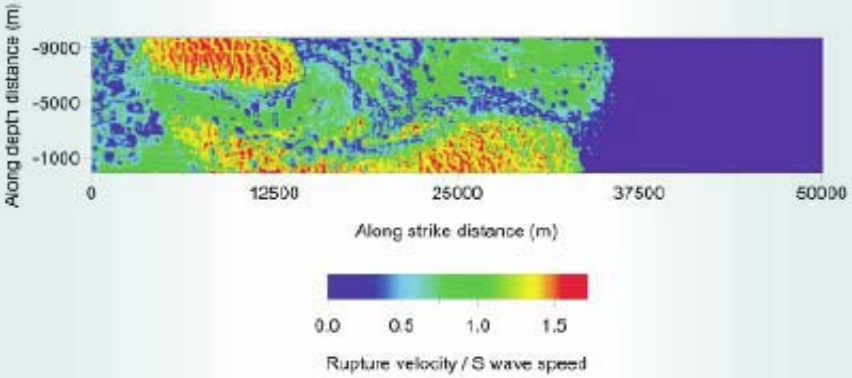
The Mechanics of Faulting: From Laboratory to Real Earthquakes



Editors

Andrea Bizzarri

Harsha S. Bhat

RESEARCH SIGNPOST

The Mechanics of Faulting: From Laboratory to Real Earthquakes

Editors

Andrea Bizzarri

Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna
Via Donato Creti, 12, 40128 Bologna, Italy

Harsha S. Bhat

Institut de Physique du Globe de Paris, 1 rue Jussieu
75238 Paris Cedex 05, France

Research Signpost, T.C. 37/661 (2), Fort P.O., Trivandrum-695 023
Kerala, India

Published by Research Signpost

Research Signpost
T.C. 37/661(2), Fort P.O.,
Trivandrum-695 023, Kerala, India

E-mail IDs: admin@rsflash.com; ggcom@vsnl.com
signpost99@gmail.com; rsignpost@gmail.com

Websites: http://www.ressign.com/home.aspx
http://www.tmres.com
http://www.journals.academicpursuits.us
http://www.signpostebooks.com

ISBN: 978-81-308-0502-3

Andrea Bizzarri

Andrea Bizzarri is a physicist with a Ph. D. degree in Geophysics (Università degli Studi di Bologna, Bologna, Italy). His expertise spreads over many different aspects of earthquake source dynamics, including the study of fault triggering phenomena; the modeling of the physical processes occurring during faulting; the inference of fault governing equations from laboratory experiments. Fundamental problems in theoretical seismology are of great importance in his research. As a part of his research, Bizzarri has been actively involved in the development of massively parallel numerical algorithms to be used in the numerical models. He is a Full Researcher at the Istituto Nazionale di Geofisica e Vulcanologia – Sezione di Bologna (INGV).

For further information: http://www.bo.ingv.it/~bizzarri; bizzarri@bo.ingv.it

Harsha S. Bhat

Harsha S. Bhat is a solid mechanician with interests in earthquake source dynamics. He obtained his Ph. D. degree in Mechanical Sciences in 2007 from Harvard University under the supervision of Prof. James R. Rice and Dr. Renata Dmowska followed by postdoctoral work at University of Southern California and California Institute of Technology under the supervision of Prof. Charles G. Sammis and Prof. Ares J. Rosakis from 2007–2011. In 2012 he joined Institut de Physique du Globe de Paris (IPGP) as a Centre National de la Recherche Scientifique (CNRS) permanent research fellow. His interests lie in earthquake source mechanics, experimental mechanics, fracture dynamics and dynamic damage mechanics.

For further information: http://www.ipgp.fr/~bhat; bhat@ipgp.fr

Contents

Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

The Mechanics of Faulting: From Laboratory to Real Earthquakes, 2012: 1-4
ISBN: 978-81-308-0502-3 Editor: Andrea Bizzarri & Harsha S. Bhat

1. Introduction

Andrea Bizzarri[1] and Harsha S. Bhat[2]
[1]*Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, Italy*
[2]*Institut de Physique du Globe de Paris, Paris, France*

Understanding the physics governing the seismogenic systems is of pivotal importance not only in the general contest of the Earth Sciences, but also in the framework of hazard assessment and risk mitigation scenarios. While statistical models accept that several aspects of the phenomenon under study are out of range and they are replaced by random processes whose behavior is described in probability terms, physical models focus on the understanding and the prediction of all the details of the considered process. In the recent years it has become clear that many competing chemical and physical mechanisms can take place during faulting (e.g., Bizzarri, 2010). On the other hand it is also evident that a large number of theoretical models for fault have been proposed, as discussed in details in Bizzarri (2011).

