Modeling earthquake source processes: from tectonics to dynamic rupture

Appendix B

V. Multidisciplinary modeling challenges (MMCs), detailed descriptions of MMC1-MMC3

MMC1. Developing constitutive laws for deformation/faulting: capturing small-scale processes

MMC1.1: Shear resistance of granular fault shear zones: localization, dilation, shear heating, healing, and fluid effects

The highly localized shear layers at the core of seismogenic faults (Figure II.MMC1) are the weakest link in the fault response to shear loading, as they have been shown to accommodate most of the shear motion across faults, often repeatedly over long time periods (Chester and Logan, 1987; Chester et al., 1993; Wibberley and Shimamoto, 2003; Sutherland et al., 2012). They are characterized by highly granular rock material, usually called fault gouge, with broad particle size distributions from several nanometers to several tens of microns (Chester et al., 2005). Clearly, the shear processes in the layers break down the rock materials to miniscule particles, many of them potentially theoretically strong, due to absence of internal defects, or at least much stronger than rocks on larger scales.

A fascinating finding is that the shear-layer width (i.e., its spatial extent in the fault-normal direction), at least in studies of mature faults exhumed from sufficient depth, can be incredibly narrow, less than a millimeter to several millimeters wide (Flinn, 1977; Sibson, 1977; Chester and Logan, 1987; Chester et al., 1993; Chester and Chester, 1998; Faulkner et al., 2003; Wibberley and Shimamoto, 2003; Sutherland et al., 2012) (Figure II-MMC1.1). Incredibly, kilometers (!) of relative shear motion can be accommodated within such narrow layers (Chester and Chester, 1998; Wibberley and Shimamoto, 2003), indicating that they can be both long-lived and capable of hosting multiple earthquakes and/or long-term slow slip in the same spatial location within the fault. Some fault cores consist of a single narrow shear zone, the case generally relevant to mature faults at seismogenic depths (Chester and Logan, 1987; Chester et al., 1993). Others contain multiple shear strands; such more complex situations are discussed in MMC1.4.

Multiple coupled physical and chemical mechanisms in the shearing layer

The relatively narrow width of the localized shear layers - and their clear importance to the overall fault resistance to shear - have made them highly suitable candidates for studies in the laboratory (Figure V-MMC1.1). For temperature and stress conditions relevant to the seismogenic portions of faults, the lab
experiments and their theoretical modelling have collectively identified a number of coupled physical and chemical mechanisms that can significantly contribute to the response of such layers.

Dilation/compaction, the associated pore pressure effects: As the shear initiates and progresses, the shearing layer dilates for initially compacted/densely packed layers and compacts for initially uncompacted/loose layers (Scuderi et al., 2014). The compressed shear layers at the mid-seismogenic depths after a period of interseismic locking are presumably densely packed and hence would initially dilate with slip. There is some evidence for slip-rate dependence of the dilation/compaction, with higher dilation for higher slip rates (Marone et al., 1990; Lockner and Byerlee, 1994; Segall and Rice, 1995; Samuelson et al., 2009). If the layer is permeated with fluids, as would be the case for many natural faults of interest, the dilation/compaction of the shear layer coupled with its overall low permeability causes significant local variations in the pore fluid pressure and effective normal stress, leading to significant variations of the resulting shear resistance (Lockner and Byerlee, 1994; Segall and Rice, 1995; Ikari et al., 2009; Samuelson et al., 2009; Faulkner et al., 2018).

Local particle contacts, their evolution and healing: Both the size and strength of local frictional contacts between particles depend on the mineral composition of the granular layer. They evolve with slip and slip rate and the associated dilation/compaction of the shearing layer. As the layer dilates or compacts, the local contact area evolves, decreasing or increasing on average, respectively, as clearly demonstrated for frictional contacts on compressed rough bare surfaces (Dieterich and Kilgore, 1994) and presumably also relevant to compressed gouge layers. The local contact strength appears to be rate-strengthening under low slip rates (Rice et al., 2001). In stationary contact (zero slip rate), the contacts grow and/or strengthen, resulting in time-dependent strength increase called healing (Dieterich, 1972; Dieterich and Kilgore, 1994; Beeler et al., 1994; Marone, 1998ab). In the presence of chemically active fluids, such as water permeating the pore space, the healing is facilitated by chemical processes such as pressure solution (Spiers et al., 1990; Bos et al., 2000; Bos and Spiers, 2001; Zhang and Spiers, 2005; Niemeijer et al., 2002, 2008).

Evolution in particle-size distribution and mineral composition, localization/delocalization: During shear, the particles can break and experience mineral transformations, such as the development of silica gels (Goldsby and Tullis, 2002; Kirkpatrick et al., 2013) or clay coatings on shear fabrics, which can dramatically weaken the layers (Goldsby and Tullis, 2002; Di Toro et al., 2004,2006; Schleicher et al., 2006, 2010; Collettini et al., 2009; Niemeijer et al., 2010b,2012; Carpenter et al., 2011, 2012, 2015; Collettini et al., 2011; Warr et al., 2014; Wojatschke et al., 2016). Perhaps promoted by such local alterations, the shear deformation tends to further localize with increasing slip even within the already narrow shear layers, especially at relatively high slip rates, sometimes into multiple strands (Poirier, 1980). The more localized shear can presumably also delocalize, e.g., in the presence of local roughness. Such localization/delocalization processes would affect the dilation/compaction as well as the effects of shear heating; the same heat input into a narrower slipping layer would generate higher local temperatures.

Shear heating and enhanced dynamic weakening: At slip rates approaching co-seismic values of ~1 m/s, shear heating input becomes large enough to raise the local temperature within the actively shearing portion of the layer (Sibson, 1973; Lachenbruch, 1980; Mase and Smith, 1985,1987; Rice 2006), with a number of potential consequences including: flash heating and dramatic weakening of the particle contacts (Archard, 1959; Rice, 2006; Beeler et al., 2008; Goldsby and Tullis, 2002,2011; Platt et al., 2015); thermal pressurization of pore fluids - that expand much more readily than the surrounding pore space - and the associated weakening due to decreasing effective confining stress, unless counteracted by fluid escaping the shearing layer due to sufficiently high permeability (Wibberley, 2002; Rempel and
These coupled mechanisms combine to create significant variations of the shear resistance with slip and slip rate, including both weakening and strengthening, as well as healing in stationary contact. The resulting shear resistance and its evolution depends on a number of factors including the stress/temperature/pore-fluid-pressure conditions, evolving mineral composition and particle distribution of the shearing layer, and, for laboratory experiments, the sample design and response of the experimental apparatus (Marone et al., 1990; Spiers et al., 1990; Lockner and Byerlee, 1994; Tsutsumi and Shimamoto, 1997; Bos et al., 2000; Bos and Spiers, 2001; Goldsby and Tullis, 2002; Di Toro et al., 2004, 2006, 2011; Zhang and Spiers, 2005; Niemeijer et al., 2002, 2008; Han et al., 2007; Samuelson et al., 2009; De Paola et al., 2011).

A number of shear-heating and alteration mechanisms can result in enhanced dynamic weakening at high, seismic slip rates as already discussed. Rupture dynamics can also play an important role. For example, when dynamic rupture propagates along a fault between two different materials in the so-called bimaterial problem, slip creates dramatic and rapid variations in normal stress behind the rupture tip and, in particular, significant reduction in normal stress in certain directions, a form of enhanced dynamic weakening (Andrews and Ben-Zion, 1997; Cochard and Rice, 2000; Ben-Zion, 2001; Rice et al., 2001; Ampuero and Ben-Zion, 2008). Such bimaterial situations are common on natural faults which often separate either different rock materials due to long history of sliding or materials with different properties due to damage (e.g., Dor et al., 2006; Schmalzle et al., 2006). Other dynamics-related mechanisms proposed include acoustic fluidization (Melosh, 1979, 1996) and elastohydrodynamic lubrication (Brodsky and Kanamori, 2001).

The shear resistance is often interpreted and reported in terms of a friction coefficient, the ratio of the applied shear stress to the effective normal stress (i.e., the normal stress minus the pore fluid pressure) applied. The outcomes are especially challenging to interpret in the presence of fluids, when the effects of the local pore fluid pressure - evolving within the sample and often impossible to measure on the relevant temporal and spatial scales - are often reported as the effective friction coefficient. This complexity highlights the need for systematic interpretation of the laboratory results through a robust constitutive framework.

Rate-and-state framework for the combined response

Fortunately, one of the most significant results in our developing understanding of the shear resistance of the fault cores - the so-called rate-and-state friction formulations - provides a powerful framework for incorporating the wealth of the accumulating laboratory findings (Figure V-MMC1.1). In the rate-and-state formulations, the shear resistance depends on the slip rate (also called slip velocity) and state variables. The standard rate-and-state formulations in common use (Dieterich, 1979ab; Rice and Ruina, 1983; Ruina, 1983; Tullis and Weeks, 1986; Scholtz, 1990; Blanpied et al., 1991, 1995; Beeler et al., 1994) were developed for low slip rates of $10^{-7}$ to $10^{-4}$ m/s, initially for slip along macroscopically flat but microscopically rough interfaces; then they were shown to generally hold for the (dry or room humidity)
**Figure V-MMC1.1. Main features of standard rate-and-state friction** which is the basis for many SEAS modeling efforts. (Top) The basic concept of static-dynamic friction has been extended in two main ways: to slip-velocity-weakening formulations and slip-weakening formulations. (“Slip rate” and “slip velocity” are often used interchangeably to refer to the rate of change of slip with time.) (Middle) More detailed laboratory experiments at slow slip rates have established that the two dependencies co-exist, leading to rate-and-state friction laws where the state variable is used to capture transient slip-dependent responses. (Bottom) In the so-called steady state, established by sufficient slip at a given slip rate, friction is rate-dependent. Velocity-strengthening friction results in aseismic slip under slow loading. On velocity-weakening interfaces, slip spontaneously accelerates over regions larger than a critical size, often called “nucleation size,” leading to seismic events and stick-slip. The nucleation size, in part, is inversely proportional to effective normal stress, increasing with increasing pore fluid pressure; hence variations in pore fluid pressure can change the stability of rate-and-state faults. Adapted from Marone (1998), Lapusta and Barbot (2012).
granular shear layers (Dieterich, 1981; Marone, 1998ab). The standard formulations encapsulate the experimental findings of (i) nearly-constant shear resistance for sufficiently large slip at a constant slip rate, (ii) an abrupt change of the shear resistance in response to an abrupt change in the slip rate, with the same sign of the change, called the direct effect, and (iii) the subsequent evolution of the shear resistance with slip at the new slip rate towards another near-constant level of shear resistance, called the evolution effect.

The formulations incorporate a state variable to capture the evolution effect; the state variable reaches a steady-state value after sufficient slip at a given slip rate, quantified through a parameter called the state evolution distance or characteristic slip. Both behaviors have been observed in laboratory experiments, with the velocity strengthening favored by low confinement, elevated temperatures, and certain mineral compositions (Marone and Scholz, 1988, Moore and Rymer, 2007; Lockner et al., 2011). Stability analyses of frictional sliding (Rice and Ruina, 1983; Dieterich, 1992; Rice et al., 2001; Rubin and Ampuero, 2005; Ampuero and Rubin, 2008) indicate that interfaces with velocity-strengthening properties are stable, i.e., tend to slip with the long-term loading slip rate in the absence of transient perturbations in loading. Velocity-weakening interfaces have an important length scale, often called nucleation size (Figure V-MMC1.1); regions larger than the nucleation size experience sequences of locked periods and earthquake ruptures, often called stick-slip, but aseismic slip can occur over portions of the interface that are smaller than the nucleation size. Indeed, during an overall locked period, the nucleation of the next dynamic rupture occurs by aseismic slip that eventually reaches the nucleation size and becomes inertially controlled, hence the name for the instability length scale.

The standard rate-and-state formulations, initially completely empirical, have been experimentally and (partially) theoretically explained as the combination of (i) the rate-strengthening nature of the local, microscopic frictional shear resistance, (ii) the gradual, rate-dependent dilatancy of the shearing layer, and (iii) the associated evolution in the area of the microscopic frictional contacts captured by the state variable (Brechet and Estrin, 1994; Heslot et al., 1994; Estrin and Brechet, 1996; Nakatani, 2001; Rice et al., 2001; Sleep, 2005; Baumberger and Caroli, 2006; Beeler et al., 2007; Bar-Sinai et al., 2014; Ikari et al., 2016). For an abrupt increase in the slip rate, the configuration of particles in the shearing layer cannot instantaneously change, but the local shear resistance of microscopic contacts instantaneously increases, as would be expected for any strain-rate-hardening failure process, either viscous or plastic, at the creeping asperity contacts. As the slip proceeds at the higher (or lower) slip rate, the shearing layer gradually dilates, resulting in the gradual decrease in the overall contact area and hence gradual reduction of the overall shear resistance. After sufficient slip that scales with the characteristic slip distance, the dilation saturates. Hence velocity weakening vs. velocity strengthening in the steady state is determined by the competition between (i) the increasing average shear resistance, per unit area, of the local friction contacts for higher slip rates and (ii) the decreasing total contact area due to the dilation.

Overall, the standard rate-and-state formulations already capture, at least conceptually, the interaction between shear resistance, dilatation/compaction of the shearing layer, and the evolution of the local friction contacts. The versatility of the formulation allows to reproduce many key observations about the earthquake source, including: shallow and deep stably creeping fault regions through velocity-strengthening friction due to low confinement and high temperature, respectively; gradual aseismic nucleation and sequences of earthquakes at the intermediate, seismogenic depth through velocity-weakening friction; fault healing in the interseismic period that naturally arises in the formulation; and a range of other phenomena, including the rate of decay in aftershock sequences, scaling properties of
repeating earthquakes, and spontaneous aseismic transients (Dieterich, 1978, 1981; Rice and Ruina, 1983; Ruina, 1983; Tullis and Weeks, 1986; Scholtz, 1990; Blanpied et al., 1991, 1995; Dieterich, 1994; Ben-Zion and Rice, 1997; Marone, 1998; Lapusta et al., 2000; Rice et al., 2001; Rubin and Ampuero, 2005; Dieterich, 2007; Liu and Rice, 2007; Ampuero and Rubin, 2008; Chen and Lapusta, 2009; Kaneko et al., 2010; Barbot et al., 2012).

The rate-and-state friction formulations typically express the shear resistance as the product of a friction coefficient and effective normal stress, which is the difference between the fault normal compressive stress and the pore fluid pressure. Indeed, shear resistance that is linearly proportional to normal stress is a hallmark of physical processes collectively called “friction.” However, such formulations break down in cases of rapidly varying normal stress; experimental results show shear resistance then does not vary completely in sync with the normal stress but rather experiences evolutionary effects, which are often described using the same state variable as in the standard rate-and-state formulations (Linker and Dieterich, 1992, Prakash and Clifton, 1992, Richardson and Marone, 1999). This extended formulation is crucial for well-posed problems and meaningful solutions in a number of situations. One of them is dynamic rupture propagating along a fault between two different materials - so called bimaterial problem - in which shear creates dramatic - and rapid - variations in normal stress behind the rupture tip (Andrews and Ben-Zion, 1997; Cochard and Rice, 2000; Ben-Zion, 2001; Rice et al., 2001; Ampuero and Ben-Zion, 2008). Such bimaterial situations are common on natural faults which often separate either different rock materials due to long history of sliding or materials with different properties due to damage (Dor et al., 2006; MMC1.2).

Self-consistent extensions of the standard rate-and-state formulations have successfully captured several laboratory and field observations related to fluid effects. When the underlying, experimentally constrained dilation/compaction evolution is coupled, through poroelasticity, with pore pressure changes, the formulation can explain nontrivial experimental findings on shear strength evolution in the presence of fluids (Segall and Rice, 1995) as well as several features of relatively recently discovered subduction-zone slow slip transients (Liu and Rubin, 2010; Segall et al., 2010). In combination with a representation for the shear-heating-induced temperature evolution and the associated changes either in the steady-state friction resistance or in the fluid pore pressure, the rate-and-state formulations have been successful in capturing the enhanced dynamic weakening effects observed in the laboratory and the low-heat, low-stress, and often low-seismicity behavior of mature fault segments (Rice, 2006; Noda et al., 2009; Schmitt et al., 2015; Jiang and Lapusta, 2016).

Hence building on the rate-and-state formulation to incorporate other known effects is a promising way forward. The most advanced current formulations incorporate at least three state variables, in addition to slip rate: the standard rate-and-state one, pore fluid pressure, and temperature (Bizzarri and Cocco, 2006; Noda and Lapusta, 2010; Schmitt et al., 2015). To incorporate other effects, one may need to develop different evolution equations for the state variables, modify the functional form of the steady-state dependencies, or add additional, physically or chemically justified, state variables.

Note that cumulative slip is not an appropriate state variable, as the system can never return to its previous states, e.g., lower slips. The effective dependence of shear resistance on slip observed in some experiments is increasingly understood as the apparent effect, due to the dependence on other variables (Bizzarri and Cocco, 2003; Rice, 2006; Lapusta and Liu, 2009; Rubino et al., 2017). For example, linear slip-weakening friction in which the shear resistance drops from the peak value to the dynamic value linearly over slip distance called Dc and then stays constant, is the simplest way to regularize the rupture tip, and hence it is a popular choice for simulating dynamic rupture (MMC2.1). Indeed, dynamic rupture experiments show that the friction evolution can be appear to be very close to linear slip weakening, but
the parameters, such as $D_c$, change for different experimental conditions, indicating that $D_c$ is not a material property but, rather, an effective quantity that depends on the rate-and-state effects, additional weakening, and dynamics of the process (Rubino et al., 2017).

**Laboratory experiments on realistic layers under realistic conditions, with fluids**

Recent developments in laboratory techniques and granular-scale modeling provide a unique opportunity to test and develop an improved set of physically-based friction constitutive laws. Laboratory experiments are now capable of capturing the evolving response of fault gouge saturated by fluids under conditions relevant to the seismogenic faults (Chester, 1995; Niemeijer et al., 2008; Bartlow et al., 2012; Faulkner et al., 2018). Focus is needed on the role of wear and fault gouge evolution, including mineral transformation and the development of clay coatings on shear fabrics that develop in-situ and can play a major role in fault weakness (Schleicher et al., 2006, 2010; Collettini et al., 2009; Niemeijer et al., 2010b, Carpenter et al., 2011, 2012, 2015; Collettini et al., 2011; Warr et al., 2014; Wojatschke et al., 2016). Fluid-induced changes in fault zone effective stress are also a root cause of induced seismicity, via waste-water disposal and fluid injection. Laboratory experiments are needed to understand the evolution of granular dilation and compaction in such cases. Fluid effects and their impact on empirical parameters need to be more fully incorporated in the constitutive framework for the fault core (e.g. Segall and Rice, 1995; Faulkner et al., 2018). Evolution of shear zone thickness and degree of localization also need to be accounted for (e.g., Elbanna and Carlson, 2014; Platt et al., 2014). More studies are also needed on the existence of the direct effect at high slip rates as well as more fundamental approaches to formulating state evolution laws, with the latter affecting the nucleation size of earthquake ruptures and healing (Tullis and Goldsby, 2003; Rubin and Ampuero, 2005; Ampuero and Rubin, 2008; Bhattacharya et al., 2015; 2017).

**Robust representations of healing, including chemical effects**

The strength recovery – so called fault healing – is incorporated in the standard rate-and-state formulations and its extensions, but that does not take into account the full range of potential heating mechanisms, such as cementation and porosity loss, which may depend on fault maturity and stress/temperature conditions. The existing studies show that shear stress has a major impact on frictional healing rates (Karner and Marone, 1998, 2001; Ryan et al., 2018), however the origin of these effects are poorly understood. Fault healing also provides an opportunity to directly compare field and laboratory studies, vis-à-vis seismic and geodetic estimates of the rate of fault healing (Scholz et al., 1986; Vidale et al., 1994; Marone et al., 1995; Tadokoro and Ando, 2002; Li et al, 2003; Peng et al, 2005; McLaskey et al, 2014), which can be used to test the derived laboratory-based friction constitutive laws.

**Thermal pressurization of pore fluids vs. dilatant hardening**

The thermal pressurization of pore fluids within the shearing layer has the potential to be the dominating mechanism of dynamic weakening for mature faults (Sibson, 1973; Lachenbruch, 1980; Mase and Smith, 1985; Rice, 2006). Yet most theories of thermal pressurization do not account for local fault non-planarity that could cause dilation and the competing effect of the increasing pore fluid pressure due to increasing pore space, the process called “dilatant hardening” It is challenging to conduct laboratory experiments at seismic slip rates of the order of 1 m/s and realistically high confining pressure while controlling the pore fluid pressure and capturing the evolution of dilatancy/compaction. Obtaining measurements of temperature evolution at the sliding surface or imaging the heterogeneous pore pressure field in the sample interior is not possible using current experimental capabilities. High fidelity computational models, informed by precise measurements of elastic, thermal and hydraulic properties of the tested specimen
and constrained by the macroscopic lab measurements, are needed for the robust inversion of the evolution of compaction and dilation in laboratory experiments at seismic slip rates.

**Modeling at the scale of the shearing layer**

Continuum-based modeling: Evolution of the shearing-layer resistance emerge from a plethora of feedback mechanisms between mechanical, hydraulic, thermal, and chemical driving forces. Theoretical models for fault zone resistance have been largely limited to the 1D approximation, often accounting for localization and porosity changes (Segall and Rice, 1995; Sleep et al., 2000; Daub and Carlson, 2008). Localization and shear heating in a model of fluid-saturated finite-width fault zone have been studied (Rudnicki and Rice, 1975; Rice et al., 2014; Platt et al., 2014). Thermal decomposition and chemical weakening has also been considered in some studies (e.g., Platt et al. 2015). Complex strain localization modes observed at lab/field scales and the spatially heterogeneous nature of the gouge layer suggest that further modeling is needed in which more complex feedback pathways between the different physical mechanisms are taken into consideration. Multi-field continuum computational formulations of the coupled thermo-hydro-mechanical problem exist and have been used in the engineering and geomechanics community at least in the quasi-static limit, which would be applicable for slow slips and earthquake nucleation. Techniques such as stabilized finite element methods (White and Borja, 2008), variational multiscale methods (Hughes et al., 1998), phase field models (Miehe and Mauthe, 2016), and mixed-formulation finite element (Phillips and Wheeler, 2007) may be helpful as well. Existing community platforms, such as MOOSE from Idaho National Lab, may be particularly useful for such modeling, as they have been specifically developed for multiphysics applications (Ma and Elbanna, 2018).

Discrete element modeling (DEM) with micromechanics and fluids: Rigorous micro-mechanical modeling coupled with statistical physics is needed to formulate the thermodynamically consistent state evolution laws in the commonly used rate-and-state formulations and their parametric dependence on slip rate, work rate, and/or stress rate. Recent developments in the geomechanics community (Nezami et al., 2004; Zhao et al., 2006; Andrade et al. 2012) show significant promise in developing DEM with realistic particle shapes. Coupling such DEM with extended finite element (XFEM) approaches (Oden and Duarte, 1997) on the particle level will allow to incorporate particle fragmentation and hence to model fault fabric evolution and localizations. To understand the rate sensitivity of the granular shear layer, more realistic inter-particle friction and contact-scale physics need to be explored, such as shear heating and feedback of temperature rise on local friction and particle interaction. Such combined DEM/XFEM simulations (and other microscopic models), coupled with statistical physics, will open new opportunities for rigorous formulations of rate and state friction and would constrain, validate, and supplement some on-going efforts in this direction (Nagata et al., 2012; Lieou et al., 2014, 2016; Hulikal et al. 2018; Ferdowsi et al., 2015; Kothari and Elbanna, 2017; Chen et al., 2017; Ma and Elbanna, 2018). DEM can be coupled with fluids, to resolve chemical reactions and alterations, swelling, mineral precipitation and dissolution, and effects of shear heating. Such models partially exist in the geomechanics community and are applied extensively in petrology as well as in CO2 sequestration applications (Sweijen et al; 2016; Bakhshian and Sahimi, 2017). These computational tools may give insights into formulating more accurate rate and state laws in the presence of fluids.

**Key future goals**

- Conduct laboratory experiments and modeling to capture the evolving response of granular fault-core layers permeated by fluids under realistic conditions, including (i) particle distribution and mineralogy inferred for natural faults, (ii) strain rates spanning from near-locked to dynamic sliding conditions, and (iii) a range of temperature/confinement/fluid pressure conditions representative of natural faults for the shallow, mid-seismogenic, and below-seismogenic depths;
Design laboratory experiments to enable insightful modeling;
Determine dominant mechanisms controlling friction at high slip rates; in particular, assess the relative importance of thermal pressurization of pore fluids and dilatant hardening.

**MMC1.2 Off-fault damage accumulation and healing**

Geological and geophysical observations of fault zones reveal that the highly localized shearing layers are surrounded by damage zones, e.g., regions of rocks with brittle fracturing on a range of different scales (Figure II-MMC1), from macroscopic and microscopic fracturing to grain-scale deformation (Chester and Logan, 1986; Chester et al., 1993; Vermilye and Scholz, 1998; Dor et al., 2006; Mitchell and Faulkner, 2009; Faulkner et al., 2011; Savage and Brodsky, 2011). The damage zone, while altered, does not accommodate much of the shear strain (Flinn, 1977; Sibson, 1977; Faulkner et al., 2010). The damage zones affect the local stresses (Faulkner et al., 2006), permeability structure (Caine et al., 1996; Sibson, 1996), and mechanical properties (Fialko et al., 2002; Hamiel and Fialko, 2007; Heap et al., 2010; Yang et al., 2015) of the fault zone, and act as a dissipative energy sink during seismic events (Rice et al., 2005; Shipton et al., 2006). Thus, the mechanical behavior, geometry, and evolution of damaged rocks have important effects on the dynamic rupture and overall deformation of fault zones. The response and evolution of damaged rocks are also important in structural geology, rock engineering, and oil exploration communities (e.g. Paterson and Wong, 2005; Zhang and Zhao, 2014).

