Introduction

**Cite this article:** Nielsen S. 2017 From slow to fast faulting: recent challenges in earthquake fault mechanics. *Phil. Trans. R. Soc. A* **375**: 20160016. http://dx.doi.org/10.1098/rsta.2016.0016

Accepted: 3 July 2017

One contribution of 6 to a Theo Murphy meeting issue ‘Faulting, friction and weakening: from slow to fast motion’.

**Subject Areas:**
geophysics

**Keywords:**
faulting, strain, friction, weakening, earthquakes

**Author for correspondence:**
S. Nielsen
e-mail: stefan.nielsen@durham.ac.uk

Faults—thin zones of highly localized shear deformation in the Earth—accommodate strain on a momentous range of dimensions (millimetres to hundreds of kilometres for major plate boundaries) and of time intervals (from fractions of seconds during earthquake slip, to years of slow, aseismic slip and millions of years of intermittent activity). Traditionally, brittle faults have been distinguished from shear zones which deform by crystal plasticity (e.g. mylonites). However such brittle/plastic distinction becomes blurred when considering (i) deep earthquakes that happen under conditions of pressure and temperature where minerals are clearly in the plastic deformation regime (a clue for seismologists over several decades) and (ii) the extreme dynamic stress drop occurring during seismic slip acceleration on faults, requiring efficient weakening mechanisms. High strain rates (more than $10^4 \text{s}^{-1}$) are accommodated within paper-thin layers (principal slip zone), where co-seismic frictional heating triggers non-brittle weakening mechanisms. In addition, (iii) pervasive off-fault damage is observed, introducing energy sinks which are not accounted for by traditional frictional models. These observations challenge our traditional understanding of friction (rate-and-state laws), anelastic deformation (creep and flow of crystalline materials) and the scientific consensus on fault operation.

This article is part of the themed issue ‘Faulting, friction and weakening: from slow to fast motion’.

1. Introduction

Faults are active at very different depths (0–30 km for the Earth’s crust, but up to 700 km for observed intermediate and deep earthquakes). Faults live under conditions of ambient pressure and temperature that cover a very wide
span—indicatively, 1–300 MPa and 20–300°C in the Earth’s crust alone; more than 10 GPa and more than 600°C for deep earthquakes.

Their range of deformation rates and styles is also very wide. A single fault may undergo slip at both extremely fast (m s\(^{-1}\), seismic rates) and extremely slow (cm yr\(^{-1}\), inter-seismic creep, up to 50 cm yr\(^{-1}\) for some slow slip events).

Different slip rates usually map onto different portions of the fault (locked versus unlocked fault patches, deep versus shallow fault areas). Such variety in fault behaviour is attributed to changes in friction, originating in variations of structural or compositional fault properties, temperature, normal stress and the presence of fluid pressure.

This theme issue is mostly focused on findings about the co-seismic fault deformation—the fast aspect of fault sliding—which is evolving into an extremely active and productive research area. In the following sections, I offer an opinion on some recent and less recent findings which are challenging our working hypothesis on earthquake slip. These findings broadly fall in three categories which are reflected in the contributions to this theme issue.

2. A fault paradigm

For perhaps the last five decades, the working hypothesis in classic fault research has broadly adopted the following assumptions:

(a) Fault deformation and earthquakes are principally associated with slip (displacement discontinuity) across a surface of negligible thickness (rather often simplified to a planar surface). Deformation accommodated within a larger volume around the fault, other than the purely elastic strain, is generally neglected when considering fault kinematics, dynamics and dissipative terms in the earthquake energy budget. Therefore, it is widely believed that faults are controlled mainly by friction on a surface (or across a gouge layer of millimetric or sub-millimetric thickness).

(b) Velocity-strengthening friction induces stable, slow sliding while velocity-weakening friction is responsible for potentially unstable, fast seismic slip. Such behaviour is often modelled using rate-and-state (R & S) friction laws, which are documented by traditional laboratory tests, where slip velocity steps are imposed using high stiffness machines. Although these allow one to explore only slow slip rates (less than mm s\(^{-1}\)), it was believed quite intuitively that velocity strengthening observed at such rates would preclude acceleration to seismic rates (more than 1 m s\(^{-1}\)) by preventing the onset of unstable behaviour. Therefore, from the phenomenological point of view, until recently a velocity strengthening material was associated with a stable sliding fault portion and vice versa, where no seismic slip was to be expected under any circumstance.

(c) From the micro-physical mechanism point of view, brittle regime is characteristic of seismic slip, while plastic or viscous deformation (crystal flow, diffusion creep, viscous shear of melt) mostly occurs in slow deformation processes either diffuse or localized.

A number of observations have come to challenge the above working paradigm.

3. Challenging observations

(a) Dissipation: is it only friction?

