accumulates from a combination of processes (Shipton and Cowie, 2003; Wilson et al., 2003; Blenkinsop, 2008; Mitchell and Faulkner, 2009).

2.3. Fault-system development

This review has largely concentrated thus far on the development of individual faults. This review has largely concentrated thus far on the development of individual faults. To understand the growth of faults and interpret relationships between displacement and length, or displacement and damage-zone development, we must also consider the interaction of faults or fault segments. For a fault zone with significant displacement, it is difficult or impossible to confidently identify the incremental steps which occurred during its development. Where syn-tectonic (growth) strata are preserved, however, the stratigraphic architecture of these units may provide constraints on the development of fault systems (e.g. Gawthorpe et al., 1997). For example, high-resolution 3D seismic surveys (e.g. Baudon and Cartwright, 2008; Lohr et al., 2008) image the geometry of growth strata. Additionally, numerical and analogue modelling allows systematic study of fault-zone and fault-system development with known boundary conditions (see papers by Henza et al., 2010; Moir et al., 2010; Schmatz et al., 2010).

As proposed by many authors (Walsh and Watterson, 1988; Peacock and Sanderson, 1991; Walsh et al., 2002; Childs et al., 2009), larger faults are the result of the growth and linkage of smaller faults. This growth and linkage process commonly produces a variety of fault-related folds, which are the continuous (ductile) deformation that accompanies and is genetically related to faulting (e.g. Anastasio et al., 1997; Cosgrove and Ameen, 2000; Wilkerson et al., 2002). Fault-related folds are important in that they indicate displacement variations on faults (e.g. relay ramps) or folds resulting from fault-tip propagation, and folds due to movement on non-planar faults (which may form due to the linkage of non-coplanar fault surfaces). Fault-related folds are associated with reverse and thrust faults (Suppe, 1983; McClay, 2004; Shaw et al., 2005), normal faults (Schlische, 1995; Janecke et al., 1998; Withjack et al., 2002), strike-slip faults (Christie-Blick and Biddle, 1985; Harding, 1990) and oblique-slip faults (Tindall and Davis, 1999; Cristallini and Allmendinger, 2001; Schlische et al., 2002). The rock properties in these folds are commonly modified; thus, some fault-related folds may be considered as part of the fault’s damage zone.

The formation of larger faults through the linkage of smaller faults influences scaling relationships (Dawers and Anders, 1995; Kim and Sanderson, 2005). For example, when two smaller faults link, the resultant length might be longer than if the fault grew as a single structure. Walsh et al. (2002) note that once some of the faults in a system become larger than others by segment linkage, they tend to dominate further deformation by increasing displacement rates; this typically results in the cessation of the activity on the smaller faults.

Early attempts to infer the scaling relationships of faults relied largely on log–log plots of maximum displacement versus length. These datasets indicate that similar displacement–length ratios occur for a wide range of scales, suggesting some scale-invariant behaviour to fault propagation and growth (Walsh and Watterson, 1988, also see Bonnet et al., 2001 for a review). The recognition of the importance of segment linkage in fault growth at least partially explained the wide scatter in the data, even on log–log plots (e.g. Cartwright et al., 1995; Dawers and Anders, 1995; Kim and Sanderson, 2005). Interestingly, analyses of the thickness of fault cores showed similar scale-invariant behaviour on log–log plots of displacement versus thickness (Section 2.2). In both cases, the validity of this approach for inferring the mechanics of fault growth or fault core-zone development (e.g. by wear or abrasion) is difficult because of the hazards of comparing datasets from different lithological and tectonic settings. Nevertheless, in the case of fault-length scaling, generalizations of data sets can provide useful upper limits to feasible displacement gradients which are important as rules-of-thumb for quality control of structural interpretations from seismic data, particularly in the case of sparse 2D lines (Freeman et al., 2010). In sedimentary basins, fault linkage typically involves the formation and eventual breaching of relay zones (see Walsh et al.,...
Recent studies of overlap and relay zones have shown how, at a variety of scales, they may evolve into fault-bounded lenses as the deformation evolves and the two segments fully connect to form a single fault (e.g. Kristensen et al., 2008; Childs et al., 2009). Similarly, asperity bifurcation (“splaying”) such as at zones of enhanced wall rock fracturing, and along-strike reconnection of these splays results in isolation of lenses of low-strain wall rock between the splays in a growing fault zone (e.g. Van der Zee et al., 2008). Both of these processes are controlled by heterogeneities of characteristic length scales ranging from grain size through bed (e.g. Wilkins and Gross, 2002; Soliva and Benedicto, 2005) and joint scale (Van der Zee et al., 2008) to structures such as “sidewall rip-outs” at the scale of strike-slip faults in the upper crust (e.g. Swanson, 2005). The scale of these might be controlled by variations in fault orientation, possibly due to fault linkage. Thus, in reality, fault thickness growth may occur by a number of discrete scale-dependent steps (Wibberley et al., 2008). Recent studies using geomorphology (Whittaker et al., 2008; Roberts et al., 2009), shallow seismic surveys (Bull et al., 2006) and numerical modelling (Cowie et al., 2006) have investigated the accumulation of slip at breached relays and the relationship between throw rate and the longevity of slip minima at linkage sites. These results have important implications for basin development and earthquake rupture propagation through fault linkage zones (Roberts et al., 2009; Walker et al., 2009).

3. Mechanics of fault zones and fault systems

Quantification of the mechanics of faulting and earthquakes comes from in situ crustal measurements, laboratory friction testing, field measurements and inversion of seismic data (Scholz, 2002). Relative mechanical properties (without quantifying the absolute stresses) may be derived from field observations and seismic data. Classic work conducted by Byerlee (1978) provides a range of friction coefficients under which fault slip should occur at crustal depths where brittle deformation dominates. A compilation of crustal stress measurements by Townend and Zoback (2000) has highlighted the general applicability of Byerlee’s law. Similarly, an analysis of the dip of seismogenic normal faults indicate that they generally fall at the lower end of the range of friction values suggested by Byerlee (Collettini and Sibson, 2001), and a similar observation holds for seismogenic reverse faults (Sibson and Xie, 1998). In this section we highlight instances where crustal fault strength appears to vary considerably from the picture outlined above. The mechanics of the earthquake process is then addressed.