The study of the fault mechanics nowadays relies on the combined efforts from laboratory experiments, geological and field observations, numerical experiments and theoretical models. All these approaches have specific limitations and advantages, so that they are somehow complemented.

The main purpose of the present book is to careful review all the most prominent aspects of the above–mentioned lines of research, in order to give to the readers a comprehensive picture of the state of the art. Of course, all the research fields are in continuous and in some cases fast development; this is one of the most exciting aspects of the earthquake physics.

Correspondence/Reprint request: Dr. Andrea Bizzarri, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, Italy. E-mail: bizzarri@bo.ingv.it

In chapter 2 Festa and Zollo present some methods and strategies to retrieve the earthquake rupture slip history on the fault plane from the inversion of observations recorded at the Earth surface. After the definition of the representation theorem they discuss the influence of the Green's functions and the source time function on the seismic/geodetic observations. To numerically solve the forward problem, they then describe the effect of the uniform and non–uniform discretizations on the computed displacement.

Robinson in chapter 3 reviews the main results pertaining to supershear earthquakes, i.e., ruptures which develop with a velocity greater than the *S* wave speed. Early theoretical work suggested that supershear ruptures are possible but unlikely to occur in the natural environment. This view was essentially unchanged until the beginning of the 21st Century, when observations of several earthquakes suggest that supershear earthquakes could occur. This observation led to several carefully constructed lab experiments that aimed to produce supershear ruptures similar to those observed by seismologists. The majority of seismologists now accept that supershear earthquakes can and do occur, and now plenty of studies try to understand the circumstances under which these earthquakes occur and the effects that they may produce.

Numerical models of dynamic fault rupture provide a convenient framework to investigate the physical processes involved in the fault rupture during earthquake and the corresponding ground motion. This is the main focus of the chapter 4, where Dalguer reviews the state of the art of the present–days numerical techniques to solve the fundamental elastodynamic equation for fault, which is the counterpart for fault systems of the second law of dynamics in classical mechanical system. In particular, two approaches of fault representation are formulated, the first is the so–called traction–at–split–node (TSN) scheme, which explicitly incorporates the fault discontinuity at velocity (and/or displacement) nodes, and the second is the inelastic–zone scheme (i.e., the so–called stress–glut (SG) method), which approximates the fault–jump conditions through inelastic increments to the stress components.

In chapter 5 Madariaga reviews the most important results obtained in fracture mechanics and seismology for antiplane cracks (i.e., mode III) in the first years of earthquake dynamics. First, he studies a model of a rupture front that moves at constant speed with the load following it, as in rupture pulses. Then he considers a model of a rupture front that appears spontaneously and propagates afterwards at variable speed. Madariaga shows that these two types of rupture behave differently as speed increases. Steady propagating rupture pulses, like dislocations, can not approach the terminal speed. On the

other hand spontaneous ruptures moving at variable speed under time independent load increase their speed steadily until they reach the shear wave speed. He then discusses the seismic radiation, the generation of high frequency waves by seismic ruptures. He also shows that the radiation can only occur when rupture speed changes. Cracks moving at constant speed simply "pull" their static elastic field behind the rupture front without emitting seismic waves. Madariaga obtains a complete solution for radiation from an arbitrarily moving crack in 2–D. Finally the chapter introduces the friction following the original work by Ida (1972) who studied different slip–dependent friction models.