Fault zones comprise one or several localized shear bands which do accommodate most of the slip [1], which have been named principal slip zones (PSZs) or principal slip surfaces. The PSZ is generally thin (50 µm–1 cm), representing high to extreme shear localization in fault strands which show evidence of having hosted seismic activity at shallow (≈ 1 km) [2–4], intermediate (≈ 4–10 km) [5–9] or great depth (upper mantle) [10–12]. Depending on fault-rock composition, the PSZ is variably constituted of ultracataclasism [2,13,14], recrystallized nano- to micro-grains [8,15,16], or
quenched melt ([3,4,17,18] and references therein). In more mature faults, the PSZ is surrounded by a fault core (typically a few decimetres) of cataclastic material. (Note that often, especially for minor fault strands, the PSZ and the fault core are assimilated.) A wider damage zone (typically a few metres, on relatively mature faults) with pervasive diffuse damage in the form of microcracks [19] secondary fault veins [20] or pulverized rock [21,22] is observed surrounding the fault core. Distributed damage has been attributed to the stress concentration around the rupture tip [23], to Mach fronts during supershear ruptures [24], or to stress fluctuations associated with fault roughness at the large or small scale (see [25] and references therein).

Friction laws derived from models or laboratory experiments consider only slip on a fault surface (or within a thin PSZ). Therefore, many earthquake models based on fault friction alone implicitly neglect off-fault dissipation. Cowie & Scholtz [26] observed from field data that the size of the breakdown zone scales with the length of the fault, therefore that energy loss per unit fault area should also scale with fault length. Additional laboratory experiments [23,27] and field studies on natural faults [20,28–31] also indicated that the width of damage zone increases with fracture length. As pointed out by Nielsen et al. [32,33], this offers an interpretation for the apparent discrepancy between fracture energy in large earthquakes (estimated from seismology), and fracture energy resulting from frictional weakening under seismic slip conditions (measured in laboratory experiments). Both fracture energies are compatible from small to moderate slip amounts (\(\Delta u \leq 0.3\) m), but appear to diverge for large slip and large earthquake magnitudes, where a larger ratio of off-fault dissipation to frictional work is expected.

Mechanical work is the product of stress and strain, and because most anelastic fault strain is accommodated by slip within the PSZ, off-fault damage in itself does not necessarily indicate a large amount of dissipated energy. On another hand, it can be argued that work involved with off-fault damage can be significant if sliding dynamic friction is low, as I will argue in the short discussion below.

In terms of mechanical work, dissipation per unit fault area due to anelastic strain \(\nu_{ij}\) can be written as

\[
W = \int_H \sigma_{ij} \gamma_{ij} \, dz.
\]

(assuming implicit summation on repeated indexes), where \(H\) is the thickness of fault zone and \(z\) the fault-normal direction. Defining a representative average for \(\sigma_{ij}\) throughout the deformation episode with \(\gamma_{ij}\) as the final strain, we may write

\[
\sigma_{ij} \equiv \frac{1}{\gamma_{ij}} \int_{\gamma_{ij}} \tau_{ij}(\nu_{ij}) \, dv_{ij}.
\]

(3.2)

The above reduces to

\[
W = \int_H \sigma_{ij} \gamma_{ij} \, dz.
\]

(3.3)

Now splitting in the contribution of the PSZ and the rest of the fault zone we may write

\[
W = \int_h \gamma_{ij} \sigma_{ij} \, dz + \int_{H'} \gamma'_{ij} \sigma'_{ij} \, dz,
\]

(3.4)

where \(h\) is the thickness of the PSZ and \(H' = H - h\) the remaining fault zone, and the prime indicates values outside the PSZ. Taking average values of strain and stress within \(h\) and \(H'\) yields

\[
W = h \gamma_{ij} \sigma_{ij} + H' \gamma'_{ij} \sigma'_{ij}.
\]

(3.5)
Strain in the PSZ is dominated by the fault-parallel shear \( \partial u_x/\partial z \) (i.e. fault slip), where \( z \) is fault normal direction and \( x \) the direction of slip; therefore, we can write the finite shear strain as

\[
\gamma_{zx} = \gamma_{xz} = \frac{1}{2} \left( -\frac{\partial u_x}{\partial z} \frac{\partial u_y}{\partial x} - \frac{\partial u_y}{\partial z} \frac{\partial u_x}{\partial x} + \frac{\partial u_x}{\partial z} + \frac{\partial u_y}{\partial x} \right)
\]

\[
\approx \frac{1}{2} \frac{\partial u_x}{\partial z} \approx \frac{1}{2} \frac{\Delta u}{H'} (3.6)
\]

where \( \Delta u \) is fault slip, and additional strain terms in the PSZ (\( \gamma_{xx}, \gamma_{yy}, \ldots \)) are negligible.