3.1. ‘Weak’ faulting

Compelling observations indicate that some faults slip under anomalously low friction coefficients, much less than those predicted by Byerlee. The notion of weak faults derives mainly from the orientation of the fault plane with respect to the maximum principal stress. Weak faults are classified as those that appear to slip even when frictional lock-up is predicted by Byerlee’s law. One issue relates to explaining continued slip on a fault rotated out of its ideal orientation. However, another problem is to explain how microfractures; (b) shows damage from linking of structures (c) shows damage from fault growth involving a ‘process zone’ (d) shows damage from continued displacement on ‘wavy’ faults; (e) shows co-seismic fracture damage, where $V_r$ is the rupture velocity and (f) illustrates how a combination of all these can produce a complicated pattern of fracture damage surrounding a fault core (based on work from Wilson et al., 2003; Blenkinsop, 2008; Mitchell and Faulkner, 2009). Note that damage generating processes highlighted in this figure can be active at different stages during the evolution of a fault zone.
some faults appear to have formed at such a non-optimal orientation (e.g. Collettini and Holdsworth, 2004).

The debate surrounding the problem of weakness on large thrusts is probably the oldest of those regarding weak faults. The paradox is that the maximum fault area that a relatively thin thrust sheet can load without breaking up internally is much less than observed in natural examples. Hubbert and Rubey (1959) suggested a solution where basal friction can be overcome by high fluid pressure in the thrust fault zone. Price (1988) on the other hand, showed that the paradox is based on the assumption that slip occurs on all of the thrust surface simultaneously, whereas evidence from earthquakes suggests that thrusts, like other active faults, operate by the overall accumulation of displacement on slip patches on different parts of the fault at different times. Furthermore, thrust zones in crystalline basement show extensive evidence for fluid-enhanced reaction weakening, encouraging a change in deformation mechanism regime from frictional sliding to shear by diffusive mass transfer and/or crystalline plasticity (Wibberley 2005). On the topic of reverse faults, Sibson (2009) recently showed that normal faults on the eastern side of the Sea of Japan reactivated in a reverse sense, as the back-arc basin has started to contract, are unfavourably oriented with respect to the regional stress field, suggesting they are weak.

Another class of faults that may be viewed as weak are low-angle normal faults (Axen, 2007). The existence of these structures is still debated, but compelling recent geological and geophysical data support their occurrence. One example can be found in central Italy, where extension and uplift is migrating from west to east, and previously active extensional structures have been exhumed and are exposed on the Island of Elba. Geological evidence suggests that the Zuccale fault on Elba, currently in a low angle orientation (∼10°), was active in this same position during movement. The evidence for this includes the formation of contemporaneous footwall conjugate extensional faults (Smith et al., 2007) and the orientation of original compressional structures and vertical opening mode extensional veins (Collettini and Holdsworth, 2004). Farther to the west, from the surface to depths around 8 km, Chiaraluce et al. (2007) imaged a microseismically active zone, the Alto Tiberina fault, oriented at a low angle. They interpret this to be a currently active low-angle extensional detachment, of which the Zuccale fault to the east is an ancient, uplifted example.

For strike-slip faults, the San Andreas fault in California is inferred to be ‘weak’ based on heat-flow data, seismological constraints and stress orientation. Early work showed that there was a conspicuous absence of any elevated heat flow from fictional heating that essentially limits the average shear stress on the fault to typical stress drops in earthquakes, 10−20 MPa (Brune et al., 1969; Lachenbruch and Sass, 1980). These data, combined with the observation that the angle between the maximum principal stress and the San Andreas fault is high (most latterly by Hickman and Zoback, 2004; Boness and Zoback, 2006), suggest that the fault moves under friction coefficients of ∼0.2 or less. The interpretation and the quality of both the heat flow and the stress orientation data are controversial (see Scholz, 2006 and references therein). Results from the San Andreas Fault Observatory at Depth (SAFOD) provided more data close to the San Andreas fault at 3 km depth at the southern limit of the creeping section near Parkfield. These data suggest the fault is indeed weak, at least in the creeping section (Hickman and Zoback, 2004; Scholz, 2006).

Regardless of whether or not weak faults exist, the debate has driven a large body of work regarding the strength of faults in general. Laboratory studies have concentrated on the search for a weak mineral phase that might realistically be present over seismogenic depths in sufficient quantities to promote weakening. The frictional strength of most single phyllosilicate phases is less than predicted by Byerlee's law (Rutter et al., 1986; Logan and Rauenzahn, 1987; Morrow et al., 1992; Scruggs and Tullis, 1998; Bos and Spiers, 2001; Saffer et al., 2001; Saffer and Marone, 2003; Moore and Lockner, 2004; Tembe et al., 2006; Takahashi et al., 2007; Crawford et al., 2008; Ikari et al., 2009), although this is not always the case (van Diggelen et al., 2010). What is not so clear at present is how the mixing of weak clay material with other phases, such as quartz affects gouge strength, and how the fluid phase may affect whether sliding is dominated by friction or another, viscous, rheology. There are few studies that have addressed the first question, but those that have suggest a gradual decrease in frictional strength with the addition of clay (Takahashi et al., 2007; Crawford et al., 2008). It also appears that natural fault rock microstructures, with inter-connected networks of weak phyllosilicate phases, are important to produce some form of weakening (Holdsworth, 2004). Collettini et al. (2009b) compared the frictional strength of intact wafers of natural fault gouge from the Zuccale fault (Elba, Italy) to mineralogically identical powders and found the material retaining the in situ microstructure is significantly weaker than its powdered counterpart. This suggests that previous experiments where powdered natural gouge was used provide an upper bound on the strength at best. However, the experiments on single mineralogical components should still provide a useful guide to the lower strength bound for natural gouge containing these phases.