In Chapter 6 Lapusta and Barbot discuss fault models based on rate– and state–dependent friction formulations that are capable of reproducing the entire range of fault slip behaviors, fueled by the increasing stream of high–quality laboratory experiments, observational data, and computational resources. A wide range of earthquake complexity can be explained within the standard (Dieterich–Ruina) rate– and state–dependent models due to their rich stability properties combined with the interactive nature of long–range elastic interactions and inherent nonlinearity of frictional response. Some of the insights provided include that the longer history of the fault the shear stress becomes higher at an asperity as well substantially reducing its effect on dynamic rupture. Another insight is that seismic and aseismic slip can occur in the same region of the fault at different times consequently resulting *i*) In a model of small repeating earthquakes a large fraction of slip in the earthquake producing patches can be aseismic, resulting in the observed scaling of seismic moment with the recurrence time; *ii*) When a rate–and–state nucleation site is perturbed by a favorable stress change, of the kind that should speed up the upcoming earthquake, the seismic event can be delayed instead due to the resulting aseismic transient slip. With the rate–and–state modeling of the Parkfield sequence of earthquakes they demonstrate the possibility of creating comprehensive physical models of fault zones that integrate geodetic and seismological observations for all stages of the earthquake source cycle.

In chapter 7 Noda illuminates about the fact that multiple lines of evidences discovered over the last couple of decades that the coseismic weakening of a fault which is much more dramatic than what is predicted from the conventional friction laws verified at low slip rates. Such a weakening undoubtedly affects the dynamic rupture propagation in various ways. For example, cosesimic weakening considered in a framework of rate–weakening has been shown to play an important role in determining the manner of dynamic rupture propagation (crack–like versus pulse–like rupture) given a pre–stress level. Moreover in the sequence of earthquakes,

the pre–stress is affected by the coseismic frictional resistance. In this chapter some of the recent studies on the significance of high velocity friction are critically reviewed.

References

1. Bizzarri, A. (2010), Toward the formulation of a realistic fault governing law in dynamic models of earthquake ruptures, in *Dynamic Modelling*, edited by A. V. Brito, pp. 167–188, ISBN: 978–953–7619–68–8. InTech (Available at http://www.intechopen.com/books/dynamic-modelling/toward-the-formulation-of-a-realistic-fault-governing-law-in-dynamic-models-of-earthquake-ruptures).
2. Bizzarri, A. (2011), On the deterministic description of earthquakes, *Rev. Geophys.*, *49*, RG3002, doi: 10.1029/2011RG000356.
3. Ida, Y. (1972), Cohesive force across the tip of a longitudinal shear crack and Griffith's specific surface energy, *J. Geophys. Res.*, *77*, 3796–3805, doi: 10.1029/JB077i020p03796.

Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

The Mechanics of Faulting: From Laboratory to Real Earthquakes, 2012: 5-62
ISBN: 978-81-308-0502-3 Editors: Andrea Bizzarri and Harsha S. Bhat

2. From data to source parameters: Kinematic modeling

Gaetano Festa and Aldo Zollo
Dipartimento di Scienze Fisiche, Università di Napoli, "Federico II", Naples, Italy

Abstract. In this chapter we present methods and strategies to retrieve the earthquake rupture slip history on the fault plane from the inversion of observations recorded at the Earth surface. After the definition of the representation theorem which relates the space-time evolution of the slip to the displacement at the surface, we discuss the influnce of the Green's functions and the source time function on the seismic/geodetic observations. To numerically solve the forward problem, we then describe the effect of the uniform and non-uniform discretizations on the computed displacement.

As concerns the inverse problem, we investigate the slip parametrization and its crucial role in retrieving the slip history on the fault plane. We discuss the possible choices for the parametrization, the interpolation strategies and the objective function to be minimized for the retrieval of the best-fit solution. We compare linearized versus global non-linear approaches for both solution search and uncertainty assessment. From careful inspection of data, we finally present some recent techniques based on time-reversal, back-projection and beamforming which are able

Correspondence/Reprint request: Dr. Gaetano Festa, Dipartimento di Scienze Fisiche, Università di Napoli, "Federico II", Naples, Italy. E-mail: festa@na.infn.it

to map selected portions of data into the source region to retrieve overall rupture characteristics, such as the fault patches which emitted the largest amount of radiation and the average rupture velocity.

We then discuss the seismic, geodetic, geological and tsunami observations and their potentiality in constraining the kinematic features of the rupture. Finally, two examples (the Tottori and Loma Prieta earthquakes) are presented.