Outside the PSZ, for an indicative bulk anelastic deformation \( \Delta u'_i \) (note that here deformation is intended as a displacement difference, as opposed to dimensionless strain), over a characteristic dimension \( H' \) we may write (neglecting quadratic terms)

\[
\gamma'_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \approx \frac{1}{2} \left( \frac{\Delta u'_i}{H'} + \frac{\Delta u'_j}{H'} \right). (3.7)
\]

Then implicitly summing all possible \( i, j \) in \( \gamma'_{ij} \sigma'_{ij} \) for the off-fault mechanical work, we may write using (3.5) and (3.6)

\[
W' = \frac{1}{2} (\Delta u'_i + \Delta u'_j) \sigma'_{ij}, (3.8)
\]

and hence summing all terms and using stress symmetry we can re-write the anelastic strain dissipation as

\[
W = \frac{1}{2} \Delta u \sigma_{xz} + \frac{1}{2} \Delta u \sigma_{xx} + \frac{1}{2} (\Delta u'_i + \Delta u'_j) \sigma'_{ij} = \Delta u \sigma_{xz} + \frac{1}{2} (\Delta u'_i + \Delta u'_j) \sigma'_{ij}. (3.9)
\]

In the particular case that the fault-parallel component is dominant in the off-fault deformation \( \Delta u' \), this may be approximated by

\[
W \approx \Delta u \sigma_{xz} + \Delta u' \sigma'_{xz}, (3.10)
\]

and reverting to the integral form (substituting \( \sigma_{ij} \) and \( \Delta u \) according to their values in (3.2) and (3.6)) we obtain

\[
W = \int u \tau_{ij}(u) \, du + \int u' \tau'_{ij}(u') \, du'. (3.11)
\]

Strikingly, this simplified formulation of (3.10) and (3.11) shows that mechanical work, and its ratio inside versus outside the PSZ, does not depend on \( h, H \) but solely on deformations \( \Delta u, \Delta u' \) and stress on- and off-fault. This aspect may facilitate field measurements with the aim to estimate dominant off-fault deformation, by simply summing slip on individual fault strands. The reason for not equating \( \sigma'_{xz} \) and \( \sigma_{xz} \) above is that both deformations are not necessarily simultaneous and therefore may take place under different stress level, as further discussed below. This expression will hold whether \( \Delta u' \) results from diffuse deformation or from the sum of localized slip on a number of fractures; continuity of traction allows to use significant average stress throughout \( H \).

I note, for consistency, that expression (3.11) for the case of fault-parallel anelastic shear alone is compatible with the more rigourously derived result of the classic J-integral applied to shear faulting [34], but (3.11) allows one to decouple on- and off-fault contributions, whether they are localized or diffuse.

Because all terms of stress are bounded by the material strength, and that \( \Delta u \gg \Delta u' \), we may expect \( W \approx \Delta u \sigma_{xz} \) and a negligible contribution to mechanical work from off-fault deformation \( \Delta u' \). However, most of the slip \( \Delta u \) takes place when the fault friction is extremely low owing to dynamic weakening. (During seismic slip \( \sigma_{xz} = \mu \sigma_n \), i.e. frictional shear stress under normal stress \( \sigma_n = \sigma_{zz} \), experiments report friction coefficient as low as \( \mu = 0.05 \) under co-seismic slip conditions.) On the other hand, stress during the rupture initiation is high and, during that initial phase in the rupture process, fault slip is still small (\( \Delta u \approx 0 \)).
Therefore, it is conceivable that under mostly low sliding friction, the off-fault term \( W' \) is not negligible in \( W \). The contribution to dissipation due to deformation in the larger volume around the fault will be significant due to the initial phase of rupture where the proximity of the rupture tip induces modest deformation \( \Delta u' \) but large stress concentrations, resulting in significant product of both, while \( \Delta u \) is still small or comparable to \( \Delta u' \). During later phases of rupture, the term \( \Delta u \sigma_{xz} \) may increase only relatively because \( \sigma_{xz} \) drops to extremely low frictional values. Note that \( W \) is not the equivalent fracture energy \( G \), which is obtained by simply subtracting the relaxed, sliding shear stress value \( \tau_r \) from the stress in the \( W \) expression:

\[
G = \int_{u}^{u'} (\tau_{xz}(u) - \tau_r) \, du + \int_{u'}^{u''} (\tau_{xz}'(u') - \tau_r) \, du' = W - \tau_r(\Delta u + \Delta u').
\] (3.12)

Therefore, using only slip on the PSZ and frictional evolution predicted for a single fault strand (left-hand integral in equation (3.12)) would result in an underestimate of the fracture energy.

As a consequence of self-similarity in fracture mechanics, fault slip \( \Delta u \) and off-fault deformation \( \Delta u' \) increase proportionally to rupture length. Slip weakening means that friction becomes less significant with on-going slip and rupture growth, while the relative importance of dissipation from off-fault damage continues to increase. Therefore, larger earthquakes should be dominated by off-fault dissipation, while small earthquakes are still dominated by friction.