Phyllosilicate phases the required strength to explain ‘weak’ faulting, such as montmorillonite or chrysotile, tend to strengthen at laboratory pressures and temperatures greater than that equivalent to a few kilometres depth (e.g. Morrow et al., 1992; Moore et al., 2004). One exception is talc, which shows remarkable weakness over the entire temperature and pressure conditions of the seismogenic crust (see Fig. 2). Escartin et al. (2008; Moore and Lockner, 2008). For this reason the discovery of serpentinites and associated talc within cuttings from the SAFOD borehole was significant (Moore and Rymer, 2007). However, it is not clear if the talc is concentrated in sufficient quantities or along narrow slip surfaces in the two currently active strands of the San Andreas fault at the SAFOD site. Testing of powdered cuttings from the SAFOD borehole (with inherent limitations, as noted above) showed that friction coefficients are lower than those predicted by Byerlee, but not sufficient to account for the observations of fault weakness (Tembe et al., 2006).

Similarly, the presence of talc in the Zuccale low-angle normal fault has been noted. It is explained by reactions involving dolomite and silica-rich fluids (Collettini et al., 2009b). However, frictional studies of powdered natural gouge from the Zuccale fault suggest that the friction coefficients are too high to explain movement, although these might represent an upper bound of strength as powdered samples were used (Smith and Faulkner, 2010). Similar results apply to other low angle normal fault systems (Numelin et al., 2007). More importantly, the discontinuous nature of the talc-rich fault gouge layers in the Zuccale fault show that, even with the presence of talc, the geometry cannot explain slip (Smith et al., 2007; Smith and Faulkner, 2010). In summary, although intrinsically weak minerals do exist within fault zones, it is currently unclear whether these can be the sole source of weakening.

Laboratory testing generally occurs at room temperature conditions and at strain rates that far exceed their natural counterparts for fault creep. One possible way to weaken faults is that additional mechanisms other than frictional ones are responsible for fault slip. These would only operate at much slower strain rates than those achievable in the laboratory (Gratier et al., 1999). The overall dominance of pressure solution creep over frictional processes in fault zones remains a possibility (Rutter and Mainprice, 1979). Geological observations of many faults suggest that the mineralogical changes during deformation in large faults...
might promote this type of behaviour (Wintsch et al., 1995; Imber et al., 2001; Holdsworth, 2004). If pressure solution is operative in very mature fault zones, the soluble phases may be removed entirely, leading renewed frictional behaviour. Experiments on rock analogues (halite and clay) showed that both frictional processes and pressure solution can operate simultaneously (e.g. Bos et al., 2000). One problem is that characteristic microstructures indicating that pressure solution creep was operative in natural fault rocks are difficult and sometimes impossible to find. Pressure solution experiments on natural rocks, or those that are likely to be present in natural fault zones, are few, and the diffusion rates for quartz, for example, are slow (Hickman and Evans, 1995; Gratier et al., 2009). Consequently it is difficult to assess the possible contribution of pressure solution creep on fault behaviour. However, if used to explain possible weakness, it will be more applicable to faults with low displacement rates (e.g. ~1 mm/year) such as those predicted for low-angle detachment faults, or in the interseismic period between earthquakes (e.g. Renard et al., 2000).

Another possibility for intrinsic fault weakening arises from the concentration of phyllosilicate phases in mature fault cores. Wintsch et al. (1995) suggested that a layer of well-oriented phyllosilicate basal planes might provide a weak horizon within faults. Experiments on phyllosilicates show they have significant mechanical anisotropy (e.g. Mares and Kronenberg, 1993; Mariani et al., 2006 and references therein). Mariani et al. (2006) showed that with muscovite polycrystals at elevated temperature and very low laboratory strain rates deformation approached linear viscous and showed no grain size dependence, suggesting some form of Harper-Dorn creep. Clearly additional experiments at lower strain rates are necessary to investigate these processes further.

High pore fluid pressures are another mechanism that may weaken faults (Fig. 7b; Hickman et al., 1995). One model involves pore pressure that varies over the seismic cycle. Immediately following an earthquake the permeability of the fault is high (Sibson, 1990). Over time, permeability falls and compaction of the fault leads to overpressure and promotes slip (Blanpied et al., 1992; Byerlee, 1993). A second model involves a fault zone that continuously acts as an impermeable barrier and high pore fluid pressure is maintained by a fluid flux from mid-crustal levels (Byerlee, 1990; Rice, 1992). The first model is more applicable to seismogenic faults, whereas the second is better suited to faults undergoing creep. Geological evidence supports both models. For the second model, low permeability fault gouges and a favourable fault zone structure suggests that long term sealing is possible if fluid sources at depth are available (Faulkner and Rutter, 2001). One drawback with these models is that the level of pore fluid pressure required to promote slip on very unfavourably oriented faults generally exceeds the minimum principal stress (Rice, 1992). Thus, the pore pressure would always be limited by hydrofracturing and fluid pressure loss before slip occurs on the fault. This problem is circumvented if the stress state within the fault zone is different to the remotely applied stress field. This condition is possible, with mechanical continuity, if the mechanical properties of the fault zone are different from the country rock (Rice, 1992; Chery et al., 2004; Faulkner et al., 2006).

One final fault weakening mechanism operates during seismic slip (dynamic weakening; Fig. 7c). If operative, dynamic weakening mechanisms can explain the lack of any heat flow anomaly but, in the absence of any other weakening mechanism, cannot explain the nucleation of dynamic events on a severely misoriented fault. The issue of dynamic rupture will be discussed in Section 3.3.

### 3.2. Earthquake nucleation

The issue of how earthquakes nucleate is important because it might produce measurable precursory phenomena that may be used in short-term earthquake prediction (Ellsworth and Beroza, 1995).
In recent years, modelling of the build up to instability has generally utilized the rate- and state-dependent friction laws that predict a period of stable, aseismic slip preceding instability. The rate and state formulation also allows a wide variety of observed fault behaviour to be modelled including aseismic fault creep, slow earthquakes and dynamic rupture (Scholz, 1998).