1. Introduction

Earthquake ruptures are generated by the relative motion (slip or dislocation) of crustal blocks, which occurs along fault surfaces embedded in the shallower Earth fragile layer. The faulting process involves a complex energy balance between the dissipation occurring along the fault and in the surrounding volume, the elastodynamic flux near the rupture tip and the radiated field, which is represented by the seismic waves propagating away from the source. Seismic waves carry on the information about the source processes and their recording at or nearby the Earth surface is used by seismologists to infer the rupture properties in terms of its space-time evolution. Geological observations indicate that the thickness of a fault is much smaller (tens to few hundreds of meters) than the wavelengths at which we investigate the rupture processes, justifying the assumption of a fault zone as an infinitely thin surface where the different dissipation mechanisms occurring during the earthquake rupture are homogenized.

The objective of this chapter is to describe methods and strategies to reconstruct the history of the rupture process on the fault plane from observations at the Earth surface. With this aim, we renounce to the description of the initial stress state to which the fault is subject before the dynamic rupture as well as to the characterization of the energy balance during the rupture propagation. We limit here the analysis to the effects of such an energy balance on the relative motion of the two sides of the fault itself. The space-time distribution of the slip is here referred to as the kinematic description of the rupture. Although it provides a partial view of the rupture process as compared to a dynamic model, kinematic descriptions of seismic ruptures give important insights into the physics of the rupture process and provide reliable estimations of the ground motion and its space-time variability.

Along this chapter, we discuss the methodologies to reconstruct the kinematic properties of the seismic rupture from the observations at the Earth surface. In the first section, we present the forward modeling, that is the computation of synthetic seismograms from a source kinematic model. After introducing the representation theorem, which relates the slip to the displacement measured at the Earth surface, we discuss the role of the

Green's functions and the source time function on the final displacement. Finally we describe the numerical solution of the representation integral and the role of the discretization in providing reliable waveforms. In the following section, we introduce the inverse problem, through the slip parametrization, the choice of the objective function, the inversion strategies with linearized versus non-linear approaches, the assessment of the uncertainties on the slip function. We finally illustrate some recent techniques based on the back-propagation of seismic waves into the source region. In the section 3, we present the seismic, geodetic, geological and tsunami observations and their potentiality in constraining the kinematic features of the rupture. Finally two examples of kinematic inversion are described in detail: the 2000, Tottori, Japan earthquake and the 1989, Loma Prieta, California earthquake.

2. Forward modeling

2.1. Representation of seismic sources

Let us consider a volume V internal to the Earth, bounded by the surface $S=\partial V$ inside which the linear elastodynamics hold, except for the fault surface Σ, along which slip occurs. We assume homogeneous boundary conditions on S, which are, for example, representative of the traction-free Earth surface. We indicate with Σ^1 and Σ^2 the two lips of the fault which move away from each other, and we define the normal **n** to the fault as the normal to Σ^1 entering Σ^2 (Figure 1). Since the amount of slip is small (centimeters to meters) as compared to the size of the process zone, which is representative of the space scale at which both the dissipation and radiation occur (tens to few hundreds of meters), we can assume the small-strain approximation to hold and refer all quantities on the faults to a reference

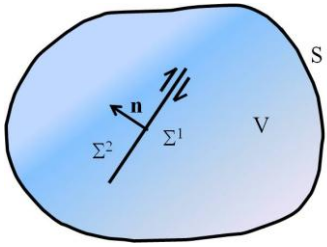


Figure 1. A seismic fault is here represented as a surface inside a volume V along which slip occurs. We separate the two lips of the fault, to interpret the displacement discontinuity across such a surface.