The importance of off-fault damage in the energy budget, and its expected increase with earthquake magnitude (or rupture length) given the above arguments, has also been investigated using numerical simulations which allow for ductile deformation off-fault when material strength is exceeded. To my knowledge, the first numerical study on this aspect was contributed by Andrews [35], where a Coulomb yield criterion was applied to the maximum resolved shear stress in any orientation within the volume surrounding the fault surface. At each numerical time step, any stress in excess of the Coulomb limit was compensated by allowing an equivalent amount of anelastic shear strain, and the corresponding anelastic work was computed. With the set of parameters used in Andrews’s study, off-fault dissipation initially exceeded on-fault dissipation when rupture length surpassed about 300 m, thereafter increasing linearly as a function of rupture length. More sophisticated models were subsequently proposed [36–42] confirming the findings of [35]. Other studies instead include explicit fault splays branching from the main fault surface [43,44] as a form of broadening the damage zone.

In addition to mechanical work dissipated by anelastic deformation, energy is also spent in the creation of new surface during rupture process. The amount of surface on the main fault only is insignificant, but does not remain such when adding a multitude of sub-faults and microcracks in the vicinity of the fault. A significant amount of new surface can be obtained only by generating a very dense network of microcracks, in particular, if comminution and thin pulverization involve a significant volume of the fault zone with creation of sub-millimetre fragments (it is readily shown that the amount of surface in a given volume of fragmented material increases as the inverse radius of the fragments). An extreme form of such pulverization is observed on a number of seismic fault outcrops [21,24,45]. No significant deformation is observed within those pulverized rocks in the vicinity of the fault core (the minerals and structures are still recognizable) but the cohesion of the rock has vanished due to the small-scale fragmentation with very high microcrack density. One interpretation is that pulverization develops under extremely high strain rates which are induced nearby the fault in the vicinity of the propagating rupture tip; in particular, in the case of super-shear rupture propagation episodes, strain rate is intensified within the Mach-cone sheet of the shear wave radiated from the rupture tip region [24].

Pulverization under high strain rates has been reproduced in the laboratory using Hopkinson impact bars to create short pulses of intense stress on rock samples [24,46–48]. In a recent example of such experimental work, Barber & Griffith [49] argue that the surface energy represents a substantial proportion of the total mechanical energy under extreme loading conditions, possibly an energy sink comparable to the amount of frictional work on a seismic fault.
Off-fault damage (due to anelastic strain, pulverization or both) provides a mechanism for stopping large earthquakes. As strain energy release rate scales with rupture length, a larger and larger fracture energy (or equivalent energy sink) is required to limit rupture propagation velocity and eventually to arrest the earthquake rupture. Because frictional dissipation remains bounded (to avoid the paradox of negative friction [33]), and because friction all but vanishes in the dynamic sliding steady state [50], larger earthquakes would never stop unless off-fault dissipation is significant. In this situation, earthquake faults—which are essentially shear cracks—become similar to mode I cracks (opening or tension cracks) where friction is absent and fracture energy is controlled by ductile deformation in a finite volume around the crack tip.

The above observations challenge the assumption (a): that the energy sinks in the earthquake rupture process are dominated by friction.

(b) Fault slip: stable, unstable or both?

Faults accommodate slip in a variety of fashions: by relatively constant slow, stable sliding; during episodic slow slip events which sometimes generate tremor; by seismic slip at high slip velocity (m s$^{-1}$) and fast rupture propagation (km s$^{-1}$) accompanied by radiation of elastic waves. Such spectrum of behaviours can be viewed as going from the most stable to the most unstable condition.

Potential to generate instability and eventual seismic rupture depends on two criteria:

(i) The steady-state frictional stress $\tau_{ss}$ on the fault decreases with slip $U$ and slip velocity $V$ (velocity weakening behaviour):

$$\frac{\partial \tau_{ss}}{\partial U} < 0 \quad \text{and} \quad \frac{\partial \tau_{ss}}{\partial V} < 0.$$  \hspace{1cm} (3.13)

(ii) The frictional relaxation is faster than the stress release due to slip. In the latter criterion, fault stiffness and a critical frictional stiffness are compared and the condition for instability may be written [51] as the inequality

$$\frac{C \mu'}{L} < -\frac{V}{D_c} \frac{\partial \tau_{ss}}{\partial V} + f(V, \ldots),$$  \hspace{1cm} (3.14)

where the left-hand side represents fault stiffness (the ratio of shear modulus $\mu'$ to the fault dimension $L$, times a dimensionless geometrical factor $C$). The right-hand side represents the velocity dependence of friction, with additional terms $f(.)$ involving inertia and the state of evolution of the fault. Because fault stiffness is positive, criterion (3.14) implicitly relies on (3.13); therefore (3.13) is a necessary but not sufficient condition for instability.

Traditionally the frictional behaviour of rocks has been determined through experiments conducted at slow sliding velocities (micrometres to millimetres per second)—these velocities are in excess of the tectonic loading rates ($\simeq$ cm yr$^{-1}$), but still orders of magnitude lower than during co-seismic slip (m s$^{-1}$). Based on these experimental results, a mathematical form of friction known as R & S law [52], with one or more evolving state variables, can be defined. With a single state variable, such friction is governed by two dimensionless parameters $a$ and $b$ (representing direct effect upon a velocity step and the subsequent evolution effect, respectively) and an evolution distance $D_c$.