**Constraints from field studies**

The width of the damage zones has been estimated from geological studies of exhumed faults as well as from seismology, geodesy and geomorphology. It is found to be depth-dependent (Figure II-MMC1). Geological studies have shown that the damage zones display an exponential decay of fracture density with distance from the fault core (Anders and Wiltshire, 1994; Vermilye and Scholz, 1998; Wilson et al., 2003; Mitchell and Faulkner, 2009) and that their width, defined by the distance over which the level of damage returns to background values, scales with the displacement on a fault, even down to very low displacements (Mitchell and Faulkner, 2009; Faulkner et al., 2011; Savage and Brodsky, 2011). These field observations of relatively deeply exhumed fault zones appear to show a saturation of damage zone widths at several hundred metres. Additional observations of damage zone width from seismology, geodesy and geomorphology show much greater extents than those from field studies (Fialko et al., 2002; Cochran et al., 2009; Jolivet et al., 2009; Wechsler et al., 2009; Allam and Ben-Zion, 2012; Lindsey et al., 2014). For example, the width of the zone with reduced elastic properties (around 50% reduction in the shear modulus) surrounding the Calico Fault is on the order of kilometres. These larger damage zone widths might be restricted to the near-surface deformation and so could be reconciled with smaller damage zone widths observed on exhumed faults using depth-dependent damage zone width, however more high-resolution studies of the mechanical structure of active damage zones are needed to better constrain the depth dependence of damage. Note that suppression of damage at greater depth is consistent with the larger values of overburden pressure which discourages distributed brittle failure according to experimental and theoretical studies discussed below.

The significance of these field observations for the processes responsible for damage zone formation have been addressed by many authors, and include fault geometrical irregularities/roughness, linking of structures, and high stresses at the tip of dynamic rupture (Pollard and Segall, 1987; Cowie and Scholz, 1992; Chester and Chester, 2000; Shipton and Cowie, 2003; Rice et al., 2005; Faulkner et al., 2011).
These processes may be acting together; e.g., the dynamic damage generation due to rupture tip can be modulated and enhanced by fault geometry/roughness, as shown by numerical studies (Dieterich and Smith, 2009; Fang and Dunham, 2013; Shi and Day, 2013).

Geodetic and seismological observations (Vidale and Li, 2003; Li et al., 2006; Brenguier et al., 2008; Kelly et al., 2013) indicate that, over the repeated processes of dynamic rupture propagation and interseismic periods, the properties of the damage zones can evolve significantly, and over a range of different timescales (Vidale and Li, 2003; Brenguier et al., 2008; Kelly et al., 2013). The increase in damage due to an earthquake rupture is manifest in decrease of seismic wave propagation speeds, increase in attenuation, and an enhanced geodetic strain. These parameters have been found to recover with time after the earthquake, typically over months to years, for the top few kilometres of the crust that the geodetic and seismological studies can resolve. Healing of damage at greater depth may be controlled by diffusional crack healing and hydrothermal precipitation (Smith and Evans, 1984; Moore et al., 1994) and may occur over a range of timescales from seconds to years.

Constraints from laboratory studies

Dependence of response on strain rate: Throughout the coseismic and interseismic periods, off-fault rocks are exposed to a broad range of loading rates. In the last 50 years, loading apparatuses have been designed to test rocks over a wide range of strain rates, from $10^{-6}$ to $10^6$ s$^{-1}$. Between strain rates of $10^{-1}$ and $10^4$ s$^{-1}$, there is significant increase in rocks failure strength and inelastic yield strength with strain rate (Grady, 1998; Doan and Gary, 2009; Kimberley et al., 2013; Zhang and Zhao, 2014; Aben et al., 2017) (Figure II-MM1C1.2). Below and above that range, the sensitivity to strain rate is very small. Note that the diagram is based mostly on high strain-rate experiments performed under uniaxial compressive loading. In the fault zones, damage is more likely to occur on the extensional side of the fault. Dynamic tests in the tensile regime are more challenging, but have been performed in the civil engineering community (Zhang and Zhao, 2014 and references therein), suggesting that tensile strength also increases with loading rate, and may depend on confining pressure (Fialko and Rubin, 1997). Designing and performing more experiments in the tensile regime at high strain rates is essential in order to constrain the behavior of damage materials in fault zones, including the strain-rate threshold for pulverization (Aben et al., 2017) and the potential effect of strain rate on elastic moduli before damage formation (Zhang and Zhao, 2014 and references therein).

Effects of cyclic loading/unloading: Fault zones rocks are exposed to cycles of loading and unloading, in which they may exceed temporarily the inelastic yielding stress state. An increasing-amplitude quasi-static cyclic loading affects elastic properties of basalts, sandstones, granite, and evaporites (Heap et al., 2010; Trippetta et al., 2013; and Yang et al., 2015). More damage was accumulated with increasing cycle number, and, except for the evaporites, the Young’s modulus showed consistent decrease with cycle number, with total decrease of 10–20%. All rocks showed significant increase of the Poisson’s ratio. Before failure, all the samples showed hardening, with increasing yield strength with cycle number. Because there were no hold periods between loading cycles, these experiments could not capture the effect of healing.

Healing: Over time, and with elevated temperature and fluids, microcracks are expected to close, heal, and become sealed by diffusional processes (e.g. Smith and Evans, 1984; Hickman and Evans, 1987), dissolution-precipitation processes (e.g. Hickman and Evans, 1992; Richard et al., 2015), and mechanical processes (Brantut, 2015). The times and spatial scales associated with these processes control how fast damaged rocks regain strength and stiffness and drive fluid flow around faults. For example, experiments
on limestone under water-saturated conditions samples (Brantut, 2015) showed that, during a static hold period of two days, the seismic wave speeds recovered by about 5% due to crack closure. Clearly, more experiments are needed to understand better the healing process in damaged rocks.

Spatial and temporal evolution of damage in experiments: Several techniques have been developed to monitor the temporal and spatial evolution of damage during loading experiments, and can help relating the evolution of damage to macroscopic behavior of lab samples. Several studies have used compressive loading experiments with the microcrack analysis to derive empirical relations between the evolution in crack density and elastic properties (e.g. Hadley, 1976; Katz and Reches, 2002; Oda et al., 2002; Takemura and Oda, 2004; Tal et al., 2016). The locations and number of acoustic emission (AE) events at different loading states show how damage evolves and localizes during the failure process (e.g. Lockner et al., 1992). Moreover, using AE events, (Goebel et al., 2017) studied the effect of fault geometry on the distribution of damage. Using a combination of image correlation technique with photography, (Tal et al., 2016) observed and quantified the production of micro fracturing during uniaxial loading of Carrara marble. The experiment showed that the damage model of Ashby and Sammis, 1990) significantly underestimated the damage that the rock could sustain before peak stress, perhaps owing to the influence of weak grain boundaries on the damage production. Photography and AE method can also be combined (e.g. Li and Einstein, 2017) to gain more understanding on damage processes. Photography requires a free surface to observe and provides only 2-D distributions. Recent advances in triaxial compression deformation apparatus design, dynamic X-ray microtomography imaging, data analysis techniques, and digital volume correlation analysis provide opportunities to monitor the in situ 4-D distribution of developing damage within rocks (Renard et al., 2018a,b) (Figure II-MMC1.2). All the experiment above were performed under small loading rate, ultra high speed can be used to monitor damage production experiments under high loading rates.

**Constitutive relations used for damage zones**

Large-scale numerical models of seismogenic faults cannot explicitly include the small-scale fractures in the damage zone and need to adopt a continuum-level constitutive descriptions. There has been considerable effort in the earthquake modeling community to study the effects of damage zones on the earthquake ruptures, with various constitutive ways of representing them.

The most straightforward way to account for some damage-zone effects is by modeling the damage as an elastic lower-velocity zone around the fault (e.g. Harris and Day, 1997; Kaneko et al., 2011; Huang et al., 2014; 2016). These studies find that reflected and head waves inside the lower-velocity damage zone interact with the earthquake rupture, modulate rupture properties such as rupture speed and slip rates, and create complex slipping modes such as multiple pulses of slip during a single dynamic event. However, such models do not capture the dissipation of energy by inelastic deformation in the damage zones, as well as the temporal and spatial changes in the rheology and extent of the damage zone.

Other models have used off-fault Mohr–Coulomb or Drucker-Prager plastic yielding to explore the effect of off-fault damage formation on the rupture process, as well as the extent of inelastic deformation (e.g. Andrews, 2005; Ben-Zion and Shi, 2005; Templeton and Rice, 2008; Ma, 2008; Dunham et al., 2011ab; Kaneko and Fialko, 2011; Gabriel et al., 2013; Shi and Day, 2013; Johri et al., 2014; Ampuero and Mao, 2017; Erickson et al., 2017; Roten et al., 2017; Wollherr et al., 2018). These models show that the plastic dissipation in the damage zones reduces stresses and slip velocities at the rupture front and leads to a coseismic slip deficit in the shallow sub-surface, where the accumulation of plastic strain is enhanced by the low overall compression. The main limitation of such damage-zone representations is that they do not model the complete feedback between the rupture and off-fault damage because, as yielded regions are
unloaded, the elastic properties recover immediately to their initial values, which is equivalent to rapid and complete healing.

Damage rheology formulations connect the evolution of elastic moduli with changes of crack density through damage variables in a continuum description. They have been used extensively to model fracturing in engineering materials and rocks (e.g. Kachanov, 1986; Krajcinovic, 1996; Allix and Hild, 2002). A thermodynamically based, nonlinear viscoelastic formulation with a scalar damage variable has been developed for evolving elastic properties of rocks in brittle rock deformation, including both degradation and healing, and validated using laboratory experiments (Lyakhovsky et al., 1997, 2011, Ben-Zion and Lyakhovsky, 2002; Hamiel et al., 2006,). The formulation has been extended to represent damage as a tensor variable that accounts for directions, numbers, and lengths of microcracks (Suzuki, 2012). Another type of damage rheology formulations aim to relate microcrack densities to macroscopic stress states by describing the micromechanical behavior of microcrack systems, using linear elastic fracture mechanics and idealized geometries of crack arrays (Brace and Bombolakis, 1963; Nemat-Nasser and Horii, 1982; Ashby and Sammis, 1990). Despite the approximations involved, the models give a good fit to laboratory rock failure observations (Ashby and Sammis, 1990). Extended microcrack models (Bhat et al., 2012) allow for more generalized stress states and include more realistic, strain-rate-dependent fracture toughness, making the constitutive response sensitive to strain rate, which is particularly important in earthquake-related processes. The extended formulation was experimentally verified by predicting the increase in failure strength of marble with loading rate for over 9 orders of strain-rate magnitude and included in dynamic ruptures simulations (Perol and Bhat, 2016; Thomas and Bhat, 2018). The main limitation of current damage-rheology constitutive relations is that they consider isolated representative volumes and have difficulty describing intensely interacting features or predicting localization instabilities (Tal et al., 2016).

Need for modeling at the scale of damage

The proposed constitutive relations for damage zones should be systematically assessed, constrained, and further developed. For rocks with fractures at scales of microns to few centimeters, the required input can be provided by lab experiments performed for different rock types at different stress conditions, temperatures, with and without fluids. A major challenge is how to combine the response at the lab scale with the effects of larger scales fractures, as well as other heterogeneities. Note that, in the original definition of crack density (Walsh, 1965), the larger fractures dominate. Simulations of earthquakes sequences and aseismic slip over a 100-km large fault segment require constitutive relations that capture effective properties and their evolution over meters to tens of meters. This may be achieved by effective medium models (e.g. Kachanov, 1992 and references therein) or by numerical techniques (e.g. Griffith et al., 2009ab). Since the problem of the discontinuous rock mass response is important in the field of rock engineering, many numerical techniques have been developed to handle discontinuities (see review by Jing, 2003), and it is worth examining whether these techniques can be used for damage volumes in fault zones. Nevertheless, these models should be validated at different scales by experimental, geodetic, and geophysical studies.

Key future goals

● Elucidate constitutive relations for damage creation during earthquake ruptures: experiments and modeling of rapid (high-strain-rate) damage to ascertain the strain-rate dependence, damage patterns, and the associated elastic moduli evolutions under representative stress/temperature/fluid conditions for the shallow, mid-seismogenic, and bottom-seismogenic depths;
● Investigate damage healing in the interseismic period: mechanical vs. chemical;
Capture all findings in constitutive relations for damage zones on 0.1-100 meter scale needed for large-scale models: beyond the laboratory scale.

MMC1.3 Rheology at and below the brittle-ductile transition and fault loading

A highly non-uniform distribution of earthquakes as a function of depth in the lithosphere (e.g., Meissner and Strehlau, 1982; Sibson, 1982; Pacheco et al., 1993; Hyndman et al., 1997; Nazareth and Hauksson, 2004) indicates a transition from predominantly seismic to predominantly aseismic deformation mechanisms. This so-called brittle-ductile transition occurs due to increasing temperature, reflects an increasing role of viscous deformation (Chester, 1995; Chester et al., 2004; Noda and Shimamoto, 2010), and has two main, potentially related manifestations: (i) the upper elastically deforming brittle (e.g., fracture-prone) rocks transition to ductile, creep-prone, visco-elastic-plastic behavior (Hobbs et al., 1986; Hirth and Kohlstedt, 1995, 1996, 2004, 2015; Niemeijer et al., 2002, 2008, 2010a; Muhuri et al., 2003; Tenthorey et al., 2003; Behr and Platt, 2014; Melosh et al., 2018) and (ii) the upper locked fault regions transition to more constantly slipping or deforming fault extensions (Romanowicz and Ruff 2002, Rolandone et al., 2004; Leonard 2010, Bruhat and Segall, 2016; Denolle et al., 2016, Jiang and Fialko, 2016; Jiang and Lapusta, 2016, 2017; Weng and Yang, 2017). (Figure II-MMC1.3) The shear deformation below the seismogenic zone is envisaged to be spread across a greater width (Sibson, 1977) and this is generally supported by observation (Melosh et al., 2018). The brittle-ductile transition is important for understanding the earthquake physics and fault behavior, due to its crucial role in fault loading - and hence earthquake rupture nucleation and propagation - and in limiting the downward extent of earthquake ruptures (Das and Scholz, 1983; Lapusta and Rice, 2003; Jiang and Fialko, 2016; Figure II-MMC3.1-2). The lower creeping bulk and fault regions may also contribute to, or even sometimes dominate, earthquake interactions (Benioff, 1951; Freed and Lin, 2001; Perfettini and Avouac, 2004).

While significant progress has been made over the last decade in our understanding of the transition between the seismogenic and aseismic parts of the lithosphere, and mechanisms by which seismically active faults are loaded in the interseismic period, a number of outstanding questions remain. Addressing these questions will require a combination of focused seismologic, geodetic, geologic etc. observations of natural faults, laboratory experiments, and numerical modeling targeting the fault and host rock behavior over the representative range of pressures, temperatures, compositions, and other relevant factors.

Importance for fault loading, fault “roots”

The more ductile rocks below the seismogenic faults respond to plate motion with relatively continuous deformation (Hetland and Hager 2005; Fay and Humphreys 2006; Takeuchi and Fialko 2012; Hearn and Thatcher, 2015), significantly affecting the loading of the shallower, seismogenic portions of the fault, and hence the stress conditions for earthquake nucleation and propagation. Indeed, given the immediate proximity of this deeper deformation to the seismogenic portion of the fault, it could be the dominating factor in fault loading.

That is why many current models of earthquake sequences in elastic bulk assume a constant slip-velocity boundary condition applied on a deeper continuation of a fault (“bottom-driven” models; e.g., Tse and Rice 1986; Lapusta et al. 2000; Lindsey and Fialko 2016). Bottom-driven models assume that shear deformation (or shear strain) is permanently localized right below seismogenic faults, possibly in the form of ductile shear zones, and broadens with depth modestly enough to result in stressing on the shallow
portions similar to the slipping interface. Strain localization in the ductile regime can result from a variety of mechanisms, including fabric development (Lister and Williams, 1979; White et al., 1980; Montesi, 2013), dynamic recrystallization (Tullis and Yund 1985; Rutter, 1999; Montesi and Hirth, 2003), thermo-mechanical coupling (Yuen et al., 1978; Brun and Cobbold, 1980; Takeuchi and Fialko, 2012; Jaquet and Schmalholz, 2018), and mineral alteration (Brodie, 1980; Gueydan et al., 2003). Indeed, there are multiple observations that support such localized deeper roots, including deep tectonic tremor (Shelly, 2010ab), postseismic afterslip (Reilinger et al., 2000), and studies of exhumed faults (Cole et al., 2007).

The ductile “roots” of faults can be produced (rather than postulated) by models which incorporate realistic visco-elastic-plastic rheologies and which are “side-driven”, with the plate velocities applied to the lateral sides of the computational domain (Sobolev et al., 2005; Takeuchi and Fialko, 2012; 2013), resulting in structures that share many similarities with geologic observations of mylonite zones (i.e., narrow regions of intense deformation) in the deeply exhumed terrains (e.g., Sibson 1977; Berthe et al. 1979; Poirier 1980; Bürgmann and Dresen 2008). This provides some justification for the use of bottom-driven models and an explanation for the geodetically observed interseismic strain rates - i.e., strain rates in the periods between large earthquakes - in the upper crust around active seismogenic faults (e.g. Thatcher 1975; Hearn et al. 2002; Fialko, 2006).

Both stable frictional slip or localized ductile shear on a deep aseismic fault root are expected to result in an elevated stressing rate at the bottom of a “locked” (seismogenic) section of a fault, making it a preferred site for the concentrated microseismicity and nucleation of large system-size ruptures (Lapusta and Rice, 2003; Barbot et al., 2012; Lindsey and Fialko, 2016). Even if the shear zone is broader, the stressing effects may be similar (Allison and Dunham, 2018), although heterogeneous structures may give rise to local variations that may be different between the more localized and more distributed deformation. Furthermore, the partitioning between the bulk and more localized extensions coupled with their constitutive response may significantly affect the ability of earthquake ruptures to penetrate deeper (Jiang and Lapusta, 2016, 2017) as well as may affect earthquake interactions.

Hence it is important to resolve the outstanding question of how much of the deeper across-fault motion is accommodated by distributed ductile flow vs. slip/flow across permanently localized deeper fault extensions, especially for faults of different maturity (or cumulative slip). More realistic models that take into account the full coupling between the seismic and aseismic parts of the lithosphere, as well as the underlying mantle, are clearly warranted, with appropriate inelastic constitutive relations for the bulk and the localized fault “roots.”

Transition in bulk rheology vs. localized fault behavior

The two most commonly considered endmember models of the seismic/aseismic transition are (i) a transition from friction-dominated brittle failure of the rocks in the seismogenic layer to distributed ductile, viscoplastic flow in the aseismic substrate (Byerlee, 1968; Kirby, 1980; Chen and Molnar, 1983), and 2) a transition from velocity-weakening to velocity-strengthening friction (Tse and Rice, 1986; Marone and Scholz 1988; Blanpied et al., 1991, 1995; ) (section MMC1.1). Both models are based on experimental data, and appeal to pressure and/or temperature-activated changes with depth.

The two models are not mutually exclusive - for example, a weak frictional-slip interface may possibly extend into the deeper regions in which the bulk deformation is dominated by viscous creep. This may explain the observed micro-seismicity and tremor on deep extensions of active faults (e.g., Shelly, 2010ab; Inbal et al., 2015). Alternatively, the deep tremor and microseismicity may be associated with locally elevated strain rates due to ductile shear zones, especially in the presence of over-pressurized
pore fluids (Nadeau and Guilhem, 2009; Hirth and Beeler 2015; Skarbek and Rempel, 2016). There is ample geologic evidence for the presence of ductile shear zones beneath major crustal faults that can result due to variety of mechanisms (Poirier, 1980; Tullis and Yund 1985; Rutter, 1999; Montesi and Hirth, 2003; Norris and Cooper, 2003; Bürgmann and Dresen, 2008), thermo-mechanical coupling (Yuen et al., 1978; Brun and Cobbold, 1980; Takeuchi and Fialko, 2012; Jaquet and Schmalholz, 2018), and mineral alteration (Brodie, 1980; Gueydan et al., 2003). Indeed, fault constitutive relations that gradually transition from friction to flow have been proposed based on experimental studies of analog materials (Kawamoto and Shimamoto, 1997; Shimamoto and Noda, 2014).

Recent experimental observations (Mitchell et al., 2016; Khamrat et al., 2018) call into question the accepted paradigm of velocity-strengthening friction of shear zones at elevated temperatures relevant to the deeper fault roots, indicating instead that the properties may be of unstable, velocity-weakening nature. They warrant further studies of the mechanisms controlling the velocity dependence of friction over the range of pressures and temperatures corresponding to the seismic/aseismic transition in the Earth’s crust. Even if an accurate description, the velocity-strengthening friction may not be ultimately responsible for the arrest of earthquakes that propagate into the nominally aseismic substrate. In case of a highly localized slip interface, various dynamic weakening mechanisms may potentially overwhelm the rate-and-state effects (Jiang and Lapusta, 2016, 2017). If so, arrest of earthquake ruptures would imply either de-localization of seismic slip near the rupture front (thereby limiting the efficiency of thermally-activated weakening mechanisms and/or increasing the effective dissipated energy), or sufficiently low deviatoric stresses below the brittle-ductile transition (Fialko, 2015).

An important question is the role of pore fluid pressure and the proper form of the effective stress law toward the base of the seismogenic zone. It has been suggested (Scholz, 1990; Hirth and Beeler, 2015) that the appropriate form is \((\sigma - \alpha p)\), where \(\sigma\) is the fault normal compressive stress, \(p\) is the pore fluid pressure, \(\alpha = 1 - \frac{A_r}{A}\), and \(A_r\) and \(A\) are the real (microscopic) and apparent (macroscopic) area of contact, respectively, with \(\alpha\) being near 1 over most of the seismogenic depths but increasingly becoming smaller than 1 for higher compression and flow-like deformation at greater depth. On the other hand, some experimental data supports the Terzaghi effective stress under flow-like deformation (Noda and Takahashi, 2016). More experiments are required to clarify the proper effective stress law near the brittle-ductile transition.

**Constitutive relations for deeper ductile rocks**

The ductile deformation of rocks at pressure and temperature conditions corresponding to the mid- to lower crust and upper mantle is typically described using viscoplastic relations. Several experimental setups, with and without fluids, combined with advanced imaging methodologies such as X-ray diffraction, Raman spectroscopy, Scanning electron microscope (SEM), and Tunneling electron microscope (TEM), have been used to relate changes in rock microstructure to its mechanical response under the relevant temperature/stress conditions, and to differentiate between different deformation mechanisms depending on temperature, stress, strain rate, and grain size. The experimental studies have identified several relevant physical mechanisms, including dislocation creep, diffusion creep, pressure solution, and grain boundary sliding. Geodetic observations at different strain rates have been enabled by interseismic and postseismic measurements, with the accelerated postseismic deformation caused by earthquake-induced stress changes.

The experimental data and geodetic observations suggest that power-law rheology relating stress, strain rate, and grain size provides the best description (e.g., Carter and Tsenn, 1987; Hirth and Kohlstedt, 1995, 1996, 2004; Rybacki and Dresen 2004; Freed and Bürgmann, 2004; Takeuchi and Fialko, 2013; Masuti et
The power-law rheology is usually associated with dislocation creep, but may also apply to the grain-size sensitive diffusion creep, provided that the grain sizes evolve to a state of equilibrium between the dynamic recrystallization and the quasi-static grain growth. Such an equilibrium implies comparable contributions of dislocation and diffusion creep to the total deformation (De Bresser et al., 1998), so that a constitutive law for ductile flow with stress-sensitive grain size has the same stress exponent as that for dislocation creep (e.g., Montesi and Hirth, 2003). It is worth noting, however, that a transient response of a ductile shear zone governed by power-law creep to sudden stress perturbations may be hard to distinguish from that of fault slip governed by rate and state friction (e.g., Barbot et al., 2009).

As viscous deformation is strongly strain-rate strengthening, variations in strain rate produced by earthquakes in the seismogenic zone above may affect the dominant operative deformation mechanisms. One potential manifestation of this is a deepening of the region of aftershocks following an earthquake, as strain rates are elevated and the deformation response is more brittle, but a progressive shallowing with time as strain rates decay and viscous deformation becomes more prominent (Rolandone et al., 2004; Jiang and Lapusta, 2016). It remains an important challenge to further constrain the rate-dependent rheology of main rock-forming minerals across the deep seismic-aseismic transition to evaluate the temporal fluctuation of the depth limit of seismicity due to changes in rate-controlling deformation mechanisms and to explain deep crustal seismicity in some tectonic settings.

The ductile rock behavior can also be strongly influenced by fluid effects, both through hydrolytic weakening and chemical alteration. There is evidence that intra-crystalline water gets into silicate phases through microcracking (Kronenberg et al 1990), suggesting that dynamic stressing through the earthquake cycle may feedback into weakening of rocks near the base of the seismogenic zone. Fluids also strongly effect recrystallization, and the formation of frictionally weak hydrated phases including saponite and talc (Moore and Rymer, 2007).

Most of the proposed viscoplastic flow rules, including the power-law rheologies, are mineral-specific, and either based on a single physical mechanism or phenomenological in nature. Yet studies of exhumed faults indicate complex mineral structure of the bulk and potential operation of multiple deformation mechanisms (Bürgmann, 2018; Hamling et al., 2017; Dascher-Cousineau et al., 2018; Melosh et al., 2018; Rowe et al. 2018). Further laboratory and modeling studies are needed to develop suitable viscoplastic constitutive laws for such realistic structures, for the wide range of relevant strain rates (see also MMC1.4).

Key future challenges
- Determine structure and constitutive response of deep localized fault “roots,” friction vs. flow: frictional slip interfaces presumably grade into ductile shear zones below the seismogenic zone, but the details of the combined structure and its constitutive response are poorly understood;
- Construct bulk flow rules for multi-mineral rocks and multiple concurrent deformation mechanisms: building on power-law rheologies;
- Investigate fluid effects: changes in flow laws and the role of effective stress in the ductile roots of fault zones.