The principles of rate and state friction are well-known and are only briefly described here (for reviews on rate and state friction behaviour see Dieterich and Kilgore, 1996; Marone, 1998). Although extrapolated to accelerating slip and instability in the case of earthquakes, the formulation is based on laboratory experiments where the response of the dynamic friction coefficient to a step increase sliding velocity is measured (see Fig. 8). There is an instantaneous response of the friction coefficient (or equally the shear traction) to the change in velocity (the rate effect), followed by time-dependent evolution over a particular slip displacement (the state effect) (Fig. 8). The rate and state friction law (Eq. (1)) is purely phenomenological and the physical processes responsible for the observed behaviour, particularly for the state evolution, are poorly known.

\[
\mu = \mu_0 + a \ln \left( \frac{V}{V_0} \right) + b \ln \left( \frac{V_0 e^\theta}{V e^\theta} \right)
\]

where \(\mu\) and \(\mu_0\) are the friction coefficient and the initial friction coefficient respectively, \(a\) and \(b\) are experimentally derived constants, \(V\) and \(V_0\) are the new sliding velocity and the initial sliding velocity respectively, \(\theta\) is the state variable and \(D_c\) is the slip weakening distance.

The evolution of the state variable is a function of time, normal stress and displacement and has units of time. It has been explained in terms of the ‘age’ of the load–supporting contacts and the time required for a new set of contacts to develop following a perturbation of the system (e.g. displacement rate, change in normal stress, etc.). Dieterich and Kilgore (1994) showed direct evidence that the state variable is related to the evolution of the area of the load supporting contacts over time and various normal stresses, termed ‘asperity creep’. Two formulations are widely used to model the state evolution, the ‘aging’ (or ‘slowness’) law and the ‘slip’ law (Eqs. (2) and (3)).

\[
\dot{\theta} = 1 - \frac{V \theta}{D_c} \quad \text{(Aging. Dieterich or Slowness law)}
\]

\[
\dot{\theta} = -\frac{V \theta}{D_c} \ln \frac{V_0 e^\theta}{V e^\theta} \quad \text{(Slip or Ruina law)}
\]

These two formulations produce quite different styles of nucleation and rupture (e.g. Ampuero and Rubin, 2008).

The range of geological materials that have been characterized involve quartz and crushed granite powders. These granular materials generally show velocity weakening behaviour (where \(a-b\) is negative, see Fig. 8) at low slip rates, opening the possibility for unstable slip (Green and Marone, 2002). Most natural fault zones contain at least a proportion of phyllosilicate minerals but experimental studies on these materials are even sparser (Morrow et al., 1992; Scruggs and Tullis, 1998; Reinen, 2000; Saffer et al., 2001; Moore and Lockner, 2008; Ikari et al., 2009; Smith and Faulkner, 2010). These studies generally suggest that phyllosilicate-rich gouges exhibit velocity strengthening behaviour at low slip rates (where \(a-b\) is positive) and thus may be associated with fault creep (Faulkner et al., 2003). Talc, in particular, exhibits inherently stable, velocity-strengthening behaviour under all conditions tested (Moore and Lockner, 2008), although these conditions do not include seismic slip speeds (~0.1 to 1 m/s). Note, however, that montmorillonite and serpentine gouge can both exhibit velocity weakening behaviour (Reinen, 2000; Saffer et al., 2001). Indeed, a recent compilation of experimental work suggests that even materials which exhibit velocity-strengthening behaviour at lower slip rates become velocity-weakening above ~0.1 m/s (Wibberley et al. 2008). Some experimental studies have shown that phyllosilicates can exhibit negative b values (Saffer and Marone, 2003; Ikari et al., 2009; Smith and Faulkner, 2010), which are difficult to interpret physically as a negative b value is generally assumed to indicate an increase in contact surface area with faster slip. Karner et al. (1997) and Blanpied et al. (1998) also report negative b values for granular quartz and granite.

The microphysical processes responsible for the observed rate- and state-dependent behaviour are thermally activated and follow Arrhenius-type behaviour (Chester, 1994; Blanpied et al., 1998; Nakatani, 2001; Rice et al., 2001). They presumably include sub-critical crack growth, crystal plasticity, diffusion and possibly reaction at grain contacts. However, the rate and state formulation does not include temperature. Chester (1994) showed that the activation energy required for wet quartz gouge was consistent with sub-critical crack growth at asperity contacts. This is supported by the conclusions of Frye and Marone (2002) that relative humidity plays a role in the granular friction of quartz and alumina as a result of chemically assisted mechanisms. This is clearly an important area for future research, for if the physical mechanisms responsible for frictional behaviour are known and characterized, then the behaviour at a wider set of environmental conditions can be better predicted. However, we note that future progress in this area of research probably needs to combine experiments with detailed microscopy (SEM and TEM) and theoretical and computational studies due to the complexity of the interactions that occur at the nanoscale (Szlufarska et al., 2008; Mo et al., 2009).

3.3. Mechanics of dynamic rupture

The properties of faults during rupture are typically studied using inversion of seismic data recorded during earthquakes to compute the slip distribution of rupture events (kinematic model). Various models allow computation of the stress field, from which the physical properties of the rupture are inferred (see review by Kanamori and Brodsky, 2004). Recently, complementary laboratory studies and field measurements have led to a better understanding of dynamic rupture, although this is currently a rapidly developing area of research.

The slip history during large earthquakes appears to take the form of a slip pulse rather than a self-similar crack-like rupture (Heaton, 1990). In the slip-pulse model, only part of the fault plane ruptures at any one time. Most seismological models now assume source-time functions of small finite duration and thus implicitly apply the slip-pulse model. However, this mode of slip still raises
a number of unanswered questions. First, the spectral signature of large earthquakes suggests that the corner frequency (lowest frequency seismic waves) scales with rupture area which runs contrary to the idea of a similar-size slip pulse regardless of the size of the rupture. Additionally, in the early stages of failure the slip-pulse model must accumulate approximately the correct amount of slip appropriate to the overall rupture size. This leads to the question of whether an earthquake event knows a priori how large it will be (Marone and Richardson, 2006). Finally, modelling using a rate-and-state framework has shown that slip pulses can only develop under a restricted set of conditions where the initial shear stress on the fault is low (Zheng and Rice, 1998). Low stress ('weak') faults might well exist (Section 3.1), but a large proportion of faults are inferred to have 'high' stress levels (Townend and Zoback, 2000). Does this mean that slip pulse-type behaviour is not possible for these faults?