The difference $(b - a)$ represents the dependence of friction on logarithm of velocity, therefore determining the velocity dependence in (3.13). Then under R & S friction and $a - b < 0$ it can be shown (assuming negligible inertia and evolution term such that $f(V, \ldots) = 0$) that the critical frictional stiffness in the right-hand side of (3.14) becomes

$$k_c = -\frac{V}{D_c} \frac{\partial \tau_{ss}}{\partial V} = \frac{\sigma_n(b - a)}{D_c},$$  \hspace{1cm} (3.15)
where $\sigma_n$ is normal stress. The condition (3.14) with (3.15) is very widely used as criterion to predict stability of fault materials. There are complications in this stability criterion when rupture is allowed to propagate ($L \neq \text{const}$). In such a case, it can be shown numerically that both $(b-a)$ and $a/b$ will control the onset of instability [53].

In the limit case where $C(\mu'/L) \approx \delta \tau/D_c$ (or $a \approx b$ under R & S law), it is believed that small fluctuations in properties combined with episodic slow slip can generate emergent, small-scale instability resulting in tremor-like, prolonged behaviour almost below the limit of instrumental detection, and in the frequency range between 1 and 10 Hz. This may happen in the transition region between the deep, stable and the shallower, unstable portion of major plate boundary faults [54].

The general expectation has been that fault patches which show steady-slip behaviour (as observed from geodetic measurements) were constituted by material with velocity-strengthening properties which would remain such throughout the seismic cycle. Given the range of velocities (and duration) where R & S friction was characterized and the emphasis on $(b-a)$ as a control parameter, slip stability criteria have often been applied to natural faults by extrapolating laboratory data of several orders of magnitude towards the lower (inter-seismic) and the higher (seismic) regimes. The measured value of $(b-a)$ in many material constituents of fault rocks is very small, and the velocity dependence in R & S laws is logarithmic; therefore, the predicted weakening upon extrapolation from laboratory to seismic velocities is modest (usually a few per cent).

However, extreme dynamic weakening of friction is now being very widely observed under fast (seismic) slip rates. Such knowledge has been established only recently, because the technical implementation of high-velocity, seismic-like conditions in laboratory experiments was achieved no earlier than the 1980s with the pioneering work of Shimamoto & Tsutsumi [55] and became widespread in the last decade (see [8,33,56–72] among others). It is notable that such high-velocity weakening has been documented even on rocks which show rate-strengthening in traditional, slow frictional tests. These comprise clay-like material which is expected in the accretionary prism traversed by the shallowest part (repeatedly interpreted as stable sliding) of seismogenic oceanic thrust faults [73].

An empirical formulation that combines the traditional R & S results with an enhanced dynamic weakening as the inverse of velocity was been proposed by Zheng & Rice [74], who used it in numerical modelling of seismic slip pulses. The experimental foundation for such an empirical formulation was subsequently demonstrated by Spagnuolo et al. [72] based on high velocity, rotary shear experiments on silica-built and carbonate-built cohesive rocks. Such formulation can be synthetized as follows:

$$\mu_{ss} = \frac{\mu_o + (a-b) \log(V/V_o)}{1 + (V/V_c)^p},$$  \hspace{1cm} (3.16)$$

where $\mu_{ss}$ is the steady-state friction coefficient. The top part of the fraction represents the usual R& S formulation of friction, where $\mu_o$ and $V_o$ refer to values for friction and sliding velocity, respectively. The denominator, on the other hand, is introduced to account for substantial weakening to kick in when sliding velocity $V$ is close to, or larger than, a characteristic velocity $V_c$. The experimental data on both high and low velocity friction were best fit by using $0.08 < V_c < 0.13 \text{ m s}^{-1}$, and an exponent $p$ indicatively in the range 0.4–1 (the higher values were obtained on carbonatic rocks), in combination with usual $a, b$ parameters for R & S friction.

To illustrate the effect of the velocity dependence in (3.16) on fault stability, we can compute the corresponding critical frictional stiffness (again assuming that we are close to the steady state and that inertia is negligible). Taking the indicative value $p = 1$ in (3.16), we obtain

$$k'_c = \frac{\sigma_n}{D_c} \left( \frac{b-a}{V/V_c+1} + \frac{\mu_o + (a-b) \log(V/V_o)}{(V/V_c + V_c/V_c + 2)} \right),$$  \hspace{1cm} (3.17)$$
where we retrieve $k_c$ as of (3.15) in the limit $V \ll V_c$, but here the critical stiffness varies and peaks substantially in the vicinity of $V = V_c$, as argued in [72]. This indicates that fault instability is enhanced if the experimentally observed velocity weakening is allowed to kick in.

It is striking that in (3.16) even if $(a - b)$ is positive, velocity weakening can be achieved at sub-seismic slip rates, and $k'_c$ will also become positive and peak close to $V = V_c$. Therefore, a slight acceleration may allow the fault to become unstable.