MMC1.4 Coupling deformation/damage and fluid effects

Fluids play an important role in many aspects of rock deformation and faulting, as already highlighted in section I and challenges MMC1.1-1.3. Here we review the main effects for both faulting and deformation of the surrounding crust. It is important to distinguish aqueous pore fluids, which reside in interstitial spaces within fault gouge as well as the bulk rock adjoining fault zones, from water that is either
chemically bound in minerals, or forms defects in silicates that promotes dislocation mobility, a phenomenon known as hydrolytic weakening.

**Fluid effects in faulting/damage zones: fluid-centric summary of related issues in MMC1.1-1.3**

Many models consider frictional fault strength to depend on the (Terzaghi) effective stress, which is the difference between the fault normal compressive stress and the pore fluid pressure. High pore pressures, and hence low effective stress, have been invoked to explain sub-horizontal thrust faults (Hubbert and Rubey, 1959), as well as slow slip events (SSE) found in many subduction zones. The famous field experiment at the Rangely oil field (Raleigh et al, 1976) is considered a confirmation of the effective stress principle.

Granular fault gouges are expected to exhibit inelastic dilatancy and compaction under shear, depending on stress state and past history of deformation. Depending on deformation rates and permeability among other factors, dilatancy can decrease fault zone pore pressure and thus exhibit a stabilizing effect; a possible mechanism for slow slip events (Liu and Rubin, 2010; Segall et al, 2010); compaction has the reverse effect. Whether or not such processes operate in the 25-40 km depth range in subduction zones remains unclear. Inclusion of such phenomena in fault zone models has been hampered by the lack of agreed upon constitutive laws; formulations from soils mechanics are largely rate independent. There are relevant laboratory experiments, although many on simple analog materials (e.g. glass bead packs). Experimental constraints at relevant conditions for earthquakes including grain fracturing, and fluid assisted cementation are required.

Rapid shearing in low permeability fault gouges is likely to lead to thermal pressurization of pore fluids and strong dynamic weakening. Thermal pressurization allows ruptures on smooth faults to propagate, in a pulse-like mode, at low levels of background ratios of shear to normal stress (Noda et al., 2009; Schmitt et al., 2015). Once generated, ruptures promoted by thermal pressurization can propagate into nominally stable fault areas (Noda and Lapusta, 2013; Jiang and Lapusta, 2016). Numerical models of this process are extremely challenging as they must resolve small spatial scale strength breakdown at the rupture tip; even more so if including flash heating. Yet such codes are needed to model earthquakes in heterogeneous environments; to understand what stops rupture, or alternatively kicks pulses into crack-like or super-shear ruptures. We also need to understand processes that can defeat thermal weakening – strong dilatancy or the possibility of hydraulic fracturing (mode I) near the rupture tip. This will require appropriate constitutive laws (see above) and well resolved numerical models.

Some laboratory experiments are strongly suggestive of thermal pressurization, although apparatus designed to test the phenomenon are very challenging to design. Thermal pressurization can be augmented by thermally induced devolatilization, which acts as an additional source of fluids (Brantut et al, 2010). There may be a trade-off between thermal pressurization and dilatancy-induced pore pressure decrease. The thickness of the strongly shearing fault core has a dominant effect on the rate of heat generation during fast slip. Many models have treated the active fault thickness as a parameter to be specified, rather than derived by fault physical processes. Some work has analyzed the active shear zone thickness as a competition between thermal pressurization, pore-pressure diffusion, and rate dependent frictional properties (Platt et al., 2014). Significant challenges will be to better understand the local scale physics and include them in larger scale rupture models (MMC1.1).

Fault zone pore fluids communicate with fluids in the host rock, potentially leading to a range of transient pore pressure behaviors. The flow is highly sensitive to the hydrogeologic architecture of faults, which is intrinsically linked to the spatially variable and evolving fault structure (MMC1.1-1.2). A major
observational challenge to the field is to robustly characterize both the in-situ permeability and storage of fault zones at the scales relevant for earthquake processes (Bense et al., 2016; Scibek et al., 2016). These scales are outside the range typically sampled by standard hydrogeological and reservoir techniques and thus specialized methods are necessary.

Hydraulic properties of near-fault rocks, notably permeability, can vary over many orders of magnitude depending on rock type, effective stress, and degree of fracturing. Samples from outcrop and drill cores have led to the conceptual model of a very low permeability fault core, surrounded by more highly permeable damage zones (Wibberley and Shimamoto, 2003; Refs) (Figure II-MMC1.4). In-situ observations could be very beneficial in constraining parameters in models, although spatial heterogeneity and possible time dependence, due to mineralization of pores and fractures, remain significant factors. Many models restrict fluid diffusion to the fault normal direction, although in some cases along-fault flow may be dominant.

Earthquakes also directly affect the hydrogeological structure. Fault rock damage (MMC1.2) can create or enhance permeability pathways that heal after an earthquake (Xue et al., 2013). As the multi-scale fracturing occurs during earthquake rupture, the newly created or re-opened cracks dynamically enhance the permeability of the surrounding medium, potentially affecting the thermal pressurization of the highly localized shear layers (MMC1.1), while the permeability is assumed to be constant in most current earthquake modeling (MMC2). Seismic waves can increase permeability even at great distance (Elkhoury et al., 2006; Manga et al., 2012). This coupling allows transient fluid pulsing that can in turn affect future earthquakes (Sibson, 1992; Fulton and Brodsky, 2016). The coupling also may create preferred hydraulic diffusivities in active fault zones that govern subsequent failure (Townend and Zoback, 2000; Xue et al., 2016).

The ductile roots of faults can also be strongly influenced by fluid effects, both through hydrolytic weakening and chemical alteration. An important question is the proper form of the effective stress law toward the base of the seismogenic zone (MMC1.3).

**Fluid effects in bulk rock deformation**

Poroelastic effects are not as commonly included in fault models, more due to computational limitations than to conviction that these effects are insignificant. The complexity arises due to the full coupling between the solid deformation and pore-fluid flow (Wang, 2000; Dunham and Rice, 2008; White and Borja, 2008; Jha and Juanes, 2014; Meng, 2017; Torberntsson et al., 2018): changes in pore-pressure act as sources of deformation, while volumetric strains act as sources of pore pressure. In addition to computational challenges, knowledge of appropriate poroelastic constants, such as permeability, are not well known, especially in cases where fractures are important. Recent experiments reveal that the coupling between deformation and pore-fluid diffusion exerts critical control over the development of fracture patterns as well as the rate of faulting and fault slip (French et al., 2016; French and Zhu, 2017).

The role of poroelasticity in fluid injection induced seismicity has been a growing interest in the last few years (Segall, 1989; Ellsworth, 2013; Segall and Lu, 2015; Goebel et al., 2017; Goebel and Brodsky, 2018). It is generally acknowledged that, for homogeneous media, the direct effective stress changes dominate in the near field, while poroelastic stress changes dominate in the far-field (Ge et al., 2009; Segall and Lu, 2015), although strongly heterogeneous permeability can alter the relative importance of these effects (Chang and Segall, 2016). At this point, it is not possible to point to definitive evidence of poroelastic effects dominating direct pore-pressure effects; a more complete suite of in situ observations is required. At the same time, robust and flexible codes that include fully coupled poro-elasticity - and
hydromechanical effects more generally - along with rate-and-state friction and other features of fault zones (MMC1.1-1.2) are needed to explore various hypotheses.

In situ permeability inferred from the space-time migration of induced earthquakes commonly exceeds that determined from compilations of other data (Manning and Ingebritsen, 1999). Whether this is a sign of poroelastic effects, propagating aseismic slip (Guglielmi, 2015), or other phenomena is presently unknown. This highlights both the challenge and opportunity of combining a well calibrated in situ-fluid injection experiment with appropriate numerical modeling tools (sections III-IN3 and III-IN4).

**Key future goals**

- Incorporate fully coupled hydromechanical (e.g., poroelastic) bulk effects into models of earthquake source processes, which are potentially key to a number of earthquake effects including induced seismicity;
- Determine constitutive relations for the evolution in hydro-mechanical properties due to near-fault damage and healing, which can have profound effects on the evolution of pore fluid pressure in the fault core and hence on the fault core shear resistance;
- Explore evolution in hydro-mechanical properties due oscillating dynamic stresses induced by wave propagation (shaking), potentially dominating effect for distant earthquake triggering.

**MMC1.5 Combining lab and field findings with modeling to construct scale-appropriate constitutive relations**

Numerical modeling of earthquake source processes involves discretization of the domain of interest into spatial elements, often called cells. For the modeling at the scale of intermediate to large earthquake ruptures to be tractable, cell sizes cannot be much smaller than 1-10 m and are often much larger, especially in 3D bulk models with 2D faults. At the same time, most laboratory experiments operate at the scales of 1-10 cm or less. This discrepancy creates the issue of upscaling the laws formulated based on laboratory experiments to the scale of the numerical discretization. Such upscaling is not straightforward, since the change in scales brings with it additional processes and structure not necessarily captured by the smaller-scale lab-based constitutive relations.

**Fault slip vs. distributed inelastic deformation**

The shear deformation within the fault shear zones is often idealized as slip - or relative shear displacement - across the faults, both for reporting field or lab observations and for modeling purposes. For example, kinematic inversions of large earthquakes routinely report the final slip distribution, given the assumed fault geometry (Simons et al., 2011; Minson et al., 2014ab; Adams et al., 2016; Ye et al., 2016; Ragon et al., 2018). While this is a clearly needed abstraction relevant to the localized shear zones of both mature faults at seismogenic depths and laboratory samples, the notion of slip implies an effective behavior over small-scale complexity (MMC1.1-1.3) and hence its use is not straightforward. This is especially true when the notion of slip is used more broadly, even for the more distributed deformation in the shallow portion of the faults (MMC1.2) and below the seismogenic zone (MMC1.3), in which cases the notion of slip needs to be used with care. For example, differentiating between slip across a relatively narrow principal fault shear zone and more distributed shear deformation on larger scales can explain some observational paradoxes, such as the shallow slip deficit on faults (Simons et al., 2002; Fialko et al., 2005; Bilham et al., 2010; Kaneko and Fialko, 2011; Erickson and Dunham, 2017).
The dominance of the notion of fault slip in earthquake source processes goes back to the conceptually powerful but simplified model of elastic rebound (Reid, 1910) and its descendants, in which faults are represented as interfaces in an otherwise elastic bulk. In such models, (i) the bulk around the faults is treated as elastic and its deformation is fully reversible, with the accumulated elastic strain energy due to plate motion released through fault slip and (ii) all dissipation (and hence irreversibility) occurs on the fault interface, usually through friction and its extensions that can be quite elaborate, laboratory-based, and coupled to shear heating and fluid processes. Such slip-based models, with appropriately selected fault constitutive relations, are valuable research tools that can be quite informative about earthquake source dynamics (Tse and Rice, 1986; Rice, 1993; Ben-Zion and Rice, 1997; Okubo, 1989; Shibazaki and Matsu'ura, 1992; Marone, 1998ab; Lapusta et al., 2000; Lapusta and Rice, 2003; Kato and Tullis, 2003; Hori et al., 2004; Kato, 2004; Duan and Oglesby, 2005; Liu and Rice, 2005; Rubin and Ampuero, 2005; Hillers et al., 2006; Ziv and Cochard, 2006; Dieterich, 2007; Ampuero and Rubin, 2008; Chen and Lapusta, 2009; Chen et al., 2010; Fang et al., 2010, 2011; Noda and Lapusta, 2010; Kaneko et al., 2011; Barbot et al., 2012; Schmitt et al., 2015; Bruhat and Dunham, 2016; Lin and Lapusta, 2018; Lui and Lapusta, 2018; MMC2). It may indeed turn out that the main aspects of the fault behavior can be captured by such models, e.g., if inelastic effects in the surrounding rock bulk are either relatively unimportant or sufficiently well-mimicked by the appropriately selected fault-shear-zone constitutive relations, as may be relevant to relatively geometrically simple and mature fault segments at seismogenic depths. For example, approximating the nonlinear and dissipative parts of a problem by appropriately selected constitutive relations on a crack interface has been a productive approach in a number of fracture-mechanics solutions for brittle solids (Burridge, 1969; Ida, 1972; Palmer and Rice, 1973; Andrews, 1976; Madariaga, 1976; Das and Aki, 1977a,b; Rice, 1980; Freund, 1990). However, such models can only be properly constructed and validated when the effects of broader inelastic bulk behavior are captured through more elaborate models with realistic, observationally and laboratory determined, fault/bulk structure and properties.

An important and open research question is to identify the fault scales and processes that can be treated within the notion of fault slip and the associated shear resistance (often called fault friction) vs. those that need to be explicitly included in models in addition to fault slip.

**Fault structural complexities and the need for constitutive relations at 1-100 m scale**

One aspect of the fault geometry that can intrinsically link the shear-layer resistance to slip and the inelastic deformation of the surrounding bulk is fault roughness or non-planarity (Brown and Scholz 1985, Power et al., 1987; Sagy et al., 2007, Dieterich and Smith, 2009; Brodsky et al., 2016; Candela and Brodsky, 2016). Defined by the average height of protrusions on the surface at a given length scale, the roughness has been shown to increase as a power-law function of the length scale of observation (Brown and Scholz 1985; Power et al., 1987; Sagy et al., 2007; Brodsky et al., 2016). Several mechanisms are expected to promote fault roughness, including the creation of faults through linkage of microfractures and dynamic effects ahead and at the rupture tip that promote fault bending and dynamic branching (Poliakov et al., 2002; Rice et al., 2005; Ponson et al., 2006; Brodsky et al., 2011; Fang and Dunham, 2013; Shi and Day, 2013; Shervais and Kirkpatrick, 2016; Zielke et al., 2017; Dascher-Cousineau et al., 2018). Still, kilometers of relative motion can be accommodated by the same narrow shear layers, as supported by some studies of the exhumed faults (Chester and Logan, 1987; Chester et al., 1993; Wibberley and Shimamoto, 2003; Sutherland et al., 2012), which means that the roughness and its variation do not reshape the slipping path at least in some cases, perhaps indicating that mature faults at seismogenic depths may be smoother, due to their large accumulated slip and formation of gouge layers. But it is clear that at least some roughness would remain for all faults under all conditions.
In the presence of fault roughness on scales comparable to the scale of slip (1-10 m in most large earthquakes), the localized slipping surface would be locally nonplanar on scales that are not tractable in large-scale 3D simulations (which use cell sizes of the order of 10-100 m or larger) (Figure II-MMC1.5). The resulting spatially variable damage on small scales, variable hydromechanical properties etc cannot be explicitly included in any large-scale modeling of societal interest. Hence the effects of such roughness, and any other variations in the slipping layers and/or fault structure on such scales, need to be incorporated through appropriately formulated constitutive relations. The resulting relations for the constitutive behavior of faults suitable for large-scale modeling should then capture not only the resistance of highly localized shear layers but also of the nearby damage zones and all the processes expected there (MMC1.2).

Furthermore, the actively shearing layers of seismogenic faults, often called fault cores, can vary greatly in width along strike, from the very localized, sub-millimeter deformation (Chester and Logan, 1987; Chester et al., 1993; Wibberley and Shimamoto, 2003; Sutherland et al., 2012) (MMC1.1) to much broader zones of the order of tens of meters. The broader zones are often manifest as multiple zones of fault gouge or cataclasite (a cohesive collections of small rock pieces within a finer-grained matrix) several metres thick embedded into a damage zone (Faulkner et al., 2003, 2008; Zoback et al., 2010); the nature of the fault zone structure is controlled by fault lithology, geometry, and cumulative offset (Faulkner, et al. 2003). Deeper fault extensions of exhumed fault systems often consist of networks of thin discrete fault strands encompassed within zones of distributed, volumetric deformation, which may occur at variable slip rates throughout the earthquake cycle (Okubo and Aki, 1987; Bürgmann, 2017; Hamling et al., 2017; Melosh et al., 2018; Rowe et al. 2018) (Figure II-MMC1.5).

The explicit inclusion of such structural complexities, on the scale of 1-100 m or less, is clearly not tractable in numerical models of earthquake processes at the larger, societally relevant scale. Hence formulating the relations between the constitutive response of such “thick” shear zones - that include multiple narrow shear layers within a wider layer of damage or viscoplastic deformation - and total slip, slip rate, etc would be the most promising way of including such structural complexities into modeling.

**Numerical experiments to capture the needed constitutive response**

The required constitutive relations at the 1-100 m scales need to come from fault/bulk modeling that includes the observed structural complexity at smaller scales. The models would use the lab-derived constitutive relations discussed in sections MMC1.1-1.3 at the scales of 1-10 cm or less, to represent the localized - but potentially rough and/or multiple - shear zones and the surrounding damaged and/or viscoplastic bulk. As the constitutive relations discussed in MMC1.1-1.3 are further developed, these models will become increasingly more realistic, coupling fault slip and rock deformation with damage creation, viscoplastic flow, fluid flow, temperature evolution, and chemical reactions. The resulting modeling tools on 1-100 m scale can be used in two ways. First, they can be used to probe the constitutive response at the 1-100 m scale by conducting “numerical experiments” on the models. Such numerical experiments would use simplified slip and slip-rate histories, just like the actual laboratory experiments do. The goal would be to connect the average stressing on the 1-100 m scale to the slip rate and other appropriately selected state variables, i.e., to create rate-and-state formulations, as have been done for highly localized shear layers (section MMC1.1). The resulting constitutive formulations can then be used on the fault interfaces in much larger-scale (10-1000 km) modeling, which would still need to include fault and bulk complexities - damage zones, viscoplastic flow, roughness - but at larger, 0.1-1 km and up, scales (section MMC2). Second, one could couple these 1-100 m scale models with larger-scale simulations in a multiscale framework.
Key future challenges

- Determine laws for effective shear resistance of localized shear layers and adjacent areas, due to the coupled evolution of localized friction of the spatially non-planar shear layer, nearby damage, and the resulting fluid flow;
- Determine laws for effective shear resistance of more complex fault cores, with multiple shear layers embedded into a damage layer, including near-surface “flower” structure;
- Distinguish between mature vs. immature faults through constitutive relations; potential candidates include the degree of local roughness, mineralogy, particle distributions, and width of the actively shearing granular layer.

Cross-cutting themes in MMC1.1-1.5:

- Systematic community-wide effort to produce coherent sets of constitutive laws for realistic fault materials under realistic fault conditions, using systematic sets of experiments, unified theoretical frameworks, and modeling at the lab scale to clarify and untangle various effects.
- Numerical experiments to construct constitutive laws on the intermediate scales of 0.1-100 m - larger than the typical lab samples but smaller than the spatial discretization in large-scale simulations - needed for tractable numerical modeling, with predictions for lab and field studies to test; such numerical experiments could capture the effects of intermediate-scale complexity and use lab-derived laws at lower scales.
MMC2. Building a coherent suite of numerical methodologies for multi-physics problems at larger scales

MMC2.1 Dynamic rupture simulations: capturing a single earthquake event

Dynamic (or spontaneous) rupture simulations focus on rupture propagation and ground motion during a single earthquake. These simulations emerged with the applications of dynamic fracture mechanics to earthquakes (Burridge, 1969; Freund 1960; Andrews, 1976a,b; Madariaga, 1976; Das and Aki, 1977a,b; Rice, 1980) and the idealization of earthquakes as shear cracks that propagate in response to rapid reduction of shear strength on the fault interface. Dynamic rupture simulations help us understand how frictional strength evolution - and more broadly the constitutive behavior of faults and surrounding bulk discussed in MMC1 - controls the rupture history, slip distribution, ability of ruptures to navigate structural complexities and jump between fault segments, and radiation of seismic waves.

Wave-mediated stress transfers along the faults are fundamentally important for the rupture propagation along the fault, so dynamic rupture simulations involve solving the equations of motion - that include the inertial term - in the rock bulk with a boundary condition on the fault or faults in terms of a friction law or, more generally, a description of the constitutive fault behavior strength evolution (MMC1.1). The rock bulk is often taken as elastic, but increasingly, inelastic processes are also included such as off-fault plasticity or damage rheology (MMC1.2). Dynamic rupture simulations solve for fault slip history, stress evolution, and radiated wavefield that satisfy this boundary-value problem. This is in contrast to the kinematic source modeling which assumes the fault slip history and computes the resulting radiated wavefield.

Many dynamic rupture simulations use a linear slip-weakening description of fault friction, in which the shear resistance drops from the peak value to dynamic value linearly over so-called critical slip $D_c$ and then stays constant. This description is widely used for several reasons. It is the simplest smoothed implementation of the classical notion that shear strength reduces from static to dynamic friction (Figure V-MMC1.1-2). It is also the simplest way to regularize the crack tip in theories of dynamic cracks, with clearly defined notion of dynamics-governing fracture energy (Ida, 1972; Palmer and Rice, 1973; Rice, 1980; Freund, 1990) which is now more broadly called “breakdown work” (Bizzarri and Cocco, 2003); the associated theoretical knowledge allows for easier design and analysis of numerical studies (Kame et al., 2003). Finally, it has been observed in some laboratory experiments when the shear resistance is plotted as a function of slip (Ohnaka et al., 1987; Ohnaka and Kuwahara, 1990; Ohnaka and Shen, 1999).

Interestingly, numerical simulations show that a commonly used form of the standard low-slip-velocity rate-and-state friction formulation (MMC1.1) translates, in an actively nucleating fault region and at the propagating rupture tip, into a dependence of shear resistance on slip similar to linear slip-weakening friction (Bizzarri and Cocco, 2003; Lapusta and Liu, 2009), reconciling the intrinsic rate-and-state nature of friction with the experimentally observed slip-weakening behaviors in some special cases. However, linear slip-weakening is clearly a significant simplification of the actual resistance of faults to dynamic rupture, as discussed in MMC1.1, especially since laboratory experiments under high slip velocities of ~1 m/s characteristic of dynamic rupture show much different behaviors than linear slip-weakening - or the low-slip-velocity rate-and-state friction - with enhanced dynamic weakening and significant additional rate effects (MMC1.1, Figure II-MMC1-1). The enhanced dynamic weakening can lead to low-heat, low-stress fault operation (MMC 3.1); significantly affects earthquake energy budget; and can generate self-healing pulse-like ruptures and other rupture modes that slip-weakening formulations do not allow (e.g., Lu et al.,...
Hence dynamic rupture simulations increasingly seek to incorporate more realistic shear resistance descriptions, the trend that should continue.

**Fundamental and quantitative insights into rupture behavior**

Dynamic rupture modeling has led to significant fundamental insights into the properties of large earthquakes. Dynamic rupture simulations were the first to confirm the analytical possibility that the rupture speed - i.e., the speed with which the rupture tip propagates along the fault - could exceed the shear wave velocity (Andrews, 1976), leading to unique ground motion characteristics (Freund, 1979; Aagaard and Heaton, 2004; Dunham and Archuleta, 2004, 2005; Bhat et al., 2007) (Figure II-MMC2.1-1). The so-called supershear rupture propagation has since been inferred for many large earthquakes (Archuleta, 1984; Olsen et al., 1997; Bouchon and Vallée, 2003; Dunham and Archuleta, 2004; Ellsworth et al., 2004; Konca et al., 2010; Yue et al., 2013) and confirmed in laboratory shear rupture experiments (Xia et al., 2004; Passelègue et al., 2013; Mello et al., 2010, 2014). Current modeling work on supershear ruptures is aimed at reconciling predicted ground motions from simulations (which, at least in homogeneous media, feature elevated shaking at high frequencies) and observations (which are similar to or possibly depleted at high frequencies relative to shaking from subshear ruptures). Potential explanations include additional complexity in the source process (Bizzarri et al., 2010) or, more likely, scattering (Vyas et al., 2018).

Furthermore, dynamic rupture modeling, constrained by matching near-source strong ground motion records (ground shaking time series recorded by accelerometers and similar instruments), provides quantitative constraints on coseismic stress changes and insights into observed rupture behavior. Early studies of the 1984 Morgan Hill (Mikumo and Miyatake, 1995) and 1992 Landers (Olsen et al., 1997) earthquakes, for example, revealed how heterogeneities in initial stress and/or friction introduce complexities in the rupture progress that are required to match ground motion records and kinematically inferred slip distributions. This effort motivated introduction of dynamic rupture models with stochastic initial stress (e.g., Mai and Beroza, 2000; 2002; Oglesby and Day, 2002). In addition, matching strong motion records with dynamic rupture simulations can be formulated as an inverse problem with stress and friction as model parameters (Peyrat and Olsen, 2004; Di Carli et al., 2010; Ruiz and Madariaga, 2011, 2013; Ruiz et al., 2017; Gallovic et al., 2018), with results offering more insight into the earthquake source process.

Significant advances have been made in using dynamic rupture simulations to understand earthquake energy balance and observed features of large subduction zone earthquakes, in particular magnitude 8 and 9 class megathrust ruptures that can span the entire seismogenic zone of subducting plate boundary faults. These achievements are further discussed in MMC3.3.

**Interaction of dynamic rupture with complex fault geometries**

Advances in source imaging reveal extremely complex rupture behavior, including ruptures jumping across step-overs to link multiple fault segments in a single event and navigating other structural complexities like bends, branches, and splay faults (Figure II-MMC2.1-2). Such behavior can be reproduced in dynamic rupture simulations. Idealized studies on relatively simple structures provide insight into controls on, for example, jumping across step-overs (e.g., Harris et al. 1991; Harris and Day, 1993, 1999; Harris et al., 2002; Aochi, 2003; Duan and Oglesby, 2006, 2007; Lozos et al., 2014; Ryan and Oglesby, 2014; Bai and Ampuero, 2017) or taking a branch (e.g., Kame et al., 2003; Oglesby et al., 2004; Duan and Day, 2008; Xu et al., 2015), which can be validated using observational compilations (e.g., Wesnousky, 2008; Biasi and Wesnouski, 2016, 2017). These idealized studies are complemented with
simulations of ruptures through the highly complex fault networks observed in nature (e.g., Meng et al., 2012; Ando and Kaneko, 2018; Wollherr et al., 2018). Validation, or at least consistency, can be achieved by comparison to inferred rupture histories from kinematic source inversions and backprojection imaging, surface slip distributions mapped by field and remote sensing studies, and by matching seismograms and deformation data. Such simulations can also be performed prospectively to quantify likely rupture behaviors and associated hazard. As one example, identification of a branched fault network adjacent to the Diablo Canyon Power Plant, a nuclear reactor facility along the California coast, raised questions regarding likelihood of the branch segment being activated (Hardebeck, 2013). These questions motivated dynamic rupture code verification efforts on branched fault geometries (Harris et al., 2018) and ongoing studies using these codes. This topic is further discussed in section MMC2.4.