Seismic records allow modelling of the mechanical properties of rupture, but, in recent years, laboratory and field measurements have provided independent constraints. All these data yield an understanding of the rupture process in terms of the energy required for the rupture but, in recent years, laboratory and field measurements have provided independent constraints. All these data yield an understanding of the rupture process in terms of the energy balance for an earthquake (Eq. (4); see Kanamori and Rivera, 2006). The change in potential energy \( \Delta U_p \) (the sum of the elastic strain energy and gravitational energy during earthquake slip) is the sum of surface energy \( U_s \) (to produce new crack surface area), kinetic energy \( U_k \) (radiated as seismic waves) and frictional energy \( U_f \) (dissipated as heat):

\[
\Delta U_p = U_s + U_k + U_f. \tag{4}
\]

Constraints on the kinetic energy are well known from modeling of seismic waves. Chester et al. (2005) estimated the fracture energy for the Punchbowl fault in California by accounting for all the fracture surface energy in both the damage zone and, more importantly, the core zone, where nano-sized particles are present and account for a significant proportion of the fracture area. Hertzian fracture models (where fracturing occurs at grain contacts of spherical grains) suggest that it is mechanically very difficult to produce such small particle sizes. Sammis and Ben-Zion (2008) suggested that shock loading and sub-critical crack growth under compressive stress, or high strain rate tensile stress may be responsible. The fracture energy expended during the formation of 'pulverized rocks' (Section 2.1) that are thought to develop during seismic slip near the surface is not currently known. However, Biegel et al. (2008) showed that off-fault damage will affect the velocity of an earthquake slip pulse.

Di Toro et al. (2005) showed that a simple analysis of pseudotachylite veins provides broad constraints on the dynamic stress during rupture. This analysis indicated very low (in terms of Byerlee friction) shear stresses driving rupture. Di Toro et al. (2006) corroborated these results with measurements, using the same protolith, of dynamic friction at seismogenic slip velocity. While frictional melting might result in low shear stresses driving slip (after the effects of viscous braking have been overcome; Tsutsumi and Shimamoto, 1997), the friction of other granular materials at high velocity were unknown until recently.

Technical developments over the past 15 years (notably in the laboratories of Shimamoto and co-workers) have allowed measurement of the stresses during high-velocity frictional testing in rotary shear apparatus. Fig. 9 shows typical results from a friction experiment conducted at seismogenic slip velocity. These data complement the data modelled from natural earthquake ruptures. A key feature of the data in Fig. 9 is the dramatic weakening from Byerlee levels of friction down to levels between 0.1 and 0.2. The reasons for this weakening are many, and dependent on the material tested. They include flash heating at asperity contacts (Bowden and Tabor, 1950), silica gel formation (Goldsby and Tullis, 2002), thermal pressurization (Hirose and Bystricky, 2007), frictional melting lubrication (Di Toro et al., 2006) and thermal decomposition (Han et al., 2007). Recent modelling of the earthquake process has started to combine and incorporate some of these additional thermal factors into rate- and state-frictional frameworks (Rempel and Rice, 2006; Rudnicki and Rice, 2006; Segall and Rice, 2006; Noda, 2008; Noda et al., 2009).

The results in Fig. 9 have important implications. First, they explain the long-standing debate on the development of frictional melting in fault zones. Simple analyses show that extreme temperatures are quickly reached due to frictional heating (see Rice, 2006 for a summary). All these models assume friction coefficients commensurate with Byerlee's law. If the shear traction required for seismic slip reduces dramatically then the frictional energy converted to heat is also dramatically reduced. It can also explain the lack of any heat flow anomaly (Section 3.1) over the seismogenic parts of the San Andreas fault.

Another feature of high velocity laboratory tests is the magnitude of \( D_f \), which is on the order of decimetres to metres. In slow frictional testing (Section 3.3) to determine rate- and state-friction parameters, the slip-weakening distance is on the order of microns. A long-standing debate focuses on the apparent discrepancy of \( D_f \) values derived from the laboratory versus values inferred from seismological data, the latter being on the order of a metre. This parameter may scale with fault roughness or fault thickness (Scholz, 1988; Marone and Kilgore, 1993). The emergence of models that account for the different physical processes that occur during seismic slip as opposed to slow laboratory frictional sliding seem to provide an answer to the discrepancy. It appears the seismically derived \( D_f \) is a different parameter to that measured at slow slip rates in the laboratory, and involves fundamentally different physics. This is supported by the range of behaviour observed in high velocity experiments, modelling and from other geological observations that include thermal pressurization (Wibberley and Shimamoto, 2005; Bizzarri and Cocco, 2006), flash heating, thermal decomposition and frictional melting. The value of \( D_f \) might be envisaged to increase as thermally activated processes continue to produce weakening with continued slip (in the same way that the boundary between surface energy and heat varies during slip as suggested by Tinti et al., 2005). It is clear that smaller critical slip distances must exist, otherwise small earthquakes could never nucleate and that a sensitive balance exists between the energy required for the work of fracture and that converted to heat. Unfortunately, testing these hypotheses with seismological data is hampered by the limited frequency bandwidth from which kinematic models of rupture are derived and also the constraints on the resolution and uniqueness of
parameters such as $D_r$ (Spudich and Guatteri, 2004; Tinti et al., 2009). Field observations by Kirkpatrick and Shipton (in press) confirm that slip weakening mechanisms are likely to be spatially and temporally variable across an earthquake fault surface.

### 4. Fluid flow in fault zones

In recent years our understanding of fluid flow through faults has advanced greatly. The typical structure of fault zones (Section 2.1), with a core and damage zone, has provided the framework within which to place laboratory measurements of the fluid flow properties of natural and synthetic fault products into context (Fig. 10). Fault-related fluid flow has also been investigated via a number of indirect data sources such as migrating seismicity at depth, shallow reservoir-induced seismicity, springs, geysers and geothermal systems. These sources have provided some first-order constraints on the rates of fluid flow in natural fault zones at depth, and at length scales unavailable to lab experiments.