However, a slight acceleration on a creeping fault section means that the fault instability somehow has already started—a chicken and egg situation of sorts—therefore the extended criterion (3.17) cannot be applied to the slow nucleation phase, but would assume some independent triggering process. An example of such triggering was illustrated in a numerical model [75] where a velocity-weakening, unstable fault patch generated the initial instability. Seismic slip then propagated into an otherwise stable creeping patch, making it dynamically weak (here the assumed mechanism for weakening was thermal pressurization of the fluids during fast slip on the creeping patch).

To corroborate these laboratory experimental and these numerical results, co-seismic slip has been observed within the shallow portion of thrust faults, which are clay-rich regions of typical velocity-strengthening behaviour. The most striking example is the large co-seismic slip observed during the Tohoku, 2009 earthquake in the up-dip part of the trench. Models based on a more traditional set of assumptions had failed to forecast co-seismic slip in such portion of the fault [76] and had to be subsequently revisited [77]. In this case, the interpretation is that the shallow portion of the fault gradually accelerated under the impulse of rupture propagating from larger depths of the fault, until it too reached a critical weakening velocity. Additionally, during the so-called tsunami earthquakes, ruptures appear to initiate within the shallow portion thrust faults, an even more paradoxical situation in terms of the expectation of stable sliding in shallow trench faults.

In response to the recent Tohoku, 2011 earthquake, Noda et al. [78] propose new ways to incorporate the enhanced dynamic weakening in a numerical model of thrust fault behaviour. Contrary to previous attempts [74], Noda et al. choose to model the enhanced weakening by incorporating a quadratic law for log velocity dependence in friction. They fit the empirical law using data from laboratory friction experiments, which were performed on samples collected from the Japan trench during the J-FAST drilling project following the earthquake. According to their model of earthquake cycle on a vertical section of the fault, essentially two types of earthquakes can occur: large catastrophic events which break through the topmost fault section to reach the ocean bottom, or intermediate events which are confined at depth to the blueshist region.

Finally, recent laboratory experiments provided evidence that velocity-hardening friction does not necessarily preclude the spontaneous nucleation of seismic-like, stick–slip instability if a sufficient level of small-scale inhomogeneity is introduced on the fault [79]. The experiments were performed on cohesive, pre-cut samples of Westerly granite, where simulated faults had been prepared by grinding the surface to achieve variable levels of roughness, and submitted to variable levels of confinement (from 30 to 200 MPa). The experimental set-up (direct shear of cohesive, pre-cut rocks) shares similarities with earlier experiments [80,81], but with a much higher confinement stress and smaller scale, as in more recent realizations [82,83]. In [79], velocity-stepping was imposed during the experiments, allowing one to measure the dimensionless parameters $a, b$ and the weakening distance $D_c$ characterizing R & S friction. Under a subset of conditions, a clearly positive value of $a - b$ was found, which indicates rate-hardening and is classically believed to generate only stable slip. Even within this very subset indicating $a - b > 0$, stick–slip and seismic instability episodes were triggered inducing partial melt of the slip surface. Using scaling arguments and the self-affine roughness of natural faults, the authors proposed that the same mechanism may trigger stick–slip instability on natural earthquake faults. The interpretation of this unstable behaviour is that when the inhomogeneity of the fault is enhanced (in this case by roughness), weak patches are formed due to normal stress fluctuations which act as stress concentrators and initiators of instability.
As reported in the experimental study proposed by Rutter & Hackston [84], a stable-sliding fault can also become seismic under the effect of a fluid injection. Here, too, we note that the effect of inhomogeneity is crucial, as rupture is triggered when the rate of injection is high, preventing pressure diffusion to a larger zone and creating a localized region of low effective normal stress. Fluid pressure can locally increase due to natural causes—the failure of a nearby seal of underground volatiles [85]—or to anthropogenic activity [86] such as injections to enhance hydrocarbure or hydrothermal productivity.

These observations challenge the stability concept (b): that fault portions showing stable creep—or within material that shows stable R & S behaviour under slow slip conditions—cannot undergo seismic slip or the nucleation of stick–slip instability.

(c) Seismic shear deformation: brittle, plastic, viscous. . .all of the above?

Deformation associated with crustal faults is typically described as a combination of dry frictional sliding on a surface (or within a gouge in a PSZ of minimal thickness), surrounded by a fault zone of intense fracturing and cataclasis; all such processes typically belong to the brittle regime. However, growing evidence of crystal plastic deformation and melting has been reported both on natural and experimentally simulated seismic faults. I discuss below the implications of these, together with other deformation and weakening mechanisms essentially triggered by an abrupt co-seismic temperature rise.