**Fundamental role of dynamic rupture modeling in understanding and predicting strong ground motions**

Dynamic rupture simulations will play an increasingly fundamental role in strong ground motion modeling and seismic hazard assessment. Early simulations with relatively uniform stress and frictional properties were successful in matching long-period ground motions, which in the near-source region are sensitive to long-wavelength slip and average rupture velocity.

Advances in computing continue to push the upper frequency limit of ground motion simulations (Figure II-MMC4.1), focusing attention on shorter-wavelength and shorter-time-scale details of the rupture process. For example, dynamic rupture models using planar fault geometry where heterogeneity of the stress fields is characterized with power-law spectra have been successfully validated against empirical data (e.g., Andrews and Barall, 2011; Andrews and Ma, 2016) and proven useful in ground motion prediction for large-scale earthquake scenarios (e.g., Roten et al., 2018). Ground motion records from hazard-scale earthquakes at frequencies greater than 1 Hz have a random character, arguably arising from complexity in the rupture source process as well as scattering along the wave propagation path. Source complexity emerges naturally in dynamic rupture simulations with stochastic initial stress, frictional properties, and/or small-scale complexities in fault geometry like surface roughness. These causative factors are likely correlated, with stress heterogeneities developing during sliding across rough fault surfaces or interaction with adjacent secondary faults. Hence dynamic rupture models can significantly contribute to our understanding and prediction of strong ground motion (MMC4.1).

**Importance of the assumed initial conditions**

Dynamic rupture simulations can elucidate the consequences of different model ingredients and physical mechanisms for the specified fault initial conditions. But significant trade-offs exist between the assumed fault constitutive relation and the assumed initial conditions. Given that, as well as relatively limited set of observations for the dynamic rupture alone, e.g., compared with the long-term simulations (MMC1.2, MMC3), dynamic rupture simulations cannot constrain many aspects of the fault physics discussed in MMC1, unless the initial conditions are known independently. Prior earthquake ruptures, slow slip, inelastic deformation, fluid flow etc all serve to redistribute stresses, pore fluid pressure, and other variables before a dynamic rupture and, in particular, make them more self-consistent with the assumed fault geometry and constitutive behavior. This is particularly important for simulations with complex fault geometries, structure, and roughness, where prior slip and deformation are likely to create stress and pore pressure patterns characteristic, at least at larger scales, for the assumed geometry and structure (Foxall et al., 1993; Fang and Dunham, 2013; Shi and Day, 2013; Zielke et al., 2017; Melosh et al., 2018). That is why it is important to develop long-term simulations of fault slip, even to study dynamic rupture, as discussed in MMC2.2.
For dynamic rupture simulations, this problem can be partially addressed by using the initial conditions, at least on larger spatial scales, motivated by long-term simulations that take into account past slip history and fault geometry, especially as such simulations further develop (MMC1.2). For example, the insight that the bottom of the seismogenic zone can have elevated shear stresses compared to the rest of the fault due to the continuous interseismic deformation of the fault roots, supported by long-term simulations (Lapusta and Rice, 2003; Jiang and Lapusta, 2016, 2017), has allowed to produce a dynamic rupture model for the 2016 Mw 7.8 Kaikoura earthquake that explained continuous rupture propagation over a complex misaligned network of faults (Ulrich et al., 2018; Figure II-MMC2.1). Another possibility for dynamic rupture simulations over complex fault geometries would be to take, as initial stress conditions, the final stress conditions after one such dynamic rupture simulation, supplemented by a projected interseismic stress increase due to regional tectonic loading. While this approach ignores the interseismic effects, it would allow the modeling to incorporate stress concentrations and other features produced by prior seismic slip.

Capturing effects of smaller-scale features on larger-scale outcomes

Despite the initial-condition limitation, dynamic rupture simulations can be used to investigate a number of key questions in earthquake source physics, as already discussed. For example, they can evaluate the relative importance of different physical mechanisms given the assumed initial conditions, e.g., shear-layer vs. damage-zone dissipation during dynamic rupture (see MMC3.3). Dynamic rupture simulations can also investigate the effect of smaller-scale processes and features on the larger-scale outcomes, by comparing simulation results for models which are the same on larger scales but have different heterogeneity, roughness, or other features on smaller scales. For cases when the responses are different on larger scales, which is possible and even likely in the highly coupled, nonlinear problems of interest, the simulations can potentially be used to determine whether the larger-scale constitutive relations can be modified to mimic the outcomes of the more broad-band models.

Developing dynamic rupture simulations to incorporate more coupled mechanisms

It is important that the dynamic rupture simulations are further developed to incorporate the most physically relevant constitutive representations of various physical phenomena as discussed in MMC1. At present, most dynamic rupture simulations incorporate advanced representations of some ingredients and highly simplified representations of others, due to a variety of reasons such as multi-scale challenges discussed in section I.3 and the associated numerical tractability, the desire to keep the model parameters to a minimum, the aim to highlight a particular effects, or capabilities of a given numerical code. For example, a model may incorporate complex fault geometry in a 3D medium but, for simplicity, linear slip-weakening friction (which is the simplest conceptual representation of fault shear resistance) and elastic bulk (Refs). The analysis of many physical effects discussed in MMC1 is still limited to simplified two-dimensional (2D) models (with one-dimensional faults) or planar interfaces in 3D.

The current community work on more systematic inclusion of a range of relevant mechanisms - realistic fault geometry, shear resistance, enhanced dynamic weakening, off-fault damage, poroelastic effects - with scale-appropriate constitutive relations (MMC1) will allow dynamic rupture models to capture coupling between different physical mechanisms and evaluate their relative importance for rupture dynamics.

Potential use in long-term simulations, machine learning
In a long-term simulation of fault slip and deformation, such as the ones discussed in MMC1.2-1.4, one can switch to a detailed dynamic rupture simulation when a dynamic event is coming up, and then switch back afterwards, potentially many times. Such approach has already been used to develop SEAS-like simulations beyond planar faults in elastic media (Kaneko et al., 2011, Duru et al., 2019). Currently, such approaches are only tractable either for relatively small-scale, geometrically simple 2D problems, or for a small number of switches in a more realistic, 3D simulations. However, as the computational resources and algorithms develop, this may be the most straightforward and promising avenue to develop long-term simulations of sequences of earthquakes and slow slip (SEAS) with realistic inertial dynamics during dynamic rupture, realistic fault geometry, and inelastic off-fault response.

Dynamic rupture simulations can also be used to develop databases of rupture scenarios for specific faults of interest, for many plausible sets of initial conditions. One can explore the use of machine-learning algorithms in conjunction with such databases to generate rupture scenarios based on the pre-event conditions in in long-term large-scale simulations. While such approaches would not be fully physical, they may bring a higher level or realism to the long-term simulations that what can be achieved without them (see also MMC1.4).

**Different implementations, code verification**

The primary computational expense of a dynamic rupture simulation comes from solving the equations of motions which, for linear elastic solids, take the form of the Navier equation of motion (e.g., Freund, 1990; Aki and Richards, 2002), almost always using explicit time-stepping methods. Spatial discretization is done with a number of methods, including finite differences (Andrews, 1976; Day, 1982; Olsen et al., 1997; Ely et al., 2009; Kozdon et al., 2013; Duru and Dunham, 2016), finite elements (Aagaard et al., 2001; Duan and Oglesby, 2005ab; Barall, 2009; Aagaard et al., 2013), spectral elements (Kaneko et al., 2011; Galvez et al., 2014), and discontinuous Galerkin (Tago et al., 2012; Pelties et al., 2014). Each discretization method comes with different ways of computing the field quantities and imposing the boundary conditions. The methods face challenges, such as i) earthquake source processes being generally ill-constrained and highly non-linear; ii) trade-offs of physical processes in terms of dominance and relevance at a given spatio-temporal scale (and in real earthquakes) related to the justification of their (most often computational) cost of their inclusion; iii) the assimilation of all available knowledge in a suitable manner for software (numerical discretisation, solvers, equations solved) and hardware (heterogeneous HPC systems, energy concerns).

Given the variety of approaches used, and as more physical mechanisms are incorporated, continuing verification of the modeling methods involved is quite important. Such verification exercises are tantamount to ensuring that the codes work as intended as they evolve with added capabilities (Harris et al., 2018). Moreover, the use of versioning software should become standard, to streamline future progress and interchange of codes, as strong collaborations are required to reach the goal of an integrated modeling effort for rupture dynamics effects.

**Need for supercomputing facilities and collaboration with computer scientists**

Recent progress in modeling of advanced rupture dynamics features and the resulting broadband ground motion in realistic 3D models have in part been fueled by access to high-performance computing facilities. Examples of advances obtained on powerful supercomputers include deciphering the complex interaction between rupture on the various non-planar segments with step-overs for the 1992 Landers, CA, earthquake (Wollherr et al., 2018), off-fault deformation and shallow slip deficit (Roten et al., 2017), and simulation of large scenarios on the southern San Andreas Fault with embedded damage zones in visco-
elastic and hysteretic nonlinear rheology (Roten et al., 2018) (Figure II-MMC2.1). The gateway to future breakthroughs in the field is critically dependent on access to sufficiently large computational platforms, with up-and-coming architectures, such as GPUs, many integrated core systems, etc. It is important for the researchers in the field to know about the existence of and gain access to such facilities, in order to accelerate progress toward more integrated solutions (see also section IV-IN1).

In addition, it is important to expand collaborations between numerical modelers and computer scientists that have proven effective at optimizing and improving the scalability of key applications for rupture dynamics and wave propagation (e.g., Cui et al., 2010, 2013; Breuer et al., 2014; Heinecke et al., 2014; Roten et al., 2016; Uphoff et al., 2017;). Further advancements in the efficiency and tractability of the simulations will come, in part, from adopting higher-order methods, adaptive mesh refinement, and efficient parallelization which geoscientists cannot efficiently implement on their own. Furthermore, such collaborations will help implement advanced workflows that are likely to play an increasingly important role in future ensemble simulations, parameter studies, and pinpointing coupling between different physical mechanisms in rupture dynamics (e.g., fault roughness and dynamic weakening, multi-material and fluid effects).

Key future goals
- Develop dynamic rupture simulations with scale-consistent representations of all relevant processes, including fault geometry and evolving fault/bulk structure/properties (shear-layer resistance, damage zones, fluid effects, MMC1.1-1.4), to capture coupling between different physical mechanisms and evaluate their relative importance for rupture dynamics;
- Explore the best ways of selecting initial conditions, based on fault-resolved regional loading and insight from long-term simulations;
- Construct workflows for ensemble dynamic rupture simulations to predict strong ground motions.

 MMC2.2 Modeling sequences of earthquakes and slow slip/deformation

Earthquake rupture propagation and resulting ground motions are highly sensitive to stresses on the fault and in the surrounding medium, as well as other conditions such as pore fluid pressure, at the initiation of rupture. While modelers must specify, using prior information or educated guesses, such initial conditions in single-event dynamic rupture simulations, these inputs are generated in a self-consistent manner in long-term simulations of fault slip/deformation that aim to produce Sequences of Earthquakes and Aseismic Slip (or SEAS).

SEAS simulations are also called earthquake cycle simulations, though that terminology should not be understood to imply periodic event sequences. The word “cycle” rather refers to the total collection of the interseismic period, earthquake nucleation, co-seismic rupture, and postseismic effects; this “cycle” of earthquake life indeed repeats, although not necessarily periodically or with an event of the same scale.

SEAS simulations typically span decades to tens of thousands of years and capture all phases of the earthquake and faulting process: interseismic loading, nucleation of an earthquake, coseismic rupture propagation, and the postseismic response. Just like single-event dynamic rupture simulations, SEAS simulations require a description of both fault strength evolution and the response of the off-fault medium. In seismogenic regions, fault strength evolution must not only include weakening to generate the coseismic phase but also, unlike in a single-event simulation, restrengthening over the interseismic period, to simulate sequences of events. The wide range of time scales in SEAS simulation necessitates
the use of adaptive time-stepping methods. Explicit Runge-Kutta methods with embedded error estimates are a common choice (Liu and Rice, 2005; Erickson and Dunham, 2013; Lambert and Barbot, 2016; Allison and Dunham, 2018,2019), though other adaptive time-stepping methods have also been used (Lapusta et al., 2000; Lapusta and Liu, 2009; Noda and Lapusta, 2010; Luo and Ampuero, 2017; Herrendoerfer et al., 2018).

Many current approaches to SEAS simulations utilize the so-called radiation-damping approximation (Rice, 1993) during the coseismic phase but ignore all other wave-mediated stress changes. Simulations that use the radiation-damping approximation are termed quasi-dynamic. The radiation-damping approximation is widely used because it is a trivial addition to the quasi-static computations that solve the equations of equilibrium, while rigorously accounting for inertial effects and the associated wave-mediated stress changes is a major computational challenge that requires tracking the wave propagation and associated patterns of dynamic stress changes.

**Success of SEAS simulations in capturing qualitative, and often quantitative, fault behaviors**

SEAS simulations have been applied across the full spectrum of faulting problems and tectonic environments. They emerged in the mid-1980s following the introduction of the rate-and-state friction laws (section MMC1.1). Prior to that time, the depth extent of crustal earthquakes was understood to be controlled by the onset of thermally activated creep that allowed rocks to accommodate tectonic motions by distributed viscous flow instead of frictional sliding on localized fault surfaces (MMC1.3). A complementary idea emerged from laboratory studies of the rate dependence of friction (Stesky, 1975; Dieterich, 1981; Tullis and Weeks, 1986; Blanpied et al., 1991, 1995), which showed a transition from velocity-weakening friction at mid-crustal temperatures to velocity-strengthening friction at temperatures characteristic of the lower crust. Theoretical stability studies (Rice and Ruina, 1983; Ruina, 1983; Dieterich, 1992) demonstrated that such transition implies the change in behavior, from stick-slip for sufficiently large velocity-weakening interfaces to stable slip with the prescribed rate for velocity-strengthening interfaces. This frictional stability transition provided an alternative explanation for the depth extent of large earthquakes and microseismicity, namely, that the fault might continue as a localized structure through the lower crust and possibly into the upper mantle, with sliding occurring aseismically in the deep velocity-strengthening portion (see section MMC1.3 for more discussion).

Rate-and-state-based SEAS models thus idealized the system as an elastic half-space containing a fault with depth-dependent frictional properties (Tse and Rice, 1986; Rice, 1993; Ben-Zion and Rice, 1995). At great depth on the fault, steady sliding at a constant plate velocity was applied to load the system, known as backslip loading. These early SEAS simulations reproduced remarkably many observed features of the earthquake cycle: recurrence intervals of hundreds of years for typical crustal thickness and plate velocities, several meters of slip per event, and a postseismic response featuring transiently elevated aseismic sliding velocities on the fault extension below the seismogenic zone. The versatility of the rate-and-state formulation comes from its combination of the always stabilizing, velocity-strengthening direct effect and potentially destabilizing state-evolution effect over characteristic slip, that together result in just broad enough collection of fault behaviors: from stable sliding in steady-state velocity-strengthening regions, to stable sliding in small enough steady-state velocity-weakening regions, to the notion of a nucleation size - which is the size that the slipping region needs to reach before slip becomes self-accelerating - to stick-slip on large-enough velocity-strengthening regions (section MMC1.1). The small-scale stability of velocity-weakening regions, in particular, is a key feature that allows for well-posed problems and well-resolved numerical solutions (Rice et al., 2001).
Note that some simulations of slow earthquake nucleation and propagation employed slip-dependent rather than rate-and-state friction (Shibazaki and Matsu’ura, 1992). The slip-dependent friction law was formulated based on a theoretical model of a fractal fault-surface wear (Matsu’ura et al., 1992), motivated by experiments (Ohnaka et al., 1987). Later, time-dependent healing of the amplitude of the surface topography was introduced (Aochi and Matsu’ura, 2002) to construct a slip- and time-dependent fault constitutive law, which was applied to earthquake-sequence simulations (Hashimoto and Matsu’ura, 2002, Hashimoto et al., 2014).

Simulations with purely rate-dependent friction were also used to simulate earthquake sequences (Shaw, 1994; Cochard and Madariaga, 1996; Myers et al., 1996), although they did not include any aseismic fault slippage and initiated dynamic rupture abruptly. Such models resulted in complex event sequences with Gutenberg-Richter-like distribution of event sizes (section MMC3.5), even on planar homogeneous faults. At the same time, early SEAS simulations of homogeneous planar faults based on rate-and-state friction predicted nearly periodic sequences of identical large earthquakes (Tse and Rice, 1986; Rice, 1993). A vigorous debate ensued regarding the origin of earthquake complexity, specifically whether nonlinearity in the frictional dynamics was sufficient to produce Gutenberg-Richter-like distribution of event sizes. The complexity in the purely rate-dependent formulations was linked to the effectively zero nucleation size in such formulations, resulting in a non-physical phenomenon of dynamic events over one cell size in the numerical model and hence discretization-dependent results (Rice, 1993), which can be more appropriately interpreted as being on strongly segmented fault (Ben-Zion and Rice, 1995). Artificial event complexity arises also in the rate-and-state models with improper nucleation-size resolution (Rice, 1993; Lapusta et al., 2000).

Ultimately, analysis and simulations demonstrated the need to regularize the rate-weakening friction with the complete rate-and-state framework, in which the velocity-strengthening direct effect stabilizes the response at short wavelengths. An important lesson is that SEAS simulations must utilize sufficiently small grid spacings to produce a solution that converges with mesh refinement and to avoid producing artificial complexity that often appears when grid spacing is too large (Rice, 1993; Lapusta et al, 2000). It was also shown that nonlinear dynamics can indeed produce some complexity in event sequences, in properly resolved well-posed simulations, if there is sufficient scale separation between the nucleation length and size of the seismogenic zone (Shaw and Rice, 2000; Lapusta and Rice, 2003). Additional complexity arises through the introduction of heterogeneity in frictional properties and effective normal stress and by the interaction of multiple fault segments.

SEAS simulations have provided continuing insights into a broad range of slip phenomena, including earthquake nucleation, the postseismic response from afterslip, patterns of seismic and aseismic slip, and aftershock sequences (Tse and Rice, 1986; Okubo, 1989; Shibazaki and Matsu’ura, 1992; Dieterich, 1992, 1994; Ben-Zion and Rice, 1997; Lapusta et al., 2000; Lapusta and Rice, 2003; Kato and Tullis, 2003; Hori et al., 2004; Kato, 2004; Duan and Oglesby, 2005; Liu and Rice, 2005; Rubin and Ampuero, 2005; Hillers et al., 2006; Dieterich, 2007; Ampuero and Rubin, 2008; Kaneko and Lapusta, 2008; Chen and Lapusta, 2009; Chen et al., 2010; Fang et al., 2010, 2011; Kaneko et al., 2011; Cattania and Segall, 2018; Lui and Lapusta, 2018).

Much work has focused on the role of heterogeneous frictional properties, specifically interspersed regions of velocity-weakening and velocity-strengthening friction. This heterogeneity, which is fixed in space, helps explain similarly sized events rupturing the same fault section, like the magnitude 6 events in the Parkfield section of the San Andreas fault (Barbot et al., 2012) and magnitude 6-8 events repeatedly rupturing seismogenic asperities in subduction zones (Kaneko et al., 2010). In the case of small repeating earthquakes commonly observed on creeping segments, the rate-and-state friction
models of small VW patches within much larger VS areas can capture their unusual scaling of moment with the recurrence time due the combined seismic and aseismic slip at the same location as well as other properties (Chen and Lapusta, 2009; Cattania and Segall, 2018; Lui and Lapusta, 2018).

Many subduction zones and some other faults around the world have been found to host large slow slip events and associated seismic tremor and low-frequency earthquakes, and SEAS simulations with rate-and-state friction and its extensions have been heavily utilized to formulate and test hypotheses to explain these phenomena (Shibazaki and Iio, 2003; Liu and Rice, 2005, 2007; Rubin, 2008; Segall et al., 2010; Shibazaki et al., 2010; Li and Liu, 2016; Luo and Ampuero, 2007) (Figure II-MMC2.2). The conceptual picture has emerged of a highly heterogeneous plate interface sliding at very low effective normal stress, with tremor and low frequency earthquakes arising when velocity-weakening patches are destabilized (section MMC3.2).

**SEAS simulations with deeper ductile rocks**

While many SEAS models capture the seismic/aseismic transition with depth by a transition from velocity-weakening to velocity-strengthening friction, it is quite clear that increasing temperatures also facilitate a transition from largely elastically deforming - and brittle - rocks in the seismogenic layer to more ductile, viscoplastic rocks in the aseismic substrate (Byerlee, 1968; Kirby, 1980; Chen and Molnar, 1983). SEAS-like simulations of layered bulk, with an elastic layer sitting on top of viscoelastic layer, laboratory-determined power-law bulk flow rheologies, and earthquakes kinematically imposed within the elastic layer (e.g., Takeuchi and Fialko, 2012) have found that distributed ductile deformation indeed occurs but still tends to localize below the fault. Overall, below the seismogenic zone, localized deeper fault extensions may take the form of localized ductile shear zones (see MMC1.3 for more discussion).

SEAS simulations are now being extended to address questions regarding the nature and partitioning of deformation at depth, the depth extent of large earthquakes, and to make connection to geologic field constraints. Work in this direction replaced velocity-strengthening rate-and-state friction below the seismogenic zone with viscous flow laws in fixed-width narrow ductile shear zones or deep fault roots, treated from the perspective of the elastic solver as interfaces (Shimamoto and Noda, 2014; Beeler et al., 2018a). Other studies have been done with a rate-and-state fault in an elastic upper crust overlying a viscoelastic half-space (Lambert and Barbot, 2016), highlighting the possible importance of transient viscous flow in the postseismic period and conditions for slow slip events at the base of the seismogenic zone. More recent SEAS simulations with rate-and-state fault friction in power-law viscoelastic solids are conducted in a manner that allows the transition depth from friction to viscous flow to be determined as part of the solution (Allison and Dunham, 2018) (Figure II-MMC1.3).

The good news from these most recent SEAS simulations is that the loading on the shallower, seismogenic part of the fault is not much affected by the exact nature of shear deformation below, as long as most of it is sufficiently localized. This means that approximations of deeper fault roots through interfaces in an otherwise elastic bulk may be able to capture many effects of the deeper shear deformation, especially if the constitutive response of the interfaces is given by more relevant effective relations for deeper fault roots, as discussed in MMC1.3 and MMC1.5. Formulating such constitutive relations and verifying them, however, requires further developments in the multi-physics modeling that incorporates both frictional faults and realistic inelastic bulk structure, which can be quite complex (Melosh et al., 2018; Rowe et al., 2018) (sections MMC1.3, 1.5).

**SEAS simulations with full inertial effects**
Some SEAS approaches account for full inertial dynamics and wave-mediated stress transfer during the rupture propagation phase. Such formulations are most efficient for simulations of planar faults embedded into a uniform elastic space, in both 2D and 3D models (Ben-Zion and Rice, 1997; Lapusta et al., 2000; Lapusta and Liu, 2009; Noda and Lapusta, 2010) although some of them have started to incorporate non-planar faults or heterogeneous bulk, at least in 2D models (Kaneko et al., 2011; Tal et al., 2016; Duru et al., 2019).

The comparison of simulations with full inertial dynamics vs. their quasi-dynamic approximations highlights important quantitative and often qualitative differences, the extent of which depends on the characteristics of generated dynamic events (which, of course, are not know a priori). In the models with the standard rate-and-state friction and relatively uniform fault properties, the fully dynamic and quasi-dynamic simulations are qualitatively similar, with crack-like ruptures and similar earthquake patterns, although the fully dynamic simulations have significantly higher slip velocities and rupture speeds (Lapusta et al., 2000; Lapusta and Liu, 2009, Thomas et al., 2014).

However, the results of the fully-dynamic and quasi-dynamic simulations become qualitatively different, even in terms of such basic outcomes as long-term event sequences and average fault prestress, for models with enhanced dynamic weakening, which may play an important role in explaining the low-stress, low-heat behavior of mature faults (section MMC3.1). In such models, the wave-mediated stress changes result in the formation of self-healing slip pulses under low-stress conditions typical of mature faults (Zheng and Rice, 1998; Noda et al., 2009). In that case, the fully dynamic SEAS simulations produce a sequence of pulse-like ruptures, whereas the quasi-dynamic SEAS simulations, unable to dynamically create a pulse-like rupture in the absence of the inertial effects, produce numerous smaller events, until a much larger crack-like event spans the entire seismogenic part of the fault (Thomas et al., 2014). Hence the quasi-dynamic simulations can lead to completely different earthquake sequence patterns.

Furthermore, the typical model-spanning events in the two types of simulations differ in average slip by almost an order of magnitude. The average shear stress on the fault is significantly higher in the quasi-dynamic simulations, which would lead to incorrect predictions about temperature evolution, especially important in models that work to implement full thermomechanical coupling with the lower, ductile substrate.

One would expect similarly dramatic differences between the fully-dynamic and quasi-dynamic simulations in other cases where wave-mediated stress transfers can lead to qualitative effects, for example, for non-planar and heterogeneous faults. Strong local heterogeneities can produce local arrest waves and cause short local rise time (Beroza and Mikumo, 1996), a phenomenon that likely cannot be correctly captured in simulations that do not reproduce waves. Another important case is supershear rupture propagation (MMC2.1); since it arises due to specific dynamic stress conditions caused by seismic waves (e.g., Andrews, 1976; Lapusta and Liu, 2008), such propagation cannot be appropriately captured in quasi-dynamic simulations.