In the original Caine et al. (1996) fault core and damage zone model of fault architecture the fault core was visualised as providing an across-fault barrier to flow and the fractured damage zone as an along/up-fault conduit. However, the varying fault architectures outlined in Section 2.1 gives rise to a much more complex set of fault zone hydraulic behaviours. The intricate structure of low and high permeability features within a fault zone can lead to extreme permeability heterogeneity and anisotropy. The permeability of a fault zone, both in-plane and perpendicular to the plane (across-fault) is governed both by the permeability of the individual fault rocks/fractures and, critically, by their geometric architecture in three dimensions (e.g. Lunn et al., 2008). For example, rocks from the fault core are commonly rich in phyllosilicates, which typically have low permeability, but only form barriers to flow if they are continuous throughout the fault plane (Faulkner and Rutter, 2001). Open fractures and slip surfaces (both within the fault core and the surrounding damage zone) have a permeability governed by their aperture distribution, which is in turn influenced by their orientation to the present day local stress field. Such fractures and slip surfaces may have a substantially greater permeability than the host rock; however, their net effect on bulk along-fault and across-fault flow, depends entirely on their connectivity and ability to cross-cut other lower permeability units.

#### 4.1. The hydraulic properties of the fault core and its influence on fluid flow

In natural faults two distinct types of gouge are present. The first are granular materials composed of broken, irregular but roughly equant clasts (in the sense that their long and short axes are approximately equal), and the second are gouges that contain some proportion of phyllosilicate material. Relatively few data on the permeability of ‘granular’ gouges are available but they tend to develop a characteristic grain size distribution (Sammis et al., 1987; Marone and Scholz, 1989) that may suggest a similar permeability development for all these materials. Zhang and Tullis (1998) measured the permeability development in synthetic quartz gouge at a normal stress of 25 MPa. They found that at shear strains up to $\sim 10$ the permeability is reduced by two to three orders of magnitude. This is in agreement with more recent findings of Crawford et al. (2008) and Main et al. (2000). Beyond this shear strain (to a shear strain $\sim 200$), Zhang and Tullis (1998) found the permeability dropped by a further two to three orders of magnitude and that a permeability anisotropy of one order of magnitude developed. This was due to the formation of localized, fine-grained Y shears. These laboratory data are in agreement with field observations and permeability measurements from boreholes that suggest a significant drop in cross-fault permeability in deformation band-dominated faults as the fault core develops through-going slip surfaces (Shipton et al., 2002, 2005).

Fault zones rich in phyllosilicate material tend to have lower permeabilities than quartz and/or framework silicate-rich gouges. Information on the fluid flow properties of phyllosilicate-rich fault zones is necessary to understand fluid flow associated with fault creep (e.g. Faulkner and Rutter, 2001) and earthquake slip (e.g. Wibberley and Shimamoto, 2005; Yamashita and Suzuki, 2009), as many large faults contain significant proportions of clays. Where the fluid-flow properties of fault zones are needed to evaluate the robustness of a fault-bounded hydrocarbon prospect or the field compartmentalizing effects of intra-reservoir faults, estimating the possible phyllosilicate content of the fault zone is critical, along with reservoir juxtaposition geometry. Based on field-observations of fault zones, the two main mechanisms that entrain phyllosilicates (typically shale and/or clays) into fault zones in layered sandstone — shale sequences are shaly or clay smearing (e.g. van der Zee and Urai, 2005) and abrasional mixing (e.g. Yielding et al.,

---

**Fig. 10.** Some physical properties of fault zones related to their structure (damage zone and fault core). (a) Single fault core and (b) multiple fault core, which illustrates the resulting complexity in characterizing the resultant properties.
The necessity of estimating fault properties from limited datasets led to algorithms for estimating fault zone composition which assume one or the other mechanism is operative. For shaly/clay smearing, the Clay Smear Potential (Webber et al., 1978) or the Shaly Smear Factor (Lindsay et al., 1993) rely on parameters such as the thickness of the shale/clay source bed, distance of a point on the fault from that source bed, and/or throw. These algorithms only predict whether or not the smear along the fault is discontinuous (likely leading to leakage) for a given fault throw and, if so, where. On the other hand, the abrasional mixing mechanism led Yielding et al. (1997) to propose the Shale Gouge Ratio (SGR) algorithm, a ratio or percentage of shale in a silicilastic fault zone, which simply assumes that any one point on the fault has a composition identical to the average composition of the sequence past which that point has slipped. In terms of sandstone — shale sequences, this is extremely practical to implement, because the net volume of clay (Vcl) logs from nearby wells can be extrapolated onto the fault (in cases of simple stratigraphy), from which SGR is calculated for all points on the fault for which wall-rock Vcl data exist. In reality, fault zones are much more complex and local small-scale variations can exist even in abrasive fault zones where the rule generally holds. However, quantitatively constraining this variation may in future help predict uncertainties in SGR-based evaluations.

In evaluating the sealing potential of a fault, analysis of juxtapositions of reservoir/carrier beds against other reservoir beds is critical. Basic geometric evaluation of such likely leak points is easily done by fault-plane mapping of hanging-wall / footwall critical. Basic geometric evaluation of such likely leak points is easily done by fault-plane mapping of hanging-wall / footwall critical. However, synthetic phyllosilicate gouges this value appears to be much lower (Zhang et al., 1999), presumably due to the nature and distribution of the authigenic clay phases that develop in natural gouges. Recent work has measured the intensity of clay fabrics by, for example, X-ray texture goniometry (Solum and van der Pluijm, 2009; Haines et al., 2009). The gouges have a generally weak preferred orientation of the clays. Solum and van der Pluijm suggest that this indicates that the clay fabrics are localized phyllosilicates. Haines et al. (2009) imply that the lack of clay fabric indicates the permeability anisotropy must also be low. However, previous work on clay fabrics shows that the permeability anisotropy is not due to clay alignment; instead this can only account for a permeability anisotropy less than an order of magnitude (Faulkner and Rutter, 1998). Alternating microlayers of porous ‘granular’ material and fine grained clay-rich material are observed in the microstructure and explain the anisotropy (Faulkner and Rutter, 1998; Faulkner, 2004). Furthermore, authigenic clay growth observed in TEM images shows growth randomly in pore space in the stress shadows of larger relic grains (Rutter et al., 1986).