configuration defined at time zero, when the two lips of the fault are at their original position. Due to the slippage, the kinematic quantities (the displacement and its time derivatives) are discontinuous across the surface Σ. We define the slip function as the difference of the Lagrangian displacement **u** across the two sides of the fault surface $\delta\mathbf{u} = \mathbf{u}^2 - \mathbf{u}^1$. On the other hand, for spontaneous ruptures the traction must be continuous across Σ. From Betti's theorem, the displacement seismogram observed at a location **x** inside the Earth or at its surface can be computed as the convolution of the slip function with the elastic response of the propagation medium [1]:

$$u_m(\mathbf{x},t) = \int_{-\infty}^{+\infty} d\tau \int_{\Sigma} \delta u_i(\vec{\xi},\tau)\, c_{ijkl} n_j \frac{\partial G_{mk}}{\partial \xi_l}(\mathbf{x}, t-\tau; \vec{\xi})\, d\vec{\xi} \tag{1}$$

In the above formula, referred to as the representation integral, **c** is the elastic coefficients tensor and is symmetric with respect to the exchange of all the indices. It has 21 independent components for a general elastic solid but can be reduced to only two coefficients for an isotropic medium: $c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$, where λ and μ are denoted as the Lamé constants. **G** is the Green's function tensor representing the impulse response of the medium. Specifically $G_{ij}(\mathbf{x},t;\vec{\xi})$ is the i-th component of the displacement recorded at the position **x** and at the time t, generated by an unidirectional impulse force acting in the j-th direction at $\vec{\xi}$ at time zero.

Using the property of reciprocity of the Green's functions, we can exchange the source and receiver positions yielding $G_{mk}(\mathbf{x}, t-\tau; \vec{\xi}) = G_{km}(\vec{\xi}, t-\tau; \mathbf{x})$. The quantity

$$c_{ijkl} \frac{\partial G_{km}}{\partial \xi_l}(\vec{\xi}, t-\tau; \mathbf{x}) n_j = T^G_{im}(\vec{\xi}, t-\tau; \mathbf{x}) \tag{2}$$

is the stress on the fault plane generated by an impulse force at **x**, contracted by the normal, that is the Green's traction $\mathbf{T}^G_m$ on the fault plane generated by an impulse force at **x** directed along the m-th direction. The representation theorem hence simplifies to

$$u_m(\mathbf{x},t) = \int_{\Sigma} \delta u_i(\vec{\xi}) * T^G_{im}(\vec{\xi}; \mathbf{x})\, d\vec{\xi} \tag{3}$$

where we suppressed the time dependence and replaced the time integral by the convolution operator, denoted here by the symbol *. For a dominant shear

faulting mechanism, which is representative of the seismic rupture, we assume that the rupture may only occur in mode II (in-plane: slip parallel to the rupture direction) or mode III (anti-plane: on-fault slip orthogonal to the rupture direction) [1]. The component of the slip normal to the fault is always zero, reducing to six the number of non zero components of the traction needed for the computation of the displacement.

For a linear system, it is not surprising that the displacement associated to the waves arriving at the Earth surface can be obtained by propagating the boundary conditions (here the source term) via a convolution. By duality, if the boundary conditions on the fault are kinematically described by the displacement, the displacement discontinuity or one of their time derivatives, the propagator involves the space derivatives of the Green's functions, that are the Green's tractions. Neumann boundary conditions, defining the evolution of the traction along the boundary, are instead propagated away from the fault by the Green's functions.

Formula (3) can be manipulated to represent an extended seismic source as a superposition of double-couple point sources, as it is common in seismology (body force equivalence). It is worth to note that such a body force representation is not unique on the fault plane, although the slip function is unique [1]. For instance, the radiated field generated by a double-couple source in the volume V is equivalent to a superposition of a point source plus a single couple. To write the representation integral (1) in terms of a distribution of double-couple sources, we start to note that only the components of **c** with $i \neq j$ contribute to the Green's tractions, since the slip vector lies on the fault plane and the normal is orthogonal to it. For an isotropic medium the observed displacement is indeed independent of the Lamé constant λ, leading to:

$$u_m(\mathbf{x},t) = \int_{-\infty}^{+\infty} d\tau \int_{\Sigma} \delta u_i(\vec{\xi},\tau)\, \mu \left(\frac{\partial G_{mi}}{\partial \xi_j} + \frac{\partial G_{mj}}{\partial \xi_i} \right) n_j d\vec{\xi} \tag{4}$$