Hence SEAS simulations that explore the effects of enhanced dynamic weakening need to be conducted with the full inertial effects, or at least better approximations of them than the standard radiation damping approach. The fully-dynamic SEAS simulations with enhanced dynamic weakening have explored the role of thermal pressurization of pore fluids, demonstrating that it is a viable explanation for the low-stress, low-heat operation of mature faults (Noda and Lapusta, 2010; Jiang and Lapusta, 2016). Such 3D simulations have also showed that enhanced dynamic weakening may lead to large seismic slips over stably creeping regions for fault properties measured in the lab (Noda and Lapusta, 2013), potentially explaining observations for the 1999 Chi-Chi and 2011 Mw 9.0 Tohoku earthquakes.
Future steps: SEAS simulations with off-fault damage, realistic fault geometry, and inertial effects

Faults are surrounded by depth-dependent and evolving damage zones; the renewal of damage during earthquake ruptures and healing in the interseismic period can significantly affect earthquake energy budget and slip dynamics (MMC1.2). SEAS simulations are uniquely positioned to evaluate the long-term effects of these processes on fault slip and damage zone structure, and hence should be developed to incorporate realistic representations of fault damage zones (MMC1.2, MMC1.5). Damage is particularly important near Earth’s surface, where lithostatic pressure and compressive stresses are low and weak materials like sediments are common. SEAS simulations with rate-and-state friction and off-fault damage, in simplified 2D models, have already been used to quantify the fraction of tectonic displacement near the surface that is accommodated by distributed deformation or plastic strain (Kaneko and Fialko, 2011; Erickson et al., 2017), and this effort should continue.

Simulation-based ground motion prediction models require realistic initial stresses (MMC2.1), and the SEAS methodology provides these in a self-consistent manner. Realistic ground motion, especially at high frequencies, emerges from complexity in the rupture process, motivating a focus on stochastic ingredients to SEAS simulations. One candidate is frictional heterogeneity, but perhaps even more promising is the introduction of realistic geometric complexity like the fractal roughness of individual fault surfaces and segmentation and branching within a multisegment fault system. Sliding on geometrically complex faults introduces and maintains stress heterogeneity, and studies demonstrate that fault roughness can significantly influence earthquake nucleation (Tal et al., 2016; Ozawa et al., 2019) and rupture propagation (Dunham et al., 2011b; Shi and Day, 2013; Withers et al., 2018) (Figure II-MMC4.1). It is presently unclear exactly how these stresses evolve over long term and how they influence rupture velocity and slip distributions during rupture propagation. The introduction of nonplanar fault geometries goes hand-in-hand with the inclusion of the off-fault damage, since stress concentrations around geometric complexities inevitably activate inelastic deformation in the brittle upper crust, enhancing the damage creation.

SEAS simulations should build on the current successes in investigating fault interactions in step-over and branched fault systems (Nielsen and Knopoff, 1998; Duan and Oglesby, 2005ab; 2006; 2007; Duan et al., 2019), which are further discussed in MMC2.4.

Note that realistic modeling of damage and geometric complexity effects requires the proper inclusion of inertial effects, or better approximations than the radiation damping term, since the inertial effects lead to much higher stress variations overall, higher stress concentration at the rupture tip, and different stress patterns, that would be important to capture to generate the correct spatio-temporal distribution of damage.

Future steps: SEAS simulations with coupled faulting and fluids

The interface between faulting and fluids should be another focus area for SEAS. Some of the fluid effects are being actively explored in SEAS simulations, specifically dilatant hardening (Rubin, 2008; Segall et al., 2010), thermal pressurization of pore fluids as mentioned above, and their interaction (Segall and Bradley, 2012).

Pore fluid pressure in fault zones controls effective normal stress and fault strength, motivating closer attention to how modelers set pore pressure in SEAS simulations. Most studies simply impose some distribution of effective normal stress; standard choices include a linear increase with depth (assuming
hydrostatic pore pressure) or a distribution that becomes independent of depth (assuming lithostatic pore pressure gradient at sufficient depth). The latter is a predicted consequence of upward fluid flow along fault zones with permeability that depends exponentially on effective normal stress (Rice, 1992) and supported by some observations (Suppe, 2014).

SEAS simulations should therefore account for fluid migration and pore pressure diffusion (or full poroelasticity) with consideration of how transport properties like permeability, porosity, and storage coefficient change with effective normal stress, healing and sealing processes, and coseismic fracturing. Geologists document evidence for extensive long-term fluid effects on faults and their slip and SEAS simulations offer the opportunity to quantitatively explore this process. For example, the damage zones of Alpine fault in New Zealand form a hydraulically active system adjacent to the fault core that likely plays a key role in controlling the evolution of the effective stress, properties, and slip on the fault (e.g., Townend et al., 2017); the fault is also a prime target for modeling due to its near-periodic large earthquakes (Berryman et al., 2012) (Figure II-MMC3.5).

Fluids are also important in the context of oil and gas production, carbon capture and storage, enhanced geothermal systems, and other anthropogenic activities in Earth’s crust that have been linked to elevated seismicity rates (Figure I-4; section III-IN3). Simulations coupling fluid flow and faulting are becoming increasingly common, and recently some have incorporated rate-and-state fault friction (Gugliemi et al., 2015; Torberntsson et al., 2018).

Need for code verification, supercomputing facilities, and collaboration with computer scientists

SEAS simulations are becoming increasingly more widespread and it is essential to verify that codes are providing convergent and adequately resolved solutions. Code verification efforts are therefore essential. Properly formulated mathematical problems should be developed to serve as community benchmarks, with multiple groups/codes providing solutions that can be compared. What may be the first SEAS code verification effort is presently underway through the Southern California Earthquake Center (Erickson et al., 2018). As for dynamic rupture simulations, access to supercomputing facilities and collaborations with computer scientists to optimize the codes is essential for rapid progress in SEAS simulations.

Key future goals

- Develop SEAS simulations that incorporate all or large combinations of relevant ingredients, including nonplanar faults, extended rate-and-state friction formulations including enhanced dynamic weakening (MMC1.1), off-fault damage and healing (MMC1.2), deeper viscoplastic regions (MMC1.3), fluid effects (MMC1.4), and wave-mediated stress transfers during co-seismic rupture;
- Employ the multi-physics simulations to evaluate the relative importance of different ingredients, for a range of parameter values, and formulate simplified representations of coupled effects;
- Use the developed simulations to constrain fault physics by reproducing multiple types of observations with the same modeling ingredients (as discussed in MMC3).

MMC2.3 Extensions of long-term simulations to fault networks

The geometrical complexity of fault systems affects rupture propagation, ground motion at all scales, and sequences of seismic events that occur, as learned from both numerical models and field observations. Long-term simulations of what can be regarded as a single fault segment, even generic, are already quite challenging (MMC2.2), due to multiple coupled physical and chemical mechanisms and a vast range of spatial and temporal scales involved (I.3, MMC1). Extending the simulations to more complex fault
geometries not only brings the obvious computational challenges of the overall size of the numerical computation but also sharpens the issues of (i) knowing the fault and bulk geometry and structure, variable with depth and along strike (MMC1), especially at fault junctions, (ii) selecting scale-appropriate constitutive laws to use, both for the fault shear resistance and the inelastic behavior of the off-fault materials (MMC1) which is amplified by geometric complexity, and (iii) capturing the interaction between the different segments or parts of the network, which potentially has multiple sources, including persistent static stress changes due to nearby events; dynamic stress changes during dynamic events which are fleeting but much larger and decay much slower; static stress changes due to aseismic processes such as postseismic slip and interseismic creep on the deeper fault roots; and potential stress interactions through distributed ductile flow at depth and coupling with fluid flow.

**Insight from dynamic rupture simulations on multi-segment faults**

The effect of multi-segment fault geometry was first explored in dynamic rupture simulations. Work of this type was pioneered in the early 1990s (Harris et al., 1991; Harris and Day, 1993) and considered, in simplified 2D dynamic models, the propagation of rupture through a stepover between disconnected fault segments in an elastic media. It found a maximum jump distance of around 5 km, consistent with field observations (Wesnousky, 2008).

The likelihood of dynamic rupture propagating over a stepover, at least as characterized from the surface observations, and other abrupt fault-geometry changes such as bends and branches, is an important ingredient in the current methodologies for seismic hazard estimation (Field et al., 2014, 2015, 2017; Shaw et al., 2018), and hence continues to be an active area of study. Later compilations of field data from a large number of surface-rupturing events (Biasi and Wesnouski, 2016, 2017) indicated that the size of the fault stepover capable of being passed by dynamic rupture is both magnitude- and mechanism-dependent. This is consistent with more recent 3D dynamic rupture simulations that indicate the dependence of the answer on the initial stress distribution and frictional parameters, the dynamics of the resulting events, and the potential connections between segments at depth (Oglesby, 2008, Lozos et al., 2014). 2D dynamic models of the dynamics of branched faults (Kame et al., 2003) indicate analogous effects: the branch that can be taken by rupture is determined by the interaction - modulated by the branching angle - of the dynamic stress field (which depends on the rupture dynamics) with the regional pre-stress field.

These relatively simple examples highlight the challenges of properly capturing dynamic rupture behavior even at a single fault discontinuity and trade-off that arise from considering a single dynamic event in isolation.

**SEAS-like simulations for multi-segment faults**

More comprehensive understanding of the behavior of a fault system will thus come from accounting for the effects of the slip history of the system, including prior earthquakes. This need leads to the development of different schemes for multi-cycle, SEAS earthquake models for geometrically complex, multi-segment faults, building on the single-segment SEAS models discussed in section MMC2.2.

There are several challenges in the modeling of earthquake faults over multiple instances of earthquakes and aseismic slip. A key physical challenge in such models is the interseismic loading and associated inelastic processes in the bulk. Long-term loading at the external boundaries of a purely elastic model (e.g., Cooke and Dair, 2011) leads to a physically high stresses and energy stored at fault discontinuities (Dieterich and Smith, 2009). Hence the modeling must include inelastic processes, which in nature take a
number of forms, from off-fault seismicity to bulk damage (MMC1.2). Some methods start with rigorous
dynamic rupture modeling, which fully accounts for inertial effects during the simulated earthquakes, and
model the interseismic loading, inelastic processes, and re-nucleation in an approximate way, e.g., using
a viscoelastic relaxation method for inelasticity (Duan and Oglesby, 2005; 2006; 2007; Duan et al., 2019;
Figure II-MMC2.3). Such models reproduce earthquakes of multiple sizes that can be examined
statistically, e.g., for the likelihood of propagation through a stepover. Results may be compared with field
observations of slip rates, paleoseismic observations of event patterns, and recurrence intervals,
constraining model parameters and informing field and paleoseismic studies. This approach has mainly
been applied to multi-segment strike-slip fault systems in two dimensions.

Earthquake simulators over fault networks

Another class of approaches, dubbed earthquake simulators, aims to simulate sequences of earthquake
events over regional-scale networks for tens of thousands of years or longer (Dieterich and Richards-Dinger, 2010; Pollitz, 2012; Richards-Dinger and Dieterich, 2012; Tullis et al. 2012a,b; Sachs et al., 2012; Ward, 2012; Shaw et al., 2018). The significant accomplishment of these approaches is in
successfully taking on a challenge of incorporating a complex fault network. To achieve that feat,
earthquake simulators adopt significant simplifications both to the physical processes considered and to
the solution procedures. Let us use one of the most advanced simulators, RSQSim, as an example
(Dieterich and Richards-Dinger, 2010; Richards-Dinger and Dieterich, 2012; Shaw et al., 2018) (Figure II-
MMC2.3). RSQSim considers the fault network to be embedded into an elastic half-space and each fault
to be loaded by a back-slip-like procedure, in which stresses on fault elements increase if they fall behind
the long-term slip rate prescribed for each fault. The interseismic processes, such as aseismic slip or
deformation on deeper fault extensions, are not included in the regional scale simulations (Richards-
Dinger and Dieterich, 2012; Shaw et al., 2018) although they could be incorporated in principle. The faults
interact by static stress changes due to co-seismic slip. Instead of the direct solution of the governing
(equilibrium) equations, which would be intractable, RSQSim employs an ingenious algorithm to update
the state of fault elements from locked to nucleating to rupturing using analytical rate-and-state-based
equations derived from single-degree-of-freedom systems (Dieterich, 1992; 2007). With this procedure,
the simulator inherits the time-dependent nucleation of the rate-and-state formulation that allows it to
reproduce the observed Omori-like decay of aftershock sequences (MMC3.5), an important requirement
for any simulation over a fault network. During a simulated earthquake event, a semi-kinematic
procedure is employed which is partially based on rate-and-state friction and partially designed to mimic
simulations of dynamic rupture with linear slip-weakening friction, employing additional parameters such
as dynamic stress overshoot factor (Richards-Dinger and Dieterich, 2012). So while the inertial effects are
not explicitly included, the simulator can mimic ruptures produced by fully dynamic rupture simulation with
linear slip weakening over a 3D planar fault with heterogeneous strength and other comparisons with
dynamic rupture simulations are ongoing. RSQSim has been successful at reproducing Gutenberg-
Richter earthquake statistics and the Omori decay of aftershocks with time (Dieterich and Richards-
Dinger, 2010; Richards-Dinger and Dieterich, 2012) and it has been used to construct hazard maps that
compare favorably with those determined by current methodologies for seismic hazard estimation (Field
et al., 2014, 2015, 2017; Shaw et al., 2018).

Such simulators can be used as a valuable research tool for understanding the effects brought by the
fault network complexity. They can examine the role of smaller faults for the properties of the largest
earthquakes that occur; study uncertainty of the problem by finding all models that match certain
statistical properties; and investigate the impact of imprecise knowledge of fault geometry.

Earthquake interaction, identifying dominating mechanisms
In reproducing earthquake sequences over fault networks, the question of earthquake interactions comes to the fore. Earthquakes clearly affect one another as evidenced by aftershock sequences and earthquake clustering. A number of physical mechanisms have been proposed to capture these effects. For example, the aftershock sequences and their properties, including the Omori’s law for their decay, have been attributed to static stress changes imposed by mainshock and rate-and-state effects (Dieterich, 1994; Gross and Bürgmann, 1998; Toda et al., 1998, 2005; Gomberg et al., 2005); increased loading rate due to aseismic processes such as postseismic slip (e.g., Benioff, 1951; Perfettini and Avouac, 2004), relaxation of the viscoelastic lower crust (e.g., Freed and Lin 2001), pore fluid motion and induced variations in fault strength (e.g., Nur and Booker, 1972; Bosl and Nur, 2002), triggering due to dynamic stress changes (e.g., Hill et al. 1993; Gomberg et al., 2003; Felzer and Brodsky, 2006), and evolution of viscoelastic damage rheology due to sudden increase in strain (e.g., Ben-Zion and Lyakhovsky 2006).

In addition to the static stress transfers due to co-seismic slip, two effects can be particularly important. One is the loading of the seismogenic faults by the deeper aseismic shear slip or deformation (section MMC1.3) which concentrates the stress at the bottom of the seismogenic layer (Figures II-MMC2.1, II-MMC3.1-2) promoting earthquake nucleation there. Such loading would be variable throughout the interseismic period, with more rapid postseismic slips and lower slip rates at other periods. Furthermore, postseismic slip fronts can carry larger stress changes along the fault than the direct coseismic stress change (Lui and Lapusta, 2016).

The second one is the dynamic triggering of nearby and distant earthquakes by seismic waves. The 1992 M7.3 Landers, California earthquake provided glimpses of the significance of dynamic triggering at a distance (e.g., Hill et al., 1993), and was among the first to convince seismologists that distant dynamic triggering is real. Earthquakes up to magnitude 5 (e.g., Gomberg et al., 1994; Gomberg et al., 2001) were triggered 100’s of km away from the M7.3 Landers source region. Detections of distant dynamic triggering of small earthquakes and effects on hydrological systems and rock properties have been repeated numerous times (Brodsky and Prejean, 2005; Manga et al., 2012; Nakata and Snieder, 2011; Vidale and Li, 2003; Xue et al., 2013; Hill and Prejean, 2014). Whereas static stress triggering (e.g., Harris and Simpson, 1992; King et al., 1995) is easy to calculate and its effects are concentrated in the near-field, dynamic stress triggering occurs both close-in (e.g., Harris et al., 1991; Kilb et al., 2000; Felzer and Brodsky, 2006) and, at great distance through propagation of surface waves. Dynamic stress triggering has been particularly noted as a trigger of slow slip events on conditionally stable portions of faults (e.g., Peng et al., 2008).

Models that study dynamics of fault systems need to account for the dominating mechanisms of such interactions to be credible in their output in terms of earthquake sequences. Hence it is important to combine a number of such mechanisms in the same simulation, to evaluate their relative effects. It is quite possible that two or more of these mechanisms have comparable importance and hence need to be considered together; for example, postseismic slip on deeper fault extensions can produce significant loading at the bottom of the seismogenic zone and promote the nucleation of an event in addition to any other interaction mechanism. The current simulators over regional fault networks are limited in that regard; e.g., RSQSim only incorporates static stress changes due to co-seismic slip. So the task of determining the dominating mechanisms falls to SEAS simulations over heterogeneous faults or small-scale fault networks, and it is important to continue develop such simulations.

**Uncertainty in fault geometry and properties**
An important additional challenge is the full characterization of the fault geometry and the physical properties of the faults and their surrounding materials. The aspect ratio and seismogenic depth of a fault can have a strong effect on its ability to allow through-going rupture at a stepover (Bai and Ampuero, 2017), on top of the well-known effects of stepover width and separation along strike (Harris et al., 1991; Harris and Day, 1993; 1999). The continuity of the fault at depth may be different from that at the surface, with significant implications for through-going rupture (Aochi, 2003). The functional dependence of the shear stress on normal stress, as well as the value of frictional parameters, can strongly affect the ability of rupture to propagate across bends and branches, stepovers, and to the surface of dip-slip faults (Meng et al., 2012; Ryan and Oglesby, 2017). The assumed processes and properties for the interseismic relaxation of stress concentrations in the bulk can have a controlling effect on the amount of stress that accumulates around discontinuities (Duan and Oglesby, 2005ab; 2006; 2007). These effects are compounded over multiple earthquake events and interseismic periods. This reality highlights the need to quantify the epistemic (model) uncertainty by modifying the fault/bulk geometry and properties within the allowable bounds.

**Future progress**

The current state of earthquake modeling on geometrically complex fault systems is that different modeling methods aim to capture different aspects of this challenging problem, each with its own unique strengths and weaknesses, associated with the sacrifices each makes for computational tractability. Systematic progress going forward will come from a number of developments, including employing justifiably simplified, scale-appropriate constitutive relations developed by smaller-scale model (MMC1 and MMC2.1-2.3); verifying larger-scale models by comparing them to more detailed smaller-scale models at the scales that both can resolve; and combined simulations where several numerical approaches are coupled with each other and, potentially, with machine learning techniques (MMC2.5).

**Key future goals**

- Continue to develop multi-physics dynamic rupture simulations on complex geometries to (i) advance our understanding of the effects of fault geometry on fault slip, energy budget, and strong ground motion, and (ii) study rupture interaction through fleeting but substantial dynamic stress changes carried by waves;
- Use SEAS modeling on single and several faults to study various proposed interaction mechanisms of earthquake rupture events, including static stress changes due to events themselves, static stress changes due to postseismic and interseismic creep/deformation, dynamic stress changes, and fluid effects including poroelastic stress changes; quantify their effects and relative importance at different distances and temporal scales; and determine dominating mechanisms that need to be included in regional-scale simulations;
- Develop methodologies for regional-scale earthquake simulators over fault networks that include simplified representations of multiple effects identified by more detailed studies; use them as research tools to understand which features of network geometries are key to include and how the model response changes for a range of parameters that are not well-constrained, such as properties of many secondary faults.

**MMC2.4 Modeling tectonics and earthquakes: variability in loading and geometry**

Tectonic processes shape many key ingredients governing earthquakes, including fault geometry, stress, strength, pore fluid pressure, and mechanical fault zone and bulk properties. In turn, earthquakes affect these controlling variables and ultimately feedback to shape tectonic features. Recent numerical advances initiated the bridging of these processes, such that both long-term subduction and short-term seismic cycle dynamics can be simulated on time scales from millions of years down to several years.
Such originally geodynamic models have recently been further developed to resolve afterslip and power law creep within the postseismic phase in complex subduction zones (Sobolev and Muldashev, 2017) and inertial dynamics and seismic wave propagation in planar and evolving strike-slip settings (Herrendörfer et al., 2018; Preuss et al., 2019). SEAS models have also been extended to include longer-term features, such as ductile shear zones and mantle flow (e.g., Lambert and Barbot, 2016; Allison and Dunham, 2018; Barbot, 2018). The development of these tectonic-earthquake sequence models (TEC-SEAS; Figure II-MM2.4) opens new research opportunities to improve our understanding and modeling of earthquake source processes in the next decades.

Stress, strength and rheology are influenced by thermal, hydrological, tectonic, lithological, petrological, and chemical, controls. These controls and feedback mechanisms can all be modelled and understood through TEC-SEAS simulations. In particular, tectonic evolution and fault zone and bulk variations in lithology, rheology, temperature, pore fluid pressure, and their thermodynamics can be conveniently solved for (e.g., Gerya and Yuen 2007; Yarushina and Podladchikov, 2015). This, in combination with the recent progress of modeling earthquake slip at fractions of a second, makes TEC-SEAS models well suited to contribute to many of the challenges outlined in section MMC1 and MMC2.2. In addition they open new opportunities to study complementary aspects through modeling fault evolution, realistic tectonic loading, fault zone structures and tectonics.

Incorporating fault geometry and its evolution

Fault geometry is one of the key variables controlling the size and recurrence of earthquakes, since it determines both potential rupture length and width and tendency to rupture through both proximity to other faults and orientation with respect to the stress field (e.g., Herrendörfer et al., 2015). Large-scale geometrical variations control interseismic coupling, loading rate, off-fault deformation, and long-term earthquake patterns (e.g., Dal Zilio et al., 2019). Geometrical irregularities, such as seamounts and fracture zones, affect earthquake location and size, interseismic coupling, aseismic creep and overriding plate deformation (e.g., Wang and Bilek, 2014). Geometrical variations and irregularities and their impact on bulk deformation are thus important to incorporate more accurately.

The evolution of fault networks dictates the geometry and connectivity of fault systems. Evolving fault simulations could help to decipher the location of unknown faults; the depth- and lateral structure and maturity of known faults; and their connectivity. These control the style and properties of fault slip (e.g., Perrin et al., 2016) and the loading rate due to slip or strain partitioning within a network of faults (e.g., Daout et al., 2016). Especially immature faults, such as the San Jacinto fault zone, are potentially susceptible to short-term geometry and stress field changes (e.g., Preuss et al., 2019.). This could affect seismic and aseismic slip characteristics. In addition, the spontaneous evolution of faults allows for a self-consistent assemblage of fault geometry, stress and strength under which evaluation of fault path selection can be assessed better (e.g., van Zelst et al., 2019).

Incorporating realistic tectonic loading

Fault stress and shear strength are key controlling parameters that govern fault slip. Besides variations due to preceding and neighboring events, these are influenced by tectonic loading and fluid flow. Tectonic loading is typically considered through a back-slip approximation or as a far-field boundary condition of accelerated or realistic velocities. However, especially in complex or convergent tectonic settings, this may not be valid (e.g., Conrad et al., 2004). For example, in a slab-controlled setting such as the Northern Apennines, minor variations in slab buoyancy or variations in lower crustal rheology are observed to significantly affect upper crustal stress seismicity types and rates (D’Acquisto et al., 2018).
This indicates that typically neglected aspects relating to subduction zone dynamics, deeper rheology, temperature and architecture may need to be considered. These processes also affect the release and flow of fluids through fault zones and the bulk. In combination with metamorphic reactions affecting fluid absorption and permeability, these affect long-term and transient pore fluid pressures and thus fault strength. Including such long-term processes ensures a more accurate representing of stress build-up and transfer and strength variations, which could have a distinct impact on fault slip.

**Incorporating fault zone and bulk structure and rheology**

Stress loading and fault stability are affected by fault and bulk structure and rheology, and heterogeneities within them (e.g., Barbot and Fialko, 2010; Naliboff et al., 2012; Kimura et al., 2012; Lotto et al., 2018; van Zelst et al., 2019; MMC1). Fault normal (and parallel) variations in lithology and rheology can be effectively incorporated and analyzed in TEC-SEAS models using a unified on- and off-fault rheological framework (e.g., Lyakhovsky et al., 2011; Herrendörfer et al., 2018). The need for this reintegration of brittle and ductile behaviour is supported by observations that brittle and ductile behaviour occur throughout fault and shear zones from the surface into the asthenosphere (e.g., Huntington et al., 2018). Such formulations in combination with large-strain simulations can help to interpret the wealth of geological records; bridge the gap between laboratory, field observations and theory; and develop and verify upscaling of constitutive relations.

**Investigating the link between long-term and earthquake-related deformation and processes**

Many cycles of inter-, co-, and postseismic deformation build impressive geological structures, such as mountain chains and peninsulas. TEC-SEAS models provide a self-consistent framework that can help understand how geological structures form and how the characteristics of future earthquakes are influenced by long-term tectonics. The location of fore-arc basins, the coast line, or mountainous topography could reveal lateral and in-depth seismogenic properties of the megathrust interface (e.g., Song and Simons, 2003; Wells et al., 2003; Fuller et al., 2006; Saillard et al., 2017). Additionally, understanding how much each period contributes to anelastic deformation provides information on how much of geodetically measured strain could be released in upcoming earthquakes (e.g., Meade, 2010; Rollins et al., 2018). Finally, such models can help to understand the interaction and feedback of earthquakes with volcanism, metamorphic reactions, and the strength of faults.