The stress history of fault gouge, particularly clay-rich gouge, has been shown to be an important control on the permeability. Bolton et al. (1998) and Zhang et al. (1999) have shown that permeability is reduced in sheared gouges that have undergone normal consolidation, but may increase their permeability if they have undergone overconsolidation and are subsequently sheared at lower effective pressure conditions. Indeed the log-linear relationship of both porosity and permeability with mean effective stress in clay-rich gouges from the Median Tectonic Line, Japan, and dependence on the anisotropy of the stress regime suggest that fluid flow modelling in such fault zones may be based around a Cam Clay-type soil mechanics framework (Wibberley et al., 2008).

The temporal evolution of the permeability of fault gouges, particularly at hydrothermal conditions, has been a recent focus of research. Pure quartz gouges (Nakatani and Scholz, 2004a; Yasuhara et al., 2005; Giger et al., 2007) and quartzofeldspathic mixtures (Olsen et al., 1998; Tenthorey et al., 1998; Morrow et al., 2001) have been tested experimentally. Permeability reduction is found to be initially rapid and progressively slows with time. The rate of permeability reduction increases with increasing temperature (Olsen et al., 1998; Tenthorey et al., 2003; Nakatani and Scholz, 2004a; Yasuhara et al., 2005; Giger et al., 2007). The processes responsible for the permeability reduction are interpreted to be solution-precipitation processes (Nakatani and Scholz, 2004a;
4.2 Damage zone permeability

The permeability of fault damage zones is governed by the host-rock permeability and the presence and geometric composition of both macro-scale fracture networks (which will increase permeability), and of low permeability deformation and compaction bands (which will decrease permeability). Fault damage zone permeability in low porosity rocks (Balsamo et al., 2010) is generally fracture-dominated and governed by the connectivity of the macro-scale fracture network (Fig. 10). This contrasts with high porosity rocks where the damage zone may be more complex and permeability is governed by the frequency and connectivity of both low permeability deformation bands and of high permeability slip surfaces (Lunn et al., 2008). Both macro-scale fracture networks and deformation bands have been shown to decrease in frequency with increasing distance from the fault core (Rawling et al., 2001; Shipton et al., 2002; Wilson et al., 2003; Mitchell and Faulkner, 2009). Wibberley and Shimamoto (2003) measured the permeability of rocks collected from the surface trace of the Median Tectonic line in Japan. Their measurements showed permeability variations over several orders of magnitude, which can be explained by the lithological variation of the fault zone and also the structural complexity. Fig. 10 shows how this structural complexity might result in permeability heterogeneity from the variation in microfracture densities observed around multiple fault cores.

The permeability of the host rock within the damage zone is controlled by the frequency and orientation of microfractures. One potential problem with direct measurement of damage zone permeability from rocks collected from the surface outcrop of fault zones is that they may be subject to weathering or modification since the fault related microfracture network was produced (Morrow and Lockner, 1994). Measurements of rocks recovered from depth from fault-zone drilling projects may overcome this problem.

Experiments indicate permeability of initially low porosity rocks taken to failure increase by two to three orders of magnitude (Simpson et al., 2001; Oda et al., 2002; Uehara and Shimamoto, 2004; Mitchell and Faulkner, 2008). For initially high porosity rocks, the permeability may significantly decrease with deformation in the damage zone (Main et al., 2000). Measurements of permeability from the intact state of rocks to their failure stress can be scaled to the levels and distribution of damage seen surrounding fault zones (see Section 2.1), if a common factor between the two can be found. For example one such common factor is the microfracture density which can be measured in the field and then compared to that in experiments at various stages of damage, where the permeability may be readily measured.

4.3 Estimating bulk fault zone permeability

The physical characteristics of fault damage zones are described in Section 2.1. Estimates of the bulk fault zone permeability, for fault-perpendicular and fault-parallel flow, are derived from numerical models that simulate flow through the fault zone (Brown and Bruhn, 1998; Jourde et al., 2002; Matthai and Belayneh, 2004; Odiing et al., 2004; Lunn et al., 2008) Such models show significant channelling of flow within fault zones into a small number of focussed flow paths. Similar channelling effects are observed in flow experiments within individual fractures (Brown et al., 1998; Beeler and Hickman, 2004). Observations from boreholes that penetrate faults, as well as along-fault and across-fault pumping tests, are necessary to determine bulk fault-zone permeability and to validate numerical models (Evans et al., 2005; Medeiros et al., 2010).

Very few studies have measured bulk fault zone permeability directly using boreholes. However, a number of secondary data sources allow estimation of along-fault permeability. Talwani et al. (1999) estimated the permeability of a shallow fault zone using sinusoidal pressure oscillations in boreholes from lake level fluctuations. Their analysis shows fault zone permeability, in faults that are subject to reservoir-induced seismicity, are between 1.1 \times 10^{-15} and 1.78 \times 10^{-15} m^2. Tadokoro et al. (2000) estimated along-strike fault (damage) zone permeabilities around 1-10 \times 10^{-15} m^2 from the migration rates of induced seismicity during borehole injection experiments in the Nojima fault zone following the Kobe 1995 earthquake. At deeper levels, Shapiro et al. (1997) found crustal permeability to be \sim 10^{-14} m^2 in the KTB borehole in the depth interval 7.5-9 km. A compilation of fault zone permeabilities derived from reservoir-induced seismicity data can be found in Talwani et al. (2007). Using the same techniques of migrating patterns of seismicity, but at deeper levels in the brittle crust, Miller et al. (2004) estimated fault zone permeability to be 4-10^{-11} m^2 immediately following the M6 Colfiorito earthquake sequence of 1997 in central Italy. Noir et al. (1997) inferred a higher fault zone permeability of 10^{-10} m^2 for the Dobi earthquake sequence in Afar in 1989.