The first contribution in the Green's function derivatives is the space derivative with respect to the normal direction. Its finite difference approximation can be written as:

$$\frac{\partial G_{mi}}{\partial \xi_j} n_j = \frac{\partial G_{mi}}{\partial \xi_n} \approx \frac{G_{mi}^2 - G_{mi}^1}{\Delta \xi_n} \tag{5}$$

The superscripts are referred to as the quantities computed on the two sides of the fault, while $\Delta\xi_n$ is the distance along the normal direction. The

discrete formula (5) represents the superposition of the displacements provided by a couple of opposite forces acting on the two sides of the fault, in the direction of the slip (red couple in Figure 2). As $\Delta\xi_n \to 0$, the distance between the forces composing the couple shrinks, giving rise to a moment on the fault, with in-plane forces and arm along the normal direction. Such a moment, which would tend to locally rotate the fault, is balanced by a second couple acting on a plane perpendicular to the fault. From inspection of the second term in the Green function derivatives, this couple is formed by two forces directed along the fault normal with the arm along the slip direction (green couple in Figure 2). Indicating with D_{mi} the m-th component of the displacement generated by such a double couple, the representation theorem can be simplified to:

$$u_m(\mathbf{x},t) = \mu \int_{\Sigma} \delta u_i(\vec{\xi}) * D_{mi}(\vec{\xi};\mathbf{x}) d\vec{\xi} \tag{6}$$

Following (6), the computation of the Green function's can be replaced by the displacement generated by a double couple of forces, whose orientation is defined by the slip and the fault normal vectors.

By properly changing the indices in the summation, formula (4) can be also rewritten as:

$$u_m(\mathbf{x},t) = \int_{-\infty}^{+\infty} d\tau \int_{\Sigma} \mu(\delta u_i n_j + \delta u_j n_i) \frac{\partial G_{mi}}{\partial \xi_j} d\vec{\xi} \tag{7}$$

The quantity $m_{ij} = \mu(\delta u_i n_j + \delta u_j n_i)$ has the dimension of a moment per area unit and it is denoted as the moment density tensor. It is symmetric by definition. For observer distances and signal wavelengths much larger than the size of the fault we can assume that the Green's function derivatives are

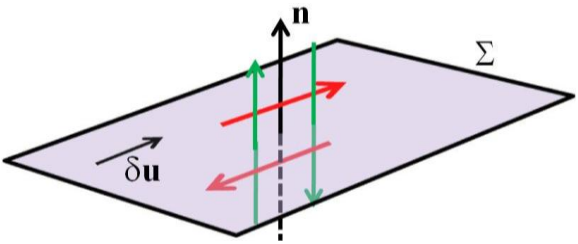


Figure 2. The double couple which can be used as an elementary source for the computation of the representation integral.

constant and bring those terms outside the surface integral (7). The integral of $\int_{\Sigma} m_{ij} d\vec{\xi} = M_0 = \mu A \delta \bar{u}$ has therefore the dimension of a moment and it is called the *seismic moment*. This quantity is an important overall measure of the earthquake size. In the far-field approximation, it is related to the amplitude of the displacement spectrum in the limit of zero frequency [1].

All the representation integrals (1-7) can be also written in the frequency domain, where the convolution reduces to a multiplication of the single Fourier transforms. For example formula (3) writes as a scalar product of the slip times the traction [2]:

$$u_m(\mathbf{x}, \omega) = \int_{\Sigma} \delta \mathbf{u}(\vec{\xi}, \omega) \cdot \mathbf{T}_m^G(\vec{\xi}, \omega; \mathbf{x}) d\vec{\xi} \tag{8}$$

The advantage of solving the representation integral in the frequency domain arises from the fact that the displacement can be computed frequency by frequency, thus making the process of inversion of band-pass filtered data more straightforward.

2.2. Green's functions

From the representation theorem, the displacement recorded at the Earth surface can be expressed as the convolution of the local source time function by the Green's tractions. Because of such a linear coupling, the determination of the source parameters from the seismograms requires an accurate modeling of the propagation effects down to the wavelengths at which we can resolve the rupture history. A complete description of the propagation medium transfer function requires the knowledge of the spatial distribution of the density, the P and S wave velocities and the anelastic attenuation parameters. Smooth wave velocity and attenuation models are generally derived from the tomographic