**Scientific and numerical issues to be addressed**

The implementation and formulation of such tectonic and rheological features can partially be achieved through exploiting expertise within existing geodynamic models and requires dedicated development. Once equipped, TEC-SEAS models, self-consistently simulating the dynamics over a variety of time scales, are suited to tackle a range of scientific issues discussed in I.3. In particular, such models could contribute to understanding spatio-temporal patterns of earthquake and, slow aseismic slip within a fault zone or specific region; differences in fault slip or strain between different tectonic environments; physical mechanisms controlling intermediate- and deep-depth earthquakes and spectrum of aseismic slip; the nature, distribution, and evolution of asperities and barriers on fault interfaces; causes of lateral and depth-dependent fault segmentation, and variations of interseismic locking through space and time. These challenges should be addressed done through systematically analyzing the self-consistent feedback between described processes and along-strike variations in stress, fluids, geometry, overriding plate and lithological structures and phase transitions. Moreover, a physically consistent, spontaneous framework introduces the opportunity to discover new processes and features, which can later be
evaluated by observations (such as a secondary zone of uplift due to megathrust earthquakes; van Dinther et al., 2019).

At the same time, significant computational challenges remain. Due to numerical tractability, the modeling is currently 2D and sometimes resolves earthquake slip over years (Figure II-MMC2.4), although some problem settings allow to resolve fault slip down to minutes (Sobolev and Muldashev, 2017) and even milliseconds for a planar non-evolving strike-slip fault over several earthquake cycles (Herrendoerfer et al., 2018). Running 2D computations while bridging from millions of years to milliseconds as well as extending the simulations to 3D are some of the challenges to be tackled in the future. Efficient computational solutions for realistic and complex settings in 2D and 3D will require collaborations with supercomputing scientists and mechanical engineers. Another promising avenue is the development and exploitation of coupled approaches with fully dynamic simulations (van Zelst et al., 2019; section MMC2.5).

Key future goals
- Develop 2D and 3D TEC-SEAS simulations that can resolve relevant processes across tectonic and earthquake scales, hydro-thermo-mechanical coupling, unified fault and bulk rheologies, and inertial effects or approximations thereof in a physically consistent, accurate, and computationally efficient manner;
- Employ TEC-SEAS simulations to understand the spatiotemporal variability of fault locking and slip/deformation, including tectonic, rheological, hydrological, thermal and chemical controls, and their feedback mechanisms;
- Use these simulations to constrain and predict fault, crustal, lithospheric and mantle deformation and causative mechanisms thereof, understand differences between tectonic regimes, interpret geological observations across scales, verify upscaling of heterogeneous fault zones, and understand long-term seismicity behaviour of different regions.

MMC2.5 Frameworks for coupling models at different scales

To bridge the wide range of temporal and spatial scales involved in earthquake source processes (section I.3), one promising - although numerically demanding - approach is to couple numerical computations at different scales. Major challenges are posed by (i) implementation choices for coupling methods - that solve different equations in different numerical implementations - in a physical meaningful way, and (ii) establishing workflows across communities working on different scales.

Two-way coupling of codes for interseismic and coseismic periods, application to simulations of fault networks

Such coupling has been successfully used in a number of studies to simulate long-term fault slip. Some of the earliest long-term fault slip simulations employed a switch from a quasi-static numerical modeling over the relatively long rupture nucleation period to a dynamic rupture simulation (Shibazaki and Matsura, 1992), an approach later extended to simulations of multiple sequences of earthquakes and interseismic periods, with inelastic bulk effects (Kaneko et al., 2011) (MMC2.2). Such two-way coupling both enables the simulation of dynamic events with the initial conditions provided by the interseismic simulation, and explores the effects of seismic events on the interseismic processes (slow fault slip, viscoelastic rock creep).
This approach can be extended to simulations of fault networks, in which a simulations of interseismic loading of a fault network is coupled to dynamic rupture simulations only on specific, activated, segments.

**One-way coupling of modeling in the subduction zone environment**

To understand the effects of long-term tectonic processes on the features of dynamic events and potential tsunami in subduction zones, one can start with one-way coupling of the developed 3D models, allowing direct model verification with available observations. Such one-way coupling, characterized by neglecting feedback mechanisms, may include the following.

1) Long-term geodynamic modeling (MMC2.3), currently specifically suitable for subduction zones, evolving over millions of years and featuring spontaneously developing faults; this approach would provide fault geometries, viscosity structure, and temperature.
2) Models of sequences of earthquakes and aseismic slip/deformation (MMC2.2) from which would come predictions of geodetic deformation, slow slip, earthquake recurrence interval (some of these permitting comparison to observations), and fault initial conditions prior to earthquakes;
3) Dynamic rupture modeling (MMC2.1) which accounts for inertial dynamics but are limited to single events, which would use the fault initial conditions from which would come predictions of strong ground motion (seismic waves) and surface or seafloor deformation (Figure II-MMC2.4);
4) Tsunami models, which would provide predictions of wave heights that can be compared to observations.

The coupling of all of these modeling approaches will enable the inclusion of appropriate constitutive relations (MMC1) for frictional and inelastic constitutive behavior governing earthquake source physics on- and off-fault during long-term fault slip and dynamic rupture, as these modeling approaches are being developed to contain the most suitable physical ingredients (MMC2.1-2.3).

**Applications of machine learning**

Replacing explicit modeling components with machine learning holds great potential for bridging scales and accelerating computations (see also III-IN2). Examples of potential applications include: mapping between distributions of pre-event fault conditions and dynamics of the resulting seismic event on a certain scale; deep networks trained to discover rules for evolving the displacement field and friction during inter-seismic deformations to accelerate simulations of interseismic periods; deep learning used to approximate dynamic rupture calculations, in equivalence to ongoing efforts in developing pseudo-dynamic source descriptions (e.g., Guatteri et al., 2004, Schmedes et al., 2013, Mai et al., 2017).

Such an approach can, for example, enable indirect inclusion of inertial effects in simulations of multiple events over regional-scale fault networks. To this end, a training database may be established using outcomes from a suite of dynamic rupture simulations, and used to train a suitably selected neural network algorithm to map pre-event fault conditions into dynamic rupture outcomes. The longer-term interseismic simulation would then call upon the trained neural network to generate dynamic rupture scenarios as needed. Based on promising findings in few dynamic rupture simulations relating the state of stress and earthquake source properties based on fracture mechanics theory (e.g. Schmedes et al., 2010, Bizzarri et al., 2012; Gabriel et al., 2013), such machine-learning algorithms are expected to shed further light on relationships between rupture characteristics, such as rupture speed and slip rate, and initial conditions related to prestress and geometry. Such approaches can be verified by comparison with dynamic rupture simulations.
Key future challenges:
● Create efficient and physically meaningful interfaces for coupling modeling approaches at different scales;
● Explore the feasibility of replacing explicit modeling components with machine learning outcomes;
● Establish workflows across communities working on different scales.

Cross-cutting themes in MMC2.1-2.5
● Using earthquake source models for systematic evaluation of all relevant physical mechanisms at the scales of interest, establishing the relative importance of different mechanisms, and identifying justifiable simplifications;
● Coupling earthquake source models at different scales with each other and/or machine learning;
● Need for supercomputing facilities, code development with computer scientists and software engineers, and code verification.
MMC3. Identifying relevant modeling ingredients by interpreting and improving a range of observations

MMC3.1 Low-heat, low-stress operation of mature faults but not the rest of the crust; thermo-mechanical coupling

The state of stress and pore pressure in Earth’s lithosphere controls the distribution of seismicity, earthquake stress drops and ground motion, structure of mountain belts and other topographic features, and even the viability of mantle convection and plate tectonics. Stress and pore pressure, whether directly measured or indirectly inferred, constitute fundamental constraints on earthquake source models. Assuming the brittle upper crust that hosts the seismogenic faults is in a frictionally critical stress state, the stress profile can be constructed that depends on the assumed distribution of pore pressure and the faulting regime, the latter controlling relative magnitudes of the principal stresses. Stress profiles constructed in this manner typically use Byerlee friction coefficients between 0.6 and 0.9 (Byerlee, 1978), which are consistent with low-velocity friction measured measured in the laboratory for most rock types (Dieterich, 1979ab, 1981a; Tullis and Weeks, 1986; Blanpied et al., 1991, 1995; Marone, 1998a) (Figure II-MMC1.1-1). Stress tensor components can then be resolved onto fault surfaces of various orientations to make predictions of which faults are active, based on proximity of the resolved shear and effective normal stress to a frictional failure condition. This exercise is the basis of the classic Andersonian theory of faulting (Anderson, 1951).

Multiple lines of evidence for low-heat, low-stress operation of mature faults

The resolved shear stresses on major plate boundary faults (Figure II-MMC3.1-1), most notably the active, mature plate-boundary San Andreas Fault (SAF), lie considerably below frictional failure of ~100 MPa calculated with the friction coefficients of 0.6-0.9 and effective normal stresses equal to overburden minus hydrostatic pore pressure at representative seismogenic depths (Townend and Zoback, 2004; Zoback et al., 2007). Likewise, for thrust-belt wedges, their extremely low dip angle, interpreted through critical taper wedge theory, again indicates their slippage at low shear stress (Suppe, 2007). The inconsistency between these observations and Andersonian faulting theory with Byerlee friction and hydrostatic pore pressure is termed the stress paradox.

Complementing the stress paradox is the heat paradox. Temperature in the lithosphere varies primarily with depth, and the simplest thermal models, based on the one-dimensional (1D) heat conduction upward from the convecting mantle with additional radiogenic heating in the crust, provide the ambient 1D geotherm. Shear heating around active faults creates a thermal anomaly or temperature increase above the ambient geotherm. Shear heating comes from frictional slip (both seismic and aseismic), off-fault plastic strain in the brittle upper crust, and viscous flow in the ductile lower crust and upper mantle. The predicted thermal anomaly has a width of the order of 10 km around faults (Lachenbruch and Sass, 1973, 1980). Shear heating and the resulting thermal anomaly depend on the absolute level of deviatoric or shear stress resisting slip and inelastic strain during deformation. Yet heat flow measurements adjacent to the SAF provide no evidence of a thermal anomaly, which together with measurement uncertainties places an upper bound on resistive shear stress of ~10 MPa (Brune et al., 1969; Henyey and Wasserburg, 1971; Lachenbruch and Sass, 1973, Lachenbruch, 1980). This is far less than an average value of ~100 MPa predicted by Byerlee friction and hydrostatic pore pressure, a discrepancy known as
the heat flow paradox. It must be recognized that heat flow constraints apply not only to the seismogenic zone but also the ductile fault root, as demonstrated by models that account for viscous shear heating in fault roots while neglecting frictional shear heating entirely (Takeuchi and Fialko, 2012). Note also that the heat flow constraint is independent of the fault shear zone width, because relevant time scales are vastly larger than thermal diffusion times across even the broadest possible shear zones. Some skepticism exists regarding the relevance of the heat flow data, derived from temperature measurements in shallow boreholes, because near-surface groundwater flow (i.e., thermal convection) might flush heat outward from the fault more efficiently than thermal conduction alone (Scholz, 2000, 2013) but there are also counter arguments (Lachenbruch and Sass, 1973, 1980).

Complementary constraints on frictional shear heating come from measurements of thermal anomalies at the 1-10 m scale in boreholes that were rapidly drilled across faults following large earthquakes (Tanikawa and Shimamoto, 2009; Fulton et al., 2013; Yang et al., 2016). These measurements provide constraints on local coseismic heat production and associated shear stress during sliding. As with regional heat flow data, these borehole measurements suggest low coseismic shear stress.

Another aspect of the fault heating problem, which is highly sensitive to shear zone width (in contrast to the constraints reviewed previously), is the possible paucity of melt signatures in exhumed faults. Frictional melt, or pseudotachylyte, is found in the core of some, but not many, faults (Sibson, 1975). Its absence bounds maximum temperature to lie below the solidus of minerals within the fault core. The maximum temperature rise is determined by both the resistive stresses during seismic shear and the width of the shear zone. Many field, laboratory, and modeling studies argue for the extreme localization of seismic shear (section MMC1.1), in which case deformation must occur with minimal resistance to prevent macroscopic melting. Controversy exists, however, because melt products called pseudotachylytes are challenging to identify in the field and may be unstable and not preserved through the exhumation process (Sibson and Toy, 2006; Rowe and Griffith, 2015). At the same time, rapid post-earthquake drilling, while limited to shallow depths, provides no evidence for pervasive macroscopic melting that would be expected from highly localized coseismic shear against high stresses (Tanikawa and Shimamoto, 2009; Fulton et al., 2013; Yang et al., 2016).

Taken together, the observations reviewed above argue for low shear resistance during both earthquake ruptures and slow creep on the SAF and other major faults.

Potential explanations for the low stresses on mature faults and role of modeling

Two main hypotheses have been advanced to explain these observations.

The first hypothesis is that these faults are always weak, both coseismically and interseismically. This might be due to elevated pore pressures throughout the earthquake cycle, in which case friction could be consistent with Byerlee values. Possible causes of fault overpressure include migration from ductile roots in the lower crust, crustal hydrology and faults as a preferred pathway for fluid migration; decrease in permeability with increasing effective normal stress; subduction of hydrated marine sediments and dehydration reactions (Sibson et al., 1975; Hyndman and Shearer, 1989; Rice, 1992). Another explanation is that pore pressure is close to hydrostatic but friction (both static and dynamic) is low. Certain gouge compositions, particularly clays and other foliated minerals, can have extremely low friction values (Moore and Rymer, 2007; Lockner et al., 2011).
The second hypothesis is that faults are only dynamically weak, that is, they have ambient pore pressures close to hydrostatic and Byerlee-level static friction coefficients, but weaken dramatically during coseismic shear by reduction in friction coefficient and/or transient elevation of pore pressure. Laboratory experiments at high sliding rates provide substantial evidence for dynamic weakening, attributable to a wide range of processes including thermal pressurization, flash heating, thermal decomposition and associated thermal pressurization, nanoparticle lubrication, silica gel formation, and macroscopic melting (section MMC1.1; Figure II-MMC1.1-1). With dynamic weakening, earthquakes nucleate within localized regions of elevated shear stress or reduced strengths, and then propagate outward into understressed regions due to dynamic weakening (Zheng and Rice, 1998; Lapusta and Rice, 2003; Noda et al., 2009; Schmitt et al., 2015), the process that can also be influenced by off-fault plasticity (Dunham et al., 2011a; Gabriel et al., 2013) and fault geometric complexity (Fang and Dunham, 2013).

It is possible that both of these hypotheses are correct, but for different faults or fault sections; they can also act in concert, enhancing the interface weakness. For example, the creeping section of the SAF is arguably weak because its core contains weak clays derived from its unique geologic history and inheritance of subduction-derived minerals (Moore and Rymer, 2007; Lockner et al., 2011). In contrast, seismogenic sections might be weak due to dynamic weakening (Rice, 2006; Noda et al., 2009; Jiang and Lapusta, 2016, 2017). And subduction plate boundary faults are likely to have overpressured fluids (Liu and Rice, 2007; Saffer and Tobin, 2011; Shelly et al., 2006) (Figure I-5), but might also be susceptible to dynamic weakening (Faulkner et al., 2011; Cubas et al., 2015).

Earthquake source modeling can make testable predictions that might distinguish between the two hypotheses for the low-stress and low-heat operation of major faults. The ambient stress level on faults and its proximity to static frictional strength influence the susceptibility of the fault to creep when subject to stress perturbations, say from nearby earthquakes, and to produce microseismicity in portions of the fault containing a heterogeneous mixture of velocity-weakening and velocity-strengthening properties. Faults that are always weak (first hypothesis) should be more sensitive and microseismically active than faults which have average stresses well below their static strength (second hypothesis) (Figure II-MMC3.1-2).

An additional opportunity for probing fault stress levels, stress heterogeneity, and proximity to failure is offered by statistical features of seismicity triggered by fluid injection or extraction (section III-IN3, V-MMC3.2), for which the forcing is relatively well known. The rate at which microseismic events occur as stress and/or pore pressure is altered depends on the (heterogeneous) background stress state relative to static strength (Maurer and Segall, 2018).

**Potentially high-stress operation of other active, less mature faults; relation to structural complexity**

Theories for stress levels on active faults must also account for the fact that many active faults are apparently operating at Byerlee friction levels (Sibson, 1985; 1994; Zoback, 1992; Townend and Zoback, 2000; Zoback and Townend, 2001; Chery et al., 2004; Zoback et al., 2007; Hurd and Zoback, 2012; Schoenball and Ellsworth, 2017). The concept of a critically stressed crust argues that deviatoric stresses are bounded by the frictional strength of pre-existing faults and fractures of all orientations (Townend and Zoback, 2000). If all such faults are permitted to slip at levels of shear stress as low as the nearby mature faults, e.g., the SAF, then crustal deviatoric stress levels would be similarly bounded. Yet stress measurements in deep drilling projects consistently find hydrostatic pore pressures and deviatoric stresses consistent with Byerlee friction (Townend and Zoback, 2000). Similarly, substantial reverse faulting and folding occurs on structures parallel to the SAF (Zoback et al., 1987), in a manner consistent with the crust adjacent to the
fault being critically stressed at Byerlee friction levels (Chery et al., 2004; Zoback et al., 2007). Finally, natural and induced seismicity in the central United States is systematically occurring on faults with Byerlee levels of friction, using stress orientations from focal mechanisms, borehole measurements, and Mohr-Coulomb failure analysis (Zoback, 1992; Hurd and Zoback, 2012; Schoenball and Ellsworth, 2017).

Why do these faults operate at Byerlee stress levels, in contrast to the low stress operation of major plate-boundary faults? The explanation is likely due to their higher degree of structural complexity, although the exact physical mechanism needs to be investigated. The simplest hypothesis is that such faults cannot experience enhanced dynamic weakening, because the extreme shear localization, a necessary condition for most dynamic weakening mechanisms, is prevented by a higher degree of structural complexity on these less active, non-plate-boundary faults. The faults might lack well-developed highly localized cores (MMC1.1) that are continuous along the fault. The alternative hypothesis is that the shear localization and associated dynamic weakening lead to extremely low shear resistance in most faults with seismic slip, but the structural complexity, in the form of fault surface roughness at scales larger than slip, branching, step-overs, and anastomosing strands of localized shear, leads to the additional resistance associated with straining around structural complexities, the latter termed backstress or roughness drag. This conceptual idea, emerging from work on fractally rough faults (Dieterich and Smith, 2009), was confirmed and quantified using rough fault dynamic rupture simulations with off-fault plasticity and on-fault dynamic weakening (Fang and Dunham, 2013). Their simulations suggest that mature faults like the SAF, having smooth and continuous slip surfaces, have negligible roughness drag and thus operate at low stresses, while less mature and more structurally complex faults operate at high stresses with resistance dominated by roughness drag. Additional modeling studies are required to resolve the full range of roughness wavelengths, explore other forms of structural complexity, and explore the consequences of both possibilities for other observables such as earthquake energy budget and microseismicity generation, in order to fully test these ideas.

**Key future goals:**

- Employ SEAS models (MMC2.2) to investigate constraints that the observed low-heat, low-shear stress operation of mature faults - together with levels of microseismicity and other observations - puts on fault physical properties including pore pressure, potential enhanced dynamic weakening of the friction resistance, as well as degree of roughness and heterogeneity;
- Investigate the differences between the mature and immature faults in that regard that can be tested through observations;
- Use the results to inform dynamic rupture simulations and investigate the potential impact on strong ground motion.

**MMC3.2 Spatio-temporal patterns of seismic/aseismic slip and distributed deformation**

The traditional view of fault slip involved two modes of fault behavior taking turns: locked conditions, enduring for decades to millennia during a period of elastic strain accumulation, and seismic rupture, catching up with long-term slip in a matter of seconds (Reid, 1910). In recent decades, this view has greatly evolved due to revolutionary advances in geodetic observational systems (Simons et al., 2002; Fialko et al., 2005; Konca et al., 2008; Perfettini et al., 2010; Hamling et al., 2017) that conclusively demonstrated the existence of aseismic slip in various environments, evidence of slow faulting in geologic outcrops, and the fundamental role of slow slip in experimental and mathematical explorations of faulting behavior (Avouac, 2015; Bürgmann, 2018). Nominally aseismic slip is evident from top to bottom of plate-boundary faults and plays an important role throughout the earthquake cycle. Slow slip can take the form
of nearly steady fault creep or occur in transient accelerations or slow slip events of varying duration and propagation speed (e.g., Ide et al., 2007). In part, episodic slow slip events have been inferred to occur at the bottom of the seismogenic zone in many subduction zones (Dragert et al., 2001; Schwarz and Rokosky, 2007; Rubinstein et al., 2010; McCaffrey et al., 2008; Beroza and Ide, 2011). There may be no exact boundary between seismic and aseismic modes of slip, and slow slip is often accompanied by repeating earthquake sequences, seismically observable tremor, and low-frequency earthquakes (Ellsworth and Dietz, 1990; Vidale et al., 1994; Nadea and Johnson, 1998; Bürgmann et al., 2000; Igarashi et al., 2003; Peng and Ben-Zion, 2005; Shelly et al., 2006, 2007; Chen et al., 2007; Brown et al., 2008, 2009).

We now recognize that many faults present a patchwork of locked and creeping patches (Figure II-MMC3.2). On the creeping sections, slow slip can be relatively steady or it can occur in the form of transient slow slip events. If fault sections accommodate a substantial portion of their slip budget by aseismic slip, their seismic potential is reduced. On the other hand, slow slip can accelerate loading on nearby locked fault sections and thus bring them closer to failure. The amount of slow slip is often small in any observational time period and thus slow slip is difficult to detect, but the changes in local stress that result from slow slip can be substantial and can trigger earthquakes. Thus, better understanding of aseismic slip appears essential for improved understanding of the dynamic earthquake process and the assessment of time-dependent earthquake hazard.

Complexity of the observed aseismic/seismic slip patterns

SEAS models of fault long-term behaviors (MMC2.2) that incorporate the important role of slow slip are informed by seismologic and geodetic observations of interseismic fault creep, spontaneous or triggered slow slip events, earthquake rupture, and postseismic afterslip (e.g., Avouac, 2015; Bürgmann, 2018, and references cited therein). Consider a recent example from the central Ecuador subduction zone (Rolandone et al. (2018) (Figure II-MMC3.2). Long-term interseismic GNSS measurements constrain the background coupling distribution on the megathrust. Several coupled sections of the fault slipped in spontaneous slow slip events accompanied by seismic swarms and repeating earthquakes. The Mw 7.8 16 April 2016 Pedernales earthquake ruptured a coupled section between 20 and 30 km depth. It was followed by afterslip up- and downdip of the rupture and triggered a slow slip event on a coupled segment about 100 km south of the mainshock rupture. A similarly rich spectrum of fault slip behavior has been observed in many of the worlds subduction zones and the partially coupled San Andreas fault near Parkfield (e.g., Barbot et al., 2012; Murray and Langbein, 2006; Shelly, 2017). In some cases, SSEs have been observed to either trigger or grow into large earthquake ruptures (e.g., Kato et al., 2012; Meng et al., 2015; Radiguet et al., 2016; Socquet et al., 2017). While in no way universal, these observations have renewed interest in studying and modeling precursory slow slip.

A growing number of observations indicate that a single fault might slip either seismically or aseismically in different circumstances, so a fault that is observed to creep should not be regarded as unable to host a large, damaging earthquake. Such behaviors have been inferred for various parts of the fault some of which slipped coseismically in the 2011 Mw 9.0 Tohoku earthquake (Ito et al., 2013; Uchida and Matsuzawa, 2013; Johnson et al., 2012; Perfettini and Avouac, 2014). Observations of this bimodal character are not unique to Japan (e.g., Bürgmann et al., 2002; Pritchard and Simons, 2006; Lin et al., 2013; Barnhart et al., 2016; Wech and Bartlow, 2014), and it has also been observed in laboratory experiments (Leeman et al., 2016; McLaskey and Yamashita, 2017). This bimodal behavior is likely responsible for certain unique seismic signatures, such as period doubling (Shelly, 2010ab) and repeating earthquake sequences that suddenly start or shut off in response to nearby large earthquakes (e.g.,
Hatakeyama et al., 2017). The unique seismic signatures can used to help constrain parameters of numerical models (e.g., Veedu and Barbot, 2016) for enhanced insight.

**Distributed deformation in the shallowest and deepest fault regions**

Slip near the bottom and below the seismogenic zone transitions into more distributed ductile shearing, involving crystal plastic deformation of some mineral phases (Sibson 1977; Yuen et al., 1978; Montesi and Hirth, 2003; Bürgmann and Dresen, 2008; Takeuchi and Fialko, 2012; MMC1.3). In the brittle-ductile transition zone, seismic slip, slow frictional slip and ductile shearing can coexist and alternate in space and time (MMC1.3). Considerable constraints on faulting processes and deep deformation mechanisms are offered by both field geology and geodesy (e.g., Freed and Burgmann, 2004; Takeuchi and Fialko, 2013). Models of long-term fault slip (MMC2.2) should connect with these constraints by making predictions of stresses, temperatures, amount of viscous strain, and the width of viscous fault roots, which are potentially constrainable by studies of foliation development and elongation of grains in mylonites, recrystallized grain piezometry, and geothermometry.

In the brittle upper crust, testable predictions can be made of the width of damage zones, intensity and location of distributed inelastic deformation, and its healing in the interseismic period, which can be compared with field geology studies quantifying microcrack density as a function of distance from the fault as well as seismic and geodetic inferences (e.g., Johri et al., 2014; Huang et al., 2016; MMC1.2).

**Microseismicity**

Microseismicity provides us with a unique a window into the stress environment, structure, and physical properties of the crust and their changes over time or due to large earthquakes. For example, the spatial distribution of seismicity provides important constraints on the geometry of fault zones and their variation with depth and along strike (e.g., Rubin et al., 1999; Waldhauser and Ellsworth, 2000; Waldhauser, 2001; Waldhauser and Schaff, 2008). Focal mechanisms of microseismicity and their changes can constrain the principal stress directions and assess their changes after large events (e.g., Hardebeck and Hauksson, 2001; Hardebeck and Shearer, 2003). Monitoring the migration of microseismicity is important for extraction of energy sources and carbon storage (e.g., Oye et al., 2013). Recently, dense seismic arrays provide precise locations of microearthquakes (e.g., Ross et al., 2017) and improved estimates of stress drop and rupture area (e.g., Denolle and Shearer, 2016; Fan and McGuire, 2018; Wetzler et al. 2018). These improvements in resolution have particularly highlighted the volumetric character of fault zones, leading to questions about whether aftershocks are dominantly occurring along the same fault surfaces as mainshocks, or instead within the surrounding damage zones (Ross et al., 2018).