4.4 Spatial and temporal variability of fault zone permeability

A number of recent studies have shown that fault zone permeability is highly heterogeneous both spatially and temporally. Studies examining the distribution of geothermal spring temperatures along the Borax Lake fault, Oregon, USA, show that neighbouring springs a few metres apart can have widely different temperatures (Fairley and Hinds, 2004). These suggest that high permeability pathways exist as discrete structures in the Borax fault damage zone, and that individual flow paths have the capability to transport fluids rapidly from depth parallel to the fault plane. Dockrill and Shipton (2010) use observations of natural leakage of CO2 along faults in Utah, USA, to show that along-fault flow is occurring at a few discrete locations along strike, and that these discrete along-fault flow channels have migrated over time.

The presence of a modern oil seep at one location on the fault also indicates the existence of discrete unconnected along-fault ‘pipes’ that provide pathways for fluid flow from lithologies at depth to the surface and are unconnected to shallower horizons. Evans et al. (2005), during an injection test in the Soultz borehole, observed that “some 95% of the flow entered the rock mass at just 10 major flowing fractures”. Do Nascimento et al. (2005) show that pressure changes as low as 0.5 kPa (equivalent to 5 cm of water head) are enough to trigger transient changes in permeability that are spatially correlated, and related to seismicity below a water reservoir.

In the hydrocarbon exploration and production industry, evidence for the spatial and temporal variation of fault permeability exists but is usually limited to indirect inferences from reservoir fluid and pressure data either side of the faults, because faults are generally avoided when drilling as they give rise to a number of drilling problems (e.g. Grauls et al., 2002). Nevertheless, studies of charge timing in fault-bounded blocks (e.g. Residual Salt Analyses of fluid inclusions in reservoir pore cements) can show an increase in the hydrocarbon buoyancy pressure differential across the fault through time, attributed to increasing sealing properties as compaction affects the fault during progressive burial. Transient increases in up-fault permeability during periodic
reactivation may lead to leakage of fault-bounded hydrocarbon traps, as has been found for several cases in the northern North Sea for example (Wiprut and Zoback, 2002). On production time scales, faults which initially sealed significant hydrocarbon-related pressure differences may become leaking as one compartment is depleted (e.g. Dincau, 1998) and stress changes related to depletion may render the fault unstable in certain cases, leading to up-fault leakage during slip (Cuisiat et al, 2010).

One exception to the general paucity of direct industry measurements of fault permeability is the Pathfinder well on Eugene Island, Gulf of Mexico, which shows an active normal growth fault in an overpressured siliciclastic setting to have a relatively high up-fault permeability of ~1 mD which sharply increases to around 1D as fluid pressure is further increased towards the minimum effective principal stress (Losh and Haney, 2006). A similar observation was made from data from an ODP borehole through the basal thrust of the Barbados accretionary prism (Screaton et al., 1996). Seismic data shot on Eugene Island at two different times over a seven year interval image the same velocities (presumably related to healing of fracture damage) following the 1992 Landers earthquake is quite rapid (<10 years) (Vidale and Li, 2003). This recovery may only partially heal fault fracture damage as the low-velocity zone surrounding faults appears to be long-lived (Cochran et al., 2009).

5. Concluding remarks

Recent advances in the study of the structure, mechanics and fluid flow properties of faults and fault systems have been reviewed. The importance of the interplay of these three properties was emphasized. We conclude this work by highlighting some of the key areas of ongoing research.

In terms of fault zone structure, the heterogeneity and along fault variability are still poorly known. For example, seismological methods are necessary to determine the structure of fault zones at depth. Hence, if we are to reconcile the structure of fault zones at depth with that observed at the surface, we must better understand the dependence of seismological parameters on the physical properties of the fault rocks. Basin-scale fault systems can benefit from revised models of fault growth related to geometric and kinematic coherence. The impact of new 3D seismic datasets has the potential to greatly enhance our understanding of fault growth by providing a detailed view of growth strata that are intimately associated with the growth of faults and development of fault systems.

Our understanding of the mechanics of faulting and earthquakes is still limited. The currently unresolved question of weak faulting has helped to focus efforts, but direct observation via fault-drilling projects is necessary to resolve this. Even then, the drill hole has to penetrate a “representative” portion of the fault zone, with all the inherent problems of fault-zone heterogeneity previously mentioned. We can improve our knowledge of the mechanics of earthquakes by integrating data from seismology, experiments and field geology. A key aspect is to improve our understanding of the manifestations of seismic slip and its thermal effects in natural fault rocks, if this is possible. This may be aided in the future with comparison with microstructures developed in high velocity experiments. Experiments (at both high and low velocity) must aim to understand the underlying physics of slip.

Fluid flow around faults is dictated by 3D fault zone structure and, as previously mentioned, this is likely to be heterogeneous. Investigations of along-fault flow, in particular within crystalline rocks at depth, show that flow can be dominated by a small number of fractures within the surrounding damage zone. Recent efforts have helped to characterize the damage surrounding faults and we can potentially reconcile this with laboratory measurements of various fault-zone components (e.g. fault core or damage zone). However, there remains a pressing need for a greater number of in situ measurements over large areas, such as those exposed within tunnels that characterize both the permeability of individual features and whole fault zones at depth. Our understanding of the temporal evolution of fault-zone permeability is still limited, but we can address this by a combination of field and laboratory measurements.

Overall, as many aspects of faulting and fault systems are highly interrelated, an integrated approach is necessary to make progress in understanding their structure, mechanics and fluid flow properties. This integrated approach will necessarily involve many different disciplines, from field geology, laboratory measurements, geophysical measurements, modelling (numerical and experimental), and direct observation of faults through drilling.

Acknowledgements

DRF thanks the Departamento de Geología, Universidad de Chile for their hospitality during the preparation of this manuscript. We thank Bob Holdsworth for providing a thorough review that improved the manuscript.

References