A classic problem arising when considering earthquake slip at velocities in excess of 1 m s$^{-1}$, occurring under the typical shear stress levels expected from dry friction (see former paragraph, R & S laws), is the intense heating and temperature rise which should occur on the fault. It is well known [17,87] that melting temperatures of the rock should be reached even for moderate size earthquakes by a simple back-of-envelope calculation for a thickness $h$ of the PSZ and thermal diffusion of a heat rate $\tau V$:

$$\Delta T = \frac{1}{\rho c \sqrt{\kappa \pi}} \mu \sigma_n V \sqrt{t} \quad \text{for } h \ll 2\sqrt{\kappa t}$$

and

$$\Delta T = \frac{1}{\rho c h} \mu \sigma_n V t \quad \text{for } h \gg 2\sqrt{\kappa t}$$

(with typical values for crystalline rock of $\kappa = 1.6 \times 10^{-6}$ m$^2$ s$^{-1}$, $\rho = 3500$ kg m$^{-3}$, $c = 1000$ J kg$^{-1}$ K$^{-1}$). Assuming a static friction angle of $\theta = 30^\circ$, and Andersonian stress state, the value of normal stress would be three-quarters of the lithostatic load for a fault near static failure [88]. Further assuming hydrostatic pore pressure at a depth of 10 km, an indicative value of effective normal stress would result in $\sigma_n \approx 180$ MPa. Under the modest weakening expected for R & S, we would still have $\mu \approx 0.56$ and for an earthquake with $V = 1$ m s$^{-1}$, $t = 1$ s we obtain $\Delta T \approx 12,800^\circ$ in the first case of (3.18) and $\Delta T \approx 2280^\circ$ assuming adiabatic shear heating within a PSZ of $h = 10^{-2}$ m in the second. In both cases, some severe alteration of the fault conditions is expected in the early phases of seismic slip, with phase transitions (melting, decomposition, amorphization, dehydration, decarbonation in the case of carbonatic rocks such as limestone or dolostone), supercritical fluid pressurization, and the thermal triggering of efficient shear strain mechanisms which therefore would mitigate further rise in temperature (as indeed no significant heat flow anomaly is detected near active faults ([89] and references therein)). These arguments have fostered an extremely active research field regarding the thermally triggered weakening mechanisms likely to take place on earthquake faults.

Two of these mechanisms are re-visited by Rice [90]: flash weakening, a process which is first encountered in early metallurgic research [91] and which can be mathematically formalized and extended to slip on fault rocks [92]; and thermal pressurization of either native fluids trapped in the fault zone or volatiles resulting from co-seismic decomposition reactions. Both models have been extensively used in earthquake rupture modelling ([75,93] and references therein). While flash weakening is widely observed under fast-slip conditions, direct experimental evidence of thermal fluid pressurization is scarce. The first unequivocal evidence of thermal pressurization...
has been documented in two experimental studies by Violay et al. [70,94], but they show that pressurization should start after several metres of slip only. Alternative weakening mechanisms are more efficient in the initial phases of slip which kick in at very early slip stages buffering the background temperature rise and initially preventing an efficient pressurization. For the types of simulated faults used in [70,94] and the extrapolation they offer to natural faults, it can be argued that pressurization is more likely a mechanism arising during large earthquakes, providing further weakening to a fault which has already achieved lubrication.

Frictional melting, on the other hand, appears as a rather intuitive consequence of co-seismic heating and represents an attractive model for dynamic fault lubrication. Its fossil product, a crypto-crystalline to vitreous solidified melt (pseudotachylyte or PST) is indeed observed on some crystalline fault sections which have been exhumed from intermediate crustal depths [10,17,95] or upper mantle shear zones [12] with frequent lateral injection veins.

It is straightforward to obtain melting of crystalline fault rocks during experiments conducted at co-seismic conditions of normal stress and slip velocity [58,68]. The frictional behaviour observed shows indeed pronounced lubrication under high (≈1 m s\(^{-1}\)) slip velocity, where a continuous layer of superheated melt is created that supports viscous shear. The melt layer remains generally thin (≈30 µm–1 mm) as excess melt is extruded to lateral veins on natural faults, or from the edges of the sample of experimental faults. Theoretical arguments indicate a modest increase of steady-state sliding shear stress \(\tau_{ss}\) with normal stress \(\sigma_n\), such that \(\sigma \propto \sigma_n^{1/4}\) [63], a trend which is confirmed by experimental investigations in high velocity rotary shear experiments [68,96]. On natural faults the average co-seismic sliding shear stress can be estimated based on the volume of melt produced [17,58,65]; this confirms extremely low (less than 0.1) equivalent friction coefficients during dynamic sliding, compatibly with the experimental results.

However, rocks undergoing fast slip show an initial weakening phase that takes place much earlier than background melting temperature is reached, and even before profuse melting is formed. This indicates that the temperature rise and the weakening occur much faster at isolated contact asperities on the sliding surface, due to their large localized stress concentrations—a behaviour clearly indicative of the aforementioned flash weakening.

While profuse melting will start to form only after finite slip (centimetres at typical mid-crustal conditions), unambiguous traces of melting are observed even for minimal slip amounts (fractions of millimetres) when conducting microstructural analysis of crystalline rocks [97,98] and even under wet conditions [70] in post-experimental samples that slipped under seismic or microseismic conditions.