Detailed studies of repeating earthquakes that occur on a number of faults have been used to study various aspects of earthquake physics and mechanics, including fault creeping velocities, postseismic slip, earthquake interaction, and stress drops (e.g., Ellsworth and Dietz, 1990; Vidale et al., 1994; Marone et al., 1995; Nadeau and Johnson, 1998; Beeler et al., 2001; Igarashi et al., 2003; Hickman et al., 2004; Imanishi et al., 2004; Matsubara et al., 2005; Allmann and Shearer, 2007; Chen et al., 2007; Dreger et al., 2007; Chen et al., 2010; Zoback et al., 2010; Abercrombie, 2014). Because of their short recurrence times and known locations, small repeating earthquakes present a rare predictable opportunity for detailed observation and study, and this has been exploited in San Andreas Fault Observatory at Depth (SAFOD) drilling project (e.g., Hickman et al., 2004; Imanishi et al., 2004; Zoback et al., 2010). One of the most interesting observations about small repeating earthquakes is the scaling between their recurrence time and seismic moment (Nadeau and Johnson 1998; Chen et al. 2007), which is much different from that of a simple conceptual model with circular ruptures, constant stress drop, and seismic slip equal to plate velocity multiplied by recurrence time.
Many large earthquakes are preceded by foreshocks [e.g., Jones and Molnar, 1976, 1979; Abercrombie and Mori, 1996; Dodge et al., 1995, 1996; Zanzerkia et al., 2003; McGuire et al., 2005; Bouchon et al., 2011, 2013; Kato et al., 2012; Brodsky and Lay, 2014], which are defined as smaller seismic events that occur within a certain distance in time and space to the main event. The physical mechanisms for foreshocks as well as their potential role in the nucleation of larger events are currently open questions. One point of view is that foreshocks are due to aftershock-like clustering of microseismicity, which occasionally results in a much larger event, as manifested by Epidemic Type Aftershock Sequence (ETAS) models (e.g., Helmstetter and Sornette, 2003). In this interpretation, the largest event is simply an aftershock of foreshocks and there is no special relationship between foreshocks and the nucleation process of a larger event. An alternative explanation is that foreshocks occur on fault patches loaded by the surrounding aseismic (slow) slip, potentially signifying the nucleation of an upcoming large event. Detailed studies in several areas have indeed shown that foreshocks require additional factors beyond typical aftershock interactions. In a foreshock sequence before the 1992 M 7.3 Landers earthquake, the foreshocks were too far to trigger each other by static stress changes, the typical explanation for aftershock interactions (Dodge et al., 1995), suggesting that another mechanism was also operational, perhaps aseismic slip. In foreshock and aftershock sequences on the East Pacific Rise transform faults, the foreshocks were too numerous in comparison with aftershocks to be explained by the same ETAS model (McGuire et al., 2005), suggesting that an additional factor was needed to trigger the foreshocks, such as slow slip transients. An analysis of foreshocks before large (M > 6.5) events from well-instrumented areas (Bouchon et al., 2013) found that there is significantly more seismicity before interplate events than before intraplate ones, with the difference perhaps being a slow nucleation process in the case of interplate events that triggers the foreshocks. Furthermore, observations from the 2011 M 9.0 Tohoku-Oki event (Kato et al., 2012) and 2014 M 8.1 earthquake in north Chile (Brodsky and Lay, 2014) indicate foreshock sequences that propagated along the fault, consistent with being triggered by slow slip; in the Tohoku-Oki case, the slow-slip explanation is further supported by the presence of repeating earthquakes among the foreshocks. The nature of fault patches that could produce foreshocks can be glimpsed from unique laboratory experiments of earthquake nucleation in a meter-scale slab of granite (McLaskey and Kilgore, 2013; McLaskey et al., 2014), which produce quasi-static accelerating slip (nucleation process) accompanied by smaller seismically detectable events, inferred to occur on persistent fault patches with properties that are different from the rest of the sample.

Since large, destructive earthquakes, and even intermediate-sized events, occur rarely on any given fault segment, it is important to numerically model the microseismicity and its relation to the other processes on the fault such as aseismic slip to extract as much information as we can about the physics of the earthquake source. Furthermore, some patterns of microseismicity may serve as precursors, provided that they can be identified as such a priori.

Consequences of fluid injection: patterns of induced slow slip and seismicity

It has long been recognized that injection of fluids into the subsurface can cause earthquakes (e.g., Evans, 1966; Rothé, 1970). The effective stress principle, initially developed for high porosity soil by Terzaghi, combined with the Mohr-Coulomb failure criterion seems to also explain, to the first order, the destabilization of faults formed in low porosity, low permeability rock formations, as demonstrated by in-situ experiments and laboratory studies (Raleigh et al., 1976; Jaeger, 1979; Zoback and Harjes, 1997). In keeping with these principles, seismicity induced by injection in deep wells is often observed to expand radially with the square root of time as expected from pore pressure diffusion (e.g., Shapiro et al., 1999). The expansion is generally not isotropic, reflecting the enhanced permeability along activated fractures (Hummel and Shapiro, 2013), an effect that can be measured in the laboratory and in-situ (Cappa et al.,
It is clear, however, that seismicity does not only occur in zones of increased pore pressure and can be triggered by poroelastic stress change, sometimes at distances of several tens of kilometers from the point of injection (Cochran et al., 2018; Goebel et al., 2017; Segall and Lu, 2015).

The mechanisms by which faults react to a fluid injection by seismic events remain poorly understood. In fact, there is accumulating evidence that fault slip triggered by a fluid injection can be aseismic. This has been observed in the laboratory and and in-situ experiments (Guglielmi et al., 2015) (Figure III-IN3) and in studies of geothermal injections (Bourouis and Bernard, 2007; Wei et al., 2015). Figure I-4(c) illustrates the example of the Brawley geothermal plant, which started operating in 2010. An intense seismic crisis occurred in 2012, including a Mw 5.4 earthquake, more than two years after the onset of energy production. Analysis of inSAR and geodetic data revealed that this earthquake was preceded and triggered by aseismic motion of a normal fault (Wei et al., 2015). The inversion revealed clearly that the zone of aseismic slip correlates with the zone of fluid injection (Figure I-4(c)). The Mw 5.4 earthquake may have been triggered indirectly by the stress redistribution imparted by the aseismic fault slip.

Successes of earthquake source modeling in matching these observations

These rich observations provide a wealth of constraints on modeling approaches for simulating sequences of earthquakes and aseismic slip (SEAS, MMC2). SEAS modeling aims to capture the observed interplay of interseismic periods and the associated aseismic fault slip—that ultimately lead to earthquake nucleation—and dynamic earthquake events themselves, in an effort to understand which physical factors control the full range of observables such as aseismic deformation, nucleation locations of earthquakes, and the recurrence times and magnitudes of major events. SEAS models simulate long periods of earthquake activity, incorporating the pre-, inter-, and postseismic slip and the resulting stress redistribution, spontaneous earthquake nucleation, and physical processes relevant to long-term slip such as interseismic healing of the fault zone and, increasingly, fluid effects, viscoelasticity, damage zones, and fault roughness. Accounting for such complexity introduces computational challenges, but is widely recognized as crucial for understanding the real Earth (MMC1) and constraining seismic hazard.

The SEAS modeling has been quite successful in reproducing the main features of the seismic-aseismic behavior of the fault zones using the general framework of rate-and-state constitutive laws formulated based on rock mechanics experiments at slow slip rates and its extensions (MMC1.1; Figures II-MMC1.1 and V-MMC1.1), as described in section MMC2.2. Faults whose shear resistance increases with increasing slip rate are called velocity-strengthening; they tend to arrest fast earthquake rupture and promote aseismic slip, unless destabilized by additional processes such as enhanced dynamic weakening. Faults whose strength decreases with increasing slip rate are called velocity-weakening; such fault regions can slip slowly over areas smaller than the nucleation size (V-MMC1.1) and accelerate into seismic ruptures over larger areas, remaining mostly locked otherwise. Postseismic slip naturally arises in this framework, as the increase in the ratio of shear to normal stresses imposed by the dynamic rupture results in larger slip rates on velocity-strengthening areas.

In particular, the observed transient slow slip events can be reproduced in SEAS models on velocity-weakening faults if they are sufficiently stable, i.e. their properties correspond to a large nucleation size (Figure V-MMC1.1-2), comparable to the extent of the SSE along dip (Liu and Rice, 2005, 2007; Liu and Rubin, 2010, Segall et al., 2010; Li and Liu, 2016,2017). The large nucleation size can be due to their slow weakening rate caused by either low effective normal stress (and, correspondingly, high pore pressure, permanent or transient) or nearly velocity-neutral friction properties.
Sliding can be further stabilized by the emergence of strengthening processes during increased slip rate (Okubo and Dieterich, 1986; Weeks, 1993; Shibazaki and Iio, 2003; Shibazaki and Shimamoto, 2007, Beeler, 2009; Matsuzawa et al., 2010, Shibazaki et al., 2010, Bar-Sinaï et al., 2014). Depending on the composition, structure, pressure, and temperature of a fault zone, different rate-strengthening micro-mechanical processes (e.g., dilatant hardening, cataclastic flow, diffusion creep, grain-boundary sliding, and dislocation creep) can contribute to the strengthening behavior (Bürgmann, 2018). A range of empirically and theoretically motivated formulations of fault zone rheology can be employed to model this behavior. For example, experimentally-motivated extensions (Segall and Rice, 1995) of the standard rate-and-state framework to include the stabilizing effect of dilatant hardening (MMC1), in which shear dilatancy (Marone et al., 1990; Lockner and Byerlee, 1994; Samuelson et al., 2009) results in increasing pore space and decreasing pore fluid pressure, enable SEAS models to reproduce SSEs more robustly (Liu and Rubin, 2010; Segall et al., 2010) than with the standard rate-and-state friction alone.

Modeling shows that rate-strengthening faults can also host transient accelerations of slow slip into SSEs if they are destabilized, for example, by poroelastic effects (Heimisson and Dunham, 2018). In fact, the occurrence of tremor on the velocity-weakening patches embedded within the otherwise velocity-strengthening fault can promote the transient accelerations there due to the collective postseismic slip of the many events that constitute tremor (Luo and Ampuero, 2017). Shear-layer compaction (which is the process opposite to dilation) has been shown to destabilize the inherently velocity-strengthening materials in laboratory experiments (Faulkner et al., 2018).

The observed overlap of seismic and aseismic slip can be explained by occasional stabilization of the seismogenic fault areas as already discussed, e.g., by transient pore fluid pressure increases. In addition, creeping regions that are velocity-strengthening at low slip rates can be ruptured dynamically and produce seismic slip due to enhanced dynamic weakening (MMC1.1), specifically thermal pressurization of pore fluids (Noda and Lapusta, 2013) (Figure II-MMC3.1-2). If the hydromechanical properties of the creeping segments are suitable and dynamic rupture is vigorous enough, enhanced dynamic weakening - for example, due to shear heating - can be activated before the rupture arrests, overwhelming the (relatively mild) velocity strengthening and resulting in overall weakening behavior that could sustain rupture propagation over substantial stretches of otherwise stable fault regions.

SEAS models based on rate-and-state friction have been quite successful in reproducing the properties of small repeating earthquakes, qualitatively as well as quantitatively (Future modeling progress)

Overall, the observations and modeling of spatio-temporal patterns of seismic/aseismic slip and distributed deformation paint a rich picture that can be explained by the interaction of the standard rate-and-state friction and/or distributed shear deformation with fluid effects and heterogeneity of the fault zone. Similar conclusions have been reached in modeling the effects of fluid injection (e.g., Guglielmi et al., 2015; Galis et al., 2017). However, many aspects of the employed models are simplified (MMC2) or poorly constrained, resulting in multiple explanations proposed for the same phenomena. One aspect that remains under-represented in the current modeling efforts is the fully coupled poroelastic effects and the evolution of hydro-mechanical properties with fault deformation, damage, and cyclic loading due to dynamic waves (MMC1.4).

Furthermore, these modeling efforts do not yet include distributed inelastic deformation (MMC1.2, MMC1.3) which can contribute to or even dominate some of the inferred interseismic phenomena typically attributed to fault slip (Deng et al., 1999; Hetland and Hager 2005; Fay and Humphreys 2006; Bürgmann
and Dresen, 2008; Takeuchi and Fialko 2012; van Dinther et al., 2013b; Lambert and Barbot, 2016; Lavier et al., 2013; Lindsey et al., 2014a,b; Hearn and Thatcher, 2015; Milliner et al., 2015; Erickson et al., 2017; Sobolev and Muldashev, 2017; Allison and Dunham, 2018; Tong and Lavier, 2018), although there are promising beginnings that combine SEAS modeling and inelastic off-fault effects (Kaneko and Fialko, 2011; Lambert and Barbot, 2016; Erickson et al., 2017; Allison and Dunham, 2018). For example, for the shallow distributed deformation, modeling has demonstrated that viscosity, hardening, and cohesion all influence the extent and magnitude of the off-fault inelastic strain at shallow depths, and all scenarios give rise to a shallow slip deficit, corresponding to ∼10% of the tectonic deformation budget (Kaneko and Fialko, 2011; Erickson et al., 2017). These numerical findings compare well to the geodetic-based observations (Meade et al., 2013) which estimate that 6% ± 9% of deformation occurs off of several major strike-slip faults. These findings suggest that cumulative inelastic deformation over the course of many events can account for a significant amount of tectonic offset in the shallow subsurface. The visco-elastic-plastic deformation at depth can accommodate part or all of the relative shear often modelled as frictional (e.g., Allison and Dunham, 2018), depending on the friction particulars and inelastic properties, both of which could be strongly influenced by temperature and fluid effects.

Future progress will come from more targeted laboratory studies and systematic modeling efforts (MMC2.2) that incorporate a range of the relevant mechanisms (MMC1) and aim to match a wide range of relevant observations. One aspect of most explanations for the observed slow slip transients present nearly in all SEAS models is significantly elevated pore fluid pressure, close to the overburden stresses expected to be 300-400 MPa at the relevant depths, with the effective normal stresses in 1-10 MPa range. Laboratory experiments under such conditions are rare. A related issue is that our current understanding of constitutive behavior comes from the patchwork of laboratory tests on different fault materials that have been tested at some conditions of interest. Given the intricate interplay of a number of mechanisms in the resulting constitutive response (MMC1), including the importance of mineral composition and chemical reactions, sustained efforts is needed to test the materials likely to be present in the faults of interest over the representative range of conditions identified by the current models, including the highly elevated pore fluid pressures. Efforts to constrain fault friction in presence of fluids from laboratory (e.g., Scuderi and Collettini, 2016; Scuderi et al., 2017; Faulkner et al., 2018), in-situ field studies (e.g., Guglielmi et al., 2015; Kneafsey et al., 2018), and induced seismicity sites (section III-IN3), in conjunction with modeling, are therefore much needed to investigate the physical and chemical processes within the fault zone.

Using modeling to verify the observational inferences of slow slip and tremor phenomena

The amount of slow slip is often small in any observational time period and thus slow slip is difficult to detect, but the changes in local stress that result from slow slip can be substantial and may trigger earthquakes. Data processing and inversion methods involve multiple steps and make modeling assumptions. Modeling results can help verify the employed approaches. One can create synthetic data based on the modeling results, add errors and proxies for other signals, check how well the existing methods would uncover the on-fault slips, and potentially suggest method improvements. This procedure is especially relevant for slip phenomena that may appear in models but have not yet been captured the observations. One possibility is that such phenomena do not exist on natural faults. But another possibility is that the employed methods cannot detect them. Using modeling outcomes in terms of seismic/aseismic slip as a verification tool for the observational inferences will help to improve our imaging of these challenging, often small-scale and transient, events.

Key future challenges:
● Determine the dominating mechanisms among several potential explanations for the various observed patterns of seismic/aseismic slip and distributed deformation, and the associated range of fault properties, by using multi-physics SEAS models (MMC2.2) with the improved constitutive relations for fault slip and inelastic deformation (MMC1);

● Constrain models of faulting in the presence of fluids by instrumenting and modeling the ongoing industrial activities that involve fluid injections, which are essentially meso-scale field experiments (see also III.IN3);

● Explore the implications of the determined fault properties on the nucleation, propagation, and arrest of large earthquake ruptures; in particular, (i) determine the conditions under which large earthquakes have identifiable precursors and/or can be triggered by slow slip transients, (ii) constrain the depth extent of large earthquake ruptures, and (iii) establish if/when earthquake ruptures can propagate through creeping regions.

MM.C3.3 Properties of earthquake rupture events, their rupture speeds, energy budget, and radiation features

Over the past decade, detailed imaging of earthquake rupture has been enabled by observations from seismic, geodetic, and tsunami monitoring networks as well as advances in data processing techniques. The combined analyses of large earthquakes now routinely includes inversions for finite-fault slip history, sources of high-frequency radiations, and the associated parameters such as rupture speed, stress drop, and various scalings (e.g., Ishii et al., 2005; Simons et al., 2011; Yue et al., 2012; Avouac et al., 2015; Duputel et al., 2015; Melgar et al., 2016; Ye et al., 2016ab; Chouet and Vallée, 2018). Modern geodetic observational systems - including GNSS (Global Navigation Satellite Systems), InSAR (Interferometric Synthetic Aperture Radar), and seafloor-based instruments - not only help to constrain the final slip but have greatly increased the spatial resolution in measuring ground deformation during earthquakes (e.g., Simons et al., 2002; Fialko et al., 2005; Konca et al., 2008; Perfettini et al., 2010; Hamling et al., 2017). Integrated seismic and geodetic measurements also enable broadband constraints on ground displacements (Bock et al., 2011).

Rupture speeds and supershear rupture propagation

The general picture illuminated by these efforts reveal that slip generally starts from a small area of the fault compared with the final earthquake rupture dimensions, at least for larger ruptures that allow for finite-fault inversions; the size of the slipping patch at the initiation of seismic slip is not resolvable by seismic inversions. The slipping region (also called rupture) propagates along the fault, with the speed of its front (known as rupture speed) being a significant fraction, 0.7-0.8, of the shear wave speed of the rock bulk, independently of the event size (although some events can break out into a speed faster than the shear wave speed, producing a supershear rupture, see MMC2.1 and below). The rupture speeds are consistent with theories of singular dynamic fracture, in which shear cracks with constant fracture energy (equal to the flow of energy into the singular crack tip) rapidly accelerate to the limiting speeds which are the shear and Rayleigh wave speeds, respectively, for the anti-plane (Mode III) and in-plane (Mode II) directions of slip (Freund, 1990). The fact that the observed rupture speeds saturate before reaching the limit may be related to the geometry of the rupture in which one dimension gets saturated by the seismogenic depth. Alternatively, it may signify that the equivalent to the fracture energy for the earthquake source, the so-called “breakdown” energy (Ida, 1972; Palmer and Rice, 1973; Rice, 1980; Freund, 1990; Bizzari and Cocco, 2006) is increasing as the rupture propagates. The latter is supported by seismological inferences (Abercrombie and Rice, 2005; Rice, 2006; Viesca and Garagash, 2015), as discussed further below.
Rupture propagation with intersonic speeds, between the shear and pressure wave speeds of the surrounding rock, has been inferred for many of the well-recorded large strike-slip events, including the 1979 Imperial Valley earthquake (Archuleta, 1984), 1992 Landers earthquake (Olsen et al., 1997), 1999 Izmit earthquake (Bouchon et al., 2000), 2001 Kunlun earthquake (Bouchon and Vallee, 2003), 2002 Denali earthquake (Ellsworth et al., 2004), 1906 San Francisco earthquake (Song et al., 2008), and others (Wang et al., 2016). The possibility of such transition had been recognized due to theoretical and dynamic rupture modeling (Burridge, 1973; Andrews, 1976) and then demonstrated in the lab (e.g. Xia et al., 2004; Rosakis et al., 2007) (Figure II-MMC2.1-1). The supershear transition can only occur in the in-plane (or Mode II) direction of shear, which coincides with the along-strike direction for strike-slip faults and along-dip directions for thrust faults such as megathrusts, explaining why it is commonly observed for the strike-slip but not subduction-zone earthquakes. Typically, supershear rupture has been observed on very straight fault segments at distances of more than tens of kilometers from the hypocenter (e.g., Xia et al., 2004; Bruhat et al., 2016 and references therein). The practical significance of supershear propagation is the much stronger shaking farther away from the fault due to the strong Mach cone that develops at the fronts of the piled-up shear waves that are left behind by the rupture. Accentuated damage from the Mach cones has been observed for events like the 2018 Palu, Sulawesi earthquake (Ulrich et al., 2019, where massive liquefaction occurred in the river valley along the southern extension of the rupture.

Hence understanding which physical factors lead to supershear propagation on strike-slip faults, such as the San Andreas Fault (SAF) in California, is quite important for seismic hazard. A number of studies examined the interaction of supershear rupture with fault heterogeneity and roughness, which can both promote the supershear transition if it is of specific types or destroy the coherence of the wave field and prevent the supershear rupture from occurring even under the otherwise favorable conditions (Dunham and Archuleta, 2004; Liu and Lapusta, 2008; Bizzarri et al., 2010; Bruhat et al., 2016). Modeling the inferred propensity of large strike-slip events to become supershear can then help constrain the type and level of heterogeneity on mature strike-slip faults.

**Earthquake breakdown energy**

Remote seismic observations have been used to indirectly quantify the energy dissipated at the rupture front, a quantity known as the breakdown work or fracture energy (Ida, 1972; Palmer and Rice, 1973; Rice, 1980; Bizzarri and Cocco, 2006). Compilations of such inferences for many events show that breakdown energy increases with earthquake magnitude or average slip (Abercrombie and Rice, 2005; Rice, 2006; Viesca and Garagash, 2015) (Figure II-MMC3.3-2). In a related finding, attempts to constrain the critical slip-weakening slip from seismic observations find that the inferred critical slip increases with the total slip of the event, being approximately half of it (Kaneko et al., 2017), which would correspond to increasing breakdown work on slip-weakening faults. Dissipation can occur both on the fault - during shear strength reduction from changes in friction coefficient as well as other weakening processes like thermal pressurization - or during off-fault inelastic deformation. The observed scaling of breakdown energy could arise from either on-fault or off-fault processes or both. If dissipation is principally on the fault, then the fracture energy scaling indicates that weakening must continuously occur with increasing slip (Abercrombie and Rice, 2005). Thermal pressurization is one process that can produce this behavior, and recent studies utilizing either an assumed constant slip velocity (Rice, 2006) or a steady state slip pulse solution (Viesca and Garagash, 2015) demonstrate consistency with observations (Figure V-MMC3.3-2). Alternatively, if dissipation dominantly occurs off of the fault during inelastic straining of the damage zone rocks, then the size and/or intensity of the inelastic region must grow with earthquake size. Dynamic rupture simulations with continuum plasticity show exactly this behavior (Andrews, 2006; Templeton and Rice, 2008; Ma and Andrews, 2010; Dunham et al., 2011a; Gabriel et al., 2013).
However, these studies are all computationally limited to perhaps two or at most three orders of magnitude in scale separation between the rupture size at initiation and when the rupture reaches the edges of the computational domain, limiting the range of event sizes that one can consider. A possible solution would be the use of adaptive mesh refinement to simulate rupture growth over the many orders of magnitude spanned by the observations constraints.

**Radiated energy: depth-dependent frequency and regional variations**

Systematic analysis of many large earthquake ruptures for kinematic finite-fault slip distributions using seismic, geodetic, and tsunami data (Ye et al., 2016a,b) has revealed relative strength of radiated energy (Ye et al., 2018), which appears to be influenced by regionally varying fault zone characteristics for large events (Figure II-MMC3.3-1). At the same time, systematic analysis of the moment rate functions for the same dataset reveals robust scaling attributes and linear growth for event stacks in varying magnitude bins (Meier et al., 2017) (Figure II-MMC3.3-2, top). Understanding the nature of these observed systematic behaviors requires investigation with dynamic rupture simulations and SEAS models. Examination of large populations of source parameter estimates indicate that moment-scaled rupture durations vary systematically with depth along plate boundary megathrusts, and there tends to be an increase in the relative strength of high frequency radiation as depth along the megathrust increases (e.g., Simons et al., 2011; Meng et al., 2011; Kiser and Ishii, 2012; Lay et al., 2012; Yao et al., 2013; Ye et al., 2016a). Given the many properties of the megathrust fault zone fluids, sediments, roughness, mineralogy, that are expected to vary with depth as pressure and temperature increase, modeling is needed to develop a comprehensive understanding of the observations. Dynamic rupture simulations have considered the role of fault heterogeneity in generating the high-frequency generation at depths (Huang et al., 2012; Galvez et al., 2014) but most other potential physical mechanisms remain unexplored.

**Shallow rupture propagation on megathrusts and continental faults**

Much recent observational and modeling focus has been on the potential for rupture propagation through the shallow portions of the fault. In the subduction zone environment, magnitude 8 and 9 class megathrust ruptures can span the entire seismogenic zone of subducting plate boundary faults, all the way to the trench. These megathrust ruptures, like the 2011 Tohoku-Oki earthquake, are responsible for generating damaging tsunamis as well as causing destruction through seismic radiation (Ni et al., 2005; Stein and Okal, 2005; Lay et al., 2005; Fujitake et al., 2011; Mori et al., 2011; Simons et al., 2011; Jiang and Simons, 2016). Dynamic rupture simulations have highlighted the importance of wave reflections from the seafloor and sea surface and associated dynamic stress changes in facilitating shallow rupture propagation (Kozdon and Dunham, 2013); these effects have also been observed in laboratory experiments (Gabuchian et al., 2013, 2017). Rupture though this region can be further enhanced due to elastic bimaterial effects (Ma and Beroza, 2008) and extreme elastic compliance of near-trench sediments (Lotto et al., 2017, 2018). Furthermore, it is also possible that dynamic weakening processes like thermal pressurization might activate during sufficiently large ruptures like the 2011 Tohoku-Oki event (Mitsui et al., 2012; Noda and Lapusta, 2013; Cubas et al., 2015). Rupture interaction with geometrical features like seamounts (Yang et al., 2012, 2013) and splay faults (Wendt et al., 2009; Xu et al., 2015; Lotto et al., 2018) has also been studied. Shallow sediments might deform inelastically during megathrust ruptures, a process that can substantially alter the seafloor uplift that generates tsunamis (Ma, 2012; Ma and Hirakawa, 2013). Three-dimensional simulations of megathrust events are now becoming possible, due to advances in parallel computing (Duan, 2012; Galvez et al., 2014; Breuer et al., 2014; Heinecke et al., 2014; Uphoff et al., 2017; Gabriel et al., 2018; Ulrich et al., 2018). An area of active research for offshore earthquakes is the coupling of dynamic rupture simulations to tsunami simulations for improved
insights into the tsunami generation process (Wendt et al., 2009; Ryan et al., 2015; Lotto and Dunham, 2015; Lotto et al., 2017; Gabriel et al., 2018; Ulrich et al., 2018).

**Dynamic source inversions constrained by physics**

Seismologists have estimated kinematic earthquake rupture parameters (slip, rupture initiation times, etc) using inverse methods for many decades (e.g., Wald et al., 1996; Custodio et al., 2005). After the radiation of seismic waves was shown to be sensitive to stress drop and friction parameters (e.g., Olsen et al., 1997; Peyrat et al., 2001; Ma et al., 2008), inversion for dynamic rupture parameters of earthquakes has been explored over the past two decades. For example, a systematic nonlinear inversion technique (the neighborhood algorithm (Sambridge, 1999) was used to estimate the stress drop distribution using rectangular subfaults for the 2000 Tottori, Japan, earthquake (Peyrat and Olsen, 2004). In order to reduce the parameter space to a more computationally manageable set, the neighborhood algorithm was used to invert for dynamic rupture parameters within a few elliptically-shaped areas on the fault (Di Carli et al., 2010; Ruiz and Madariaga, 2011, 2013; and Ruiz et al., 2017). However, resolution of the estimated dynamic parameters is limited by typically small amounts of near-field strong-motion data available and the presence of correlation between rupture parameters (Corish et al., 2007). Uncertainties in the crustal structure surrounding the fault contribute additional sources of error in the estimation. Nonetheless, a more complete exploration of the parameter space including uncertainties (e.g., Gallovic et al., 2018) with the help of today’s powerful supercomputers and more abundant strong motion data is bound to provide further constraints on the dynamic rupture parameters in the near future.

**Key future goals**

- Use SEAS models constrained by other observations as well as dynamic rupture models with the initial conditions informed by the SEAS models to investigate the origin of the observed properties of seismic radiation in large events, including enhanced high frequency radiation from depth and systematic patterns over the subduction zones;
- Use the modeling to understand the potential of large earthquakes to break to the surface, especially in subduction zones, and to become supershear;
- Develop procedures for checking and improving finite-fault inversions based on earthquake source simulations, including dynamic rupture inversions.

**3.4 Magnitude-invariant stress drops over the entire range of earthquake sizes**

Earthquakes lead to the overall reduction of shear stress on the ruptured fault area, and stress drop - which represent the average difference between the initial and final shear stress for the earthquake source - is an important parameter which has long been studied in seismology. Inferences of static stress drops exhibit an invariance over seventeen orders of earthquake magnitude (Ide and Beroza, 2001; Abercrombie and Rice, 2005; Allmann and Shearer, 2009; Goodfellow and Young, 2014; McLaskey and Lockner, 2015; Cocco et al, 2016; Ye et al, 2016a) (Figure II-MMC3.4). Static earthquake stress drops range between 0.1 and 100 MPa, with a median of about 3 MPa. There is also no systematic dependence on depth for stress drops taken as a whole, although recent observations have found some increase of stress drops with depth in certain regions (Oth, 2013; Uchide et al., 2014; Goebel et al., 2015; Trugman and Shearer, 2017).

**Epistemic uncertainties in estimates of stress drop**
The large scatter in stress drop values is attributed to both natural variability and epistemic (i.e., model-based) uncertainties. The epistemic uncertainties may be considerable. For most earthquakes, especially small ones, the finite fault inversions discussed in MMC 3.3 are not feasible, and other approaches have to be used to decipher the earthquake source properties. The moment magnitude can be reliably estimated from long-period seismic waves; various methods are used to obtain the source duration from the seismograms. Once the moment and duration of the earthquake source are known, seismologists typically use the circular crack model of an axisymmetrically expanding rupture with a constant rupture speed and uniform stress drop (Eshelby, 1957; Madariaga, 1976) to infer the stress drop. If the source model correctly represents the earthquake, uncertainties in these relations may bring a factor of 10 if the rupture velocity is not known (Kaneko and Shearer, 2014), or if there is source directivity (Kaneko and Shearer, 2015; Ross and Ben Zion, 2016), or simply if the focal sphere is not well sampled (Kaneko and Shearer, 2015). In such cases, as well as for elliptical sources, use of second-moment methods allows to capture the directivity and eccentricity of the source and improve the stress drop estimates (McGuire and Kaneko, 2018), which exhibits less variability in stress drop than corner frequency estimates. If the earthquake source shape is very different from an elliptical patch and/or exhibits strong heterogeneity, the estimates of stress drop may be biased by more than an order of magnitude (Noda et al, 2013; Brown et al, 2015; Adams et al, 2016; Lin and Lapusta, 2018). Nonetheless, estimates of stress drop from finite-fault inversions, for larger events, are close to those inferred from moment-duration relations and provide a median stress drop of 5 MPa (Adams et al, 2016; Cocco et al, 2016). These estimates are critical for determining the seismic energy budget (Kanamori and Rivera, 2006).

The recovery of a broadband seismic pulse, either in spectral or time domain, that captures the source duration and other properties, is the first step in the moment-duration stress drop estimation approaches. Central to the earthquake problem, far-field wave propagation effects, and in particular attenuation and scattering, ought to be removed from the observed seismogram. Both theoretical (Boatwright and Choy 1986) and empirical (Mueller 1985; Shearer et al, 2006; Abercrombie 2015; Baltay et al, 2014; Denolle and Shearer 2016; Ross and Ben-Zion 2016; Trugman and Shearer 2017) approaches suffer from inaccuracies in velocity and attenuation models or from limited frequency bandwidth. Synergistic activities are necessary to establish the epistemic uncertainties of the multiple methodologies to reconstruct the source shape and infer stress drop (Abercrombie and Shearer, 2018).

Improving earthquake source models used for stress drop estimation

To reduce the epistemic uncertainties due to oversimplified earthquake source models, we need to improve the inference of stress drop given the seismic observations of moment, source duration (or corner frequency), source time function or spectral shapes, and of finite fault models. Currently, most observational studies use the uniform, circular, constant-stress-drop and constant-rupture-speed source model (Eshelby, 1957; Brune, 1970; and Madariaga, 1976) to predict the stress drop from the seismic moment and P- and S-wave pulse duration. However, the assumption of a circular rupture area is unlikely to hold faults with heterogeneity (Lin and Lapusta, 2018), including large multi-segment earthquakes (e.g. the M7.8 2016 Kaikoura Earthquake, Ando and Kaneko, 2018). Furthermore, the assumption of a smooth source-pulse shape fails to explain certain observations (Uchide and Imanishi, 2016). We need more realistic source models that incorporate finite fault effects such as source directivity, e.g. a double-corner frequency source model (Denolle and Shearer 2016); fault geometry (Noda et al, 2013; Kaneko and Shearer 2015; Lin and Lapusta, 2018); and rupture style, e.g. crack- vs pulse-like rupture (Wang and Day 2017). The use of the second-moment methods is an example of such an improvement (McGuire and Kaneko, 2018). Improving the inference of stress drop with finite fault models by solving for slip distributions that minimize both data misfit and stress drop is another promising direction (Adams et al., 2016).
Earthquake source simulations can explore the relation of seismologically estimated stress drop to fault properties

The invariance of stress drop is sometimes interpreted as due to a geometrical invariance, where by the geometry of the fault and of the slip follows a self-similar scaling (Prieto et al., 2004). However, such similarity should break at the scale of earthquakes that saturate the seismogenic depth (Romanowicz and Ruff 2002, Leonard 2010, Denolle et al., 2016, Weng and Yang, 2017) but the magnitude-invariance of stress drops persists even for the largest magnitudes. Self-affine fault roughness predicts a decrease of stress drop with earthquake size (Candela et al., 2011), which is not observed.

Note that the scale invariance of stress drop holds when combining inference from multiple data sets; however, individual studies tend to exhibit a scaling with an increasing trend (Figure II-MMC3.3, Ide and Beroza, 2001; Oth, 2013; Trugman and Shearer 2017; Wu et al., 2018). This can be explained by data selection and finiteness of the frequency bandwidth (Ide and Beroza, 2001; Abercrombie 2015). But it is also possible that aftershocks have lower stress drops than main shocks (Shaw et al., 2015; Bindi et al., 2018), possibly due to the co-seismic damage of near-fault bulk (Brenguier et al, 2008; Thomas et al, 2017; Bindi et al, 2018).

It is tempting to interpret the seismologically estimated stress drops as the actual decrease from the initial to final shear stress across the fault, as the name implies, yet the relation, even theoretically, is more complicated. As the earthquake rupture propagates through the potentially heterogeneous stress field with evolving strength that depends on many local fault characteristics (MMC1), and finally arrests, the local stress change is bound to be heterogeneous across the rupture area, with a potentially non-negligible area of negative stress drops (e.g. stress increase) around the rupture perimeter (Noda et al., 2013; Schaal and Lapusta, 2019). A single number representing this potentially complex distribution needs to average over it. The moment-duration-based estimates discussed above are actually weighted averages of the distribution (Madariaga, 1979; Noda et al., 2013). The stress drops are then often used in earthquake energy considerations (Kanamori and Brödsky, 2004; Kanamori and Rivera, 2006), but that use implies a different weighted average, and the two averages can be quite different for significant fault heterogeneity (Noda et al., 2013).

In short, the (average) stress drop is process-dependent, i.e., it depends on the rupture dynamics. So its relation to fault properties can only be reliably explored through earthquake source simulations.

Furthermore, multiple processes on and off the fault that contribute to the stress evolution and hence earthquake stress drop are inherently scale-dependent. Fault roughness is observed to be self-affine, with no intrinsic length scale (Candela et al, 2011). Because the interpretation of the relation between stress change and fault topography induces a scaling of stress drop with magnitude, fault roughness may not dominate at all scales (Cocco et al, 2016). Dynamic weakening mechanisms (MMC1.1) are scale-dependent, and recent interpretation of the scaling of seismic observations suggest that they dominate in large earthquakes (Rice, 2006; Viesca and Garagash, 2015). If these interpretations of observations hold, the effective constitutive parameters of faults are indeed scale-dependent. Furthermore, the dominance of dynamic weakening effects at large scale may imply either lower pre-stress levels before larger earthquakes (Lambert et al., 2018), or strong healing mechanisms, or both to maintain a scale-invariant stress drop. Strong healing mechanisms are invoked in pulse-like ruptures that likely prevail in large earthquakes, with rise times that are much shorter than the source duration (Heaton 1990; Melgar and Hayes, 2017), and healing mechanisms may also be scale-dependent. Off-fault damage and dissipation (MMC1.2) is expected to grow for larger earthquakes (Andrews, 2006; Templeton and Rice, 2008; Ma
and Andrews, 2010; Dunham et al., 2011a; Gabriel et al., 2013) and to reduce radiation in seismic waves, a potential counter effect to dynamic weakening mechanisms (Dunham et al, 2011ab; Ma and Hirakawa 2013; Roten et al., 2018). Building on the current accomplishments, the multi-physics modeling of the earthquake source (MMC2) will soon be able to explore the interplay between these scale-dependent mechanisms that, together with rupture dynamics, yield scale-invariant stress drops.

**Relation of the scatter in stress drop to fault heterogeneity**

While some of the stress drop scatter is likely due to the epistemic uncertainty, the variations in stress drops may also hold valuable information about the variability in fault properties that combine to produce stress drops, as discussed above. How much can the variability in stress drops tell us about fault heterogeneity (Allman and Shearer, 2007)? The variability of high-frequency ground motion is lower than that of inferred stress drops (Cotton et al, 2013). What is the spectrum of plausible heterogeneity on the fault that can explain the variability in ground motions? Does it reproduce as well the range in stress drops?

**Significant modeling and numerical challenge**

Modeling earthquake sequences over a large range of magnitudes with a broad set of physical mechanisms is numerically quite challenging. Several earthquake source simulations have produced up to four orders of magnitude that exhibit observed statistical properties of earthquakes (Zielke and Arrowsmith 2008). The next step, fueled by the developments highlighted in MMC1 and MMC2, would be to incorporate more realistic fault geometry, near-source structure and its rheology, frictional behaviors at seismic slip rates, and full inertial effects while creating a catalog of simulated earthquakes that span a range of magnitudes and explore the scaling and variability in simulated stress drops.

**Key future challenges:**

- Improve observational constraints on stress drops, in part, by improving the earthquake source models used; use earthquake source simulations to determine how much of the scatter could be physical and what the observed scatter tells us about fault heterogeneity;
- Systematically investigate the dependence of stress drops on depth with improved data and improved, simulation-informed methods, to clarify whether there is indeed no systematic depth dependence in most cases, or whether the overall data hides important local dependencies as suggested by some recent studies;
- Determine the combinations of fault physical mechanisms and properties - many of which are found to be scale- and depth-dependent - that result in the invariance of stress drop over the entire range of earthquake magnitudes and potentially with depth; a key challenge is to formulate multi-physics SEAS models that can reproduce a wide range of earthquake sizes in a numerically tractable way.

**3.5 Gutenberg-Richter law, Omori’s law for aftershocks, and other statistics on regional scales**

As earthquake source simulations advance to incorporate complex fault geometries, and develop the ability to simulate a range of seismic events, they will need to consider constraints based on key statistical characteristics of seismic sequences inferred from observations, including distribution of earthquake sizes and their temporal patterns.
Frequency-magnitude distribution of seismicity

Perhaps the best known empirical law in seismology is the Gutenberg-Richter law (Gutenberg and Richter 1944) (Figure V-MMC3.5), which describes the distribution of earthquake magnitudes (or frequency-magnitude distribution). It states that the number of earthquakes $N$ with magnitude at least equal to $M$ is given by $\log N(M) = a - bM$, where parameter $a$ is related to the overall number of earthquakes in the region considered and parameter $b$ controls the relative frequency of small and large earthquakes. This relationship has been confirmed at global and regional scales, even though some variations exist. Typical values of parameter $b$ are close to 1, but values in the range $0.4 < b < 2.0$ have been reported (Wiemer and Wyss 2002), with some systematic variations across tectonic settings (Schorlemmer et al., 2005) which have been attributed to the dependence on differential stress. Given the relationship between earthquake moment magnitude and seismic moment (Hanks and Kanamori 1979), this is a power-law in terms of the seismic moment. Consequently, the Gutenberg-Richter relation has been interpreted as a manifestation of a scale-invariant underlying process (Rundle 1989), and linked to the fractal nature of faults surfaces (Aki, 1981). The power-law distribution appears to hold down to the smallest scales accessible to observations, including microearthquakes down to Mw -1.3 (Boettcher et al., 2009) and laboratory experiments of acoustic emission at the millimeter scale (Uhl et al., 2015).

While seismicity on a global and regional scale is consistently well fit by the Gutenberg-Richter relation, some exceptions emerge when considering more limited seismogenic sources. For example, some sequences of small (Mw < 3) periodic earthquakes present very little variation in magnitude (Nadeau and Johnson 1998; Chen et al., 2007). On a larger scale, there is evidence that individual fault segments experience a higher fraction of “system-size” events (i.e. earthquakes that rupture the entire segment) than would be predicted by the Gutenberg-Richter scaling (Schwartz and Coppersmith, 1984; Wesnousky, 1994; Parsons, 2018), even though these claims have been questioned (Page, Alderson, and Doyle, 2011; Page and Felzer, 2015). Such behavior is often called “characteristic.”

Temporal patterns of seismicity

Another important property of the collective behavior of earthquakes, and potentially diagnostic of the underlying physics, is their temporal distribution. Temporal patterns can be broadly divided into three categories: (i) Periodic or quasi-periodic, characterized by regular intervals between events; (ii) Clustered, characterized by periods of more intense activity; and (iii) Poissonian, characterized by random occurrence. Each type of behavior is described in more detail below.

Quasi-periodicity: Earthquakes can be described as the rapid release of elastic energy accumulated on the timescale of years (Reid, 1910). Assuming that the stressing rate between earthquakes is constant; each earthquake releases the same amount of stress; and the stress threshold for instability is constant in time, this simple model leads to simple periodic cycles. However, quasi-periodic sequences are extremely rare in nature. While some small quasi-periodic events have been observed on creeping faults in various locations worldwide (Nadeau and Johnson 1998; Chen et al., 2007), larger quakes are typically characterized by a lower degree of periodicity. In spite of the popularity of early quasi-periodic models of earthquake occurrence (Shimazaki and Nakata, 1980; Thatcher 1984; Bakun and McEvilly 1984), subsequent studies demonstrated that predictions based on the concepts of quasi-periodic seismic cycles typically do not pass statistical tests (Rong, Jackson, and Kagan 2003; Jackson and Kagan 2006), leading to strong criticism of these concepts (Kagan, et al., 2012; Geller et al., 2015). These critiques notwithstanding, studies based on paleoseismic and instrumental data indicate that quasi-periodic earthquakes of moderate and large magnitudes may occur on certain types of faults, such as isolated,
geometrically simple continental faults or oceanic strike-slip faults (McGuire 2008; Parsons 2008; Scharer et al. 2010; Berryman et al., 2012; Cochran et al., 2017; Howarth et al., 2018).

Clustering: A well-known property of earthquakes is that they cluster in space and time, and aftershock sequences are the most ubiquitous example of this behavior. The temporal decay of aftershocks follows a power law, discovered by Omori after the 1981 Nobi earthquake and later modified by Utsu (Utsu et al. 1995); $N(t) = K/(t + c)^p$ where $N$ is the number of earthquakes per unit time, $t$ is the time from the mainshock, and $K$, $c$ and $p$ are constants; $p$ is generally close to 1. The number of aftershocks grows exponentially with mainshock magnitude (Utsu and Seki 1955); most aftershocks occur on the mainshock fault and within few rupture lengths, but aftershocks have also been observed at distances of several fault lengths up to thousands of kilometres (Felzer and Brodsky 2006). Several physical mechanisms have been proposed to explain aftershocks: static stress transfer from the mainshock fault (King et al 1995; Das and Scholz 1981) are invoked to explain the spatial distribution of aftershocks, and when combined with the response of a population of seismogenic sources governed by rate-state friction (Dieterich 1994), result in time-dependent seismicity rates in agreement with Omori’s law. Other processes have been invoked to explain aftershock triggering: these include dynamic stresses (Felzer and Brodsky 2006), which can trigger seismicity at much greater distance from the mainshock than static stress changes; time-dependent processes such as afterslip (Perfettini and Avouac 2004; Schaff et al., 1998), fluid flow (Cocco 2002), and viscoelastic relaxation (Freed and Lin 2001).

Poissonian: Once earthquake clusters are removed from a seismicity catalog, the remaining events appear to exhibit Poisson behavior (Gardner and Knopoff 1974; van Stiphout et al., 2012), or in other words, they occur randomly in time. However, the validity of the declustering methodology and the conclusion that seismicity is Poissonian has been questioned (Luen and Stark 2012; Mulargia et al., 2017). The Poisson distribution serves as a null hypothesis for testing whether a sequence exhibits clustering or periodicity.

**Insight from earthquake source simulations and future progress**

What are the physical origins of the Gutenberg-Richter relation and the Omori-Utsu law of aftershock decay? Numerical modeling of earthquake source processes has already started to address these questions.

The earliest simulations of earthquake sequences (MMC2.2) dealt with planar faults in a homogeneous elastic medium, with some small degree of large-scale heterogeneity, such as depth-dependent frictional parameters (Tse and Rice 1986; Rice 1993). When correctly discretized, these rate-and-state-based friction models produced sequences of quasi-periodic, system-size, and hence characteristic, events (Rice 1993). Later, it was established that even in such relatively homogeneous models, some complexity in the form of smaller events appears when the simulated domain is much larger than the nucleation size (Shaw and Rice, 2000; Lapusta and Rice, 2003).

In contrast, inherently discrete models of faults such as cellular automata (Olami et al., 1992) or ensemble of blocks connected by springs (Burridge and Knopoff, 1967) produced a power-law distribution of earthquake size, analogous to what is observed in nature. Nominally continuum models can also produce a range of earthquake sizes if they are not correctly discretized, since numerical cells can fail independently of one another in a manner analogous to discrete models (Rice, 1993; Lapusta et al., 2000), a result recently confirmed by a benchmarking exercise by the Simulations of Earthquake Sequences and Aseismic Slip (Erickson et al., 2018). However, note that under-resolved simulations do not provide a correct solution to the elastic continuum equations they are built to solve.
Rupture models including a degree of heterogeneity produce a more realistic distribution of earthquake sizes; in fact, discrete models can be seen not as an approximation of a continuum, but rather as highly heterogeneous system, with the frequency-magnitude distribution reflecting the range of scales in the underlying heterogeneity (Ben-Zion and Rice, 1993, 1995; Ben-Zion, 2008). Indeed, SEAS simulations with spatial variations in frictional properties (such as a fractal distribution of the characteristic slip distance of the rate-and-state friction, which results in a fractal distribution of the nucleation size on the simulated fault) can reproduce realistic magnitude distributions (Hillers et al., 2006).

While the spatial distribution of frictional properties is not well constrained, an undisputed source of fault heterogeneity is geometrical roughness (Candela et al. 2009), an aspect that has recently seen significant progress in terms of simulation capabilities (MMC2.1, MMC2.2). In particular, fault roughness promotes rupture arrest hence generates a distribution of sub-system size ruptures (Fang and Dunham 2013; Tal and Hager, 2016), as so far explored in 2D simulations. The same roughness characteristics may lead to less diverse behavior in 3D models, since unfavorable patches block the 1D fault in a 2D simulations (insuring rupture arrest) but would become compact patches in 3D that the rupture can surround and continue, although it is clear that sufficient degree of roughness would result in diverse sequences even in 3D simulations. The large and heterogeneous stresses after rupture on a rough fault also provide an explanation for aftershocks, both on-fault and off-fault.

Alternatively, or in addition to heterogeneity on a single fault, the collective behavior of earthquakes on a regional scale is affected by the distribution of fault sizes; for example, an ensemble of faults each with characteristic behaviour can presumably generate seismicity with a wide range of scales and an observed power-law distribution. It is also possible that mature vs. immature faults (MMC3.1) are characterized by different levels of roughness and heterogeneity, and hence have different local size distributions that add up to the Gutenberg-Richter relation over larger, fault-network scales.

The high resolution required to correctly model the frictional behaviour of faults in SEAS-like simulations limits the geographical area and number of faults that can be included in a fully resolved earthquake source model. For example, a model of the California fault system, as represented in the Uniform California Earthquake Rupture Forecast (Field et al., 2014, 2015, 2017) and discretized with a cell size of 1-10 m, would require $10^8-10^{11}$ fault elements, and the number of just static stress interaction between elements scales with the number of elements squared. Simulations of well-resolved earthquake sequences comprising a large number of earthquake rupture events - which would be needed to match the statistical observations discussed - clearly exceed current computational capabilities.

To overcome this limitation, the so-called earthquake simulators (MMC2.4) discretize regional fault networks using a much larger cell size (and hence making the model inherently discrete, as discussed above) and sacrifice a number of the physical aspects such as wave-mediated stress transfer or aseismic fault slip (Tullis et al. 2012b); a set of such models developed for California presented good agreement with the statistical properties of observed seismicity (Tullis et al. 2012a). In spite of their limitations, these models provide insight into the origin of the statistical behaviour of seismic sequences and the role of geometric fault complexity.

Further development of SEAS models to include realistic features on a single fault segment (MMC2.2) and to study how the behavior scales with increasingly more complex and realistic fault networks (MMC2.4) will be essential to make continuing progress. Given that all potentially relevant fault and bulk processes (MMC1) cannot be resolved in sufficient detail at the scale of fault networks, progress in using smaller-scale models to develop scale-appropriate constitutive relations for simulations at the fault-
network scale becomes key. These advances can be combined with the idea of coupling models at different scales, including the use of machine learning (MMC2.5). Two questions are of particular interest. The first one is the identification of the dominating processes for earthquake interaction, as in part expressed in the Omori’s aftershock decay law, which can be explored by such simulations, as already discussed in MMC2.3. The second one is the properties of largest events in the system and whether they could in fact be quasi-periodic, or different from the overall statistics in other ways. There are a number of physical reasons for the largest events to be special, due to their potential preferential occurrence on plate-boundary, mature faults with low-heat, low-stress operation which is special compared to other faults in the crust (MMC3.1).

Key future challenges:
- Use SEAS modeling on single and several faults to study various proposed interaction mechanisms of earthquake rupture events (MMC2.3) and determine dominating mechanisms that need to be included in regional-scale simulations;
- Use SEAS models developed for several segments as well as improved earthquake simulators developed for fault networks (MMC2.3) to investigate the properties of the largest events that occur on mature plate-boundary faults with low-heat, low-stress operation (MMC3.1) and whether they deviate from the overall properties of seismicity, e.g., by being quasi-periodic;
- Investigate the differences between the event statistics on mature vs. immature faults.

Cross-cutting themes in MMC3.1-3.5
- Using multiple types of observations to constrain multi-physics modeling; developing simplified representations of coupled physical processes that are consistent with observations, for use in larger-scale models;
- Elucidating similarities and differences in the structure and properties between mature vs. immature faults as well as continental vs. megathrust environments;
- Using modeling outcomes to verify and improve methods employed for observational inferences which are often based on highly simplified models.