It is also argued that melting is not very generally observed on seismic fault outcrops. There are several possible explanations for the causes of such paucity. First, the melting process may be a rarity which occurs only under a specific set of conditions (immature faults and dry crystalline rocks in continental crust environment). However, it can be rebutted that many of the accessible outcrops expose only the part of the fault which has been seismically active at shallow depth, where normal stress and frictional heating are not sufficient to produce melting, but which are not representative in the budget of mechanical energy release contributing to the rupture. Fossil seismic faults which have been exhumed from depth are more rare. PST veins are often extremely thin and difficult to observe and recognize in the field, unless expected and specifically investigated. Also, the amorphous material constituting PST is easily altered, recrystallized (for example, transformed into chlorite or epidote), overprinted or destroyed [99]. Strikingly, products of co-seismic melting and PST were observed on samples which were drilled from active faults in the months after an earthquake [3,4], although the depth was relatively modest (a few hundred metres).

Lastly, on carbonate rock composed predominantly of calcite or dolomite (limestones, marbles or dolostones which host many crustal earthquakes in the Mediterranean area, among others) shear strain appears to localize within a thin (≈50–100 µm) PSZ layer with extreme grain reduction and crystal plastic deformation taking place, but no melting. Indeed, thermal dissociation is reached in carbonate rocks at quite lower temperatures than melting, as argued
in [59,60,100]. In the advanced stages of slip, the PSZ exhibits a densely stacked polygonal aggregate of small (few tens to hundreds of nanometres) crystal grains, with structure typical of grain-boundary sliding plasticity, a regime where superplastic behaviour (the capacity to accommodate finite plastic strain under high deformation rate) has been reported for ceramics. Such structures have been observed in experimental simulated faults after sliding at seismic slip velocity [8,16] but also on samples of natural faults from tectonic areas involving carbonate rocks [7,16].

Plastic flow laws generally depend on grain size $D$ and on temperature $T$; strain rate $\dot{\gamma}$ and shear stress $\tau$ can be equated by

$$A \tau^n = \dot{\gamma} e^{H/RT} D^b,$$

(3.19)

where $H$ is creep activation energy and $A$ a dimensional normalization factor. In the case of grain boundary sliding the exponents are $n = 1$ and $2 < b < 3$, with slightly increasing values if dislocation creep component is present [8]. Within the validity domain of (3.19), observing that the value of $b > 1$ with $D \ll 1$ and the exponential decay with temperature, we may expect strain accommodation to be ever more efficient as temperature increases and grain size decreases. However, this type of flow law has seldom been explored at strain rates in excess of $1 \text{s}^{-1}$, while in the case for the co-seismic slip in carbonate PSZs the inferred values can be of the order of $\dot{\gamma} = 10^4 \text{s}^{-1}$. Therefore, while extreme weakening is observed in combination with dynamic recrystallization within the co-seismic PSZ, deformation takes place in an uncharted, high-end territory of strain rates, most probably involving flow laws and mechanisms previously unreported.

Plastic flow has also been invoked as the mechanism for deep- and intermediate-focus earthquakes, within silicate-build constituents of the subducting slab, on the double Wadati–Benioff seismicity planes. In this case, it is proposed that the superplastic behaviour is triggered by phase transformation, for example, olivine–spinel transition under pressure increase or dehydration embrittlement of serpentine.

A review of experimental evidence of the latter phase transformation in connection with seismic slip is proposed by Green [101], who also discusses broader implication for earthquakes. As pointed out by Green, nanometric material weakening has been an intensely debated topic, but remains to date poorly understood. Hypotheses on its possible origin have been proposed such as the inverse Hall–Petch effect; whereas grain-size reduction normally hardens the material by inhibiting dislocation creep, it can enhance other weakening processes such as the grain boundary sliding as explored in [8] and briefly discussed in the preceding paragraph. In any case, a growing body of evidence supports the idea that plastic- and melt-related flow does take place on some, if not all, earthquake faults.

These observations challenge assumption (c): that seismic slip should belong to purely brittle regime rather than plastic or viscous regimes.

4. Conclusion

What emerges is a considerably more complex picture of fault behaviour and on the controls of stable versus unstable slip than that offered by the paradigm of §2. It will be a non-trivial endeavour for the Earth science community, or indeed for a multi-disciplinary science community, to explore the variety of micro-physical mechanisms responsible for the enhanced dynamic weakening which is maintained at high slip velocity, but also to highlight the way that the weakening mechanisms are triggered in the first place. A quick glance at some of the ongoing research can be gathered in the present issue on slow to fast faulting.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. I acknowledge the support of the Royal Society for publishing this theme issue and for the help and support in the organization and hosting of the related conference. I am indebted to the many
reviewers who helped to improve the papers included in this issue. I thank G. Di Toro for indicating some useful additional references. I acknowledge R. E. Holdsworth for encouraging me with the conference, and J. R. Rice, T. Mitchell and A. Schubnel for co-chairing it.

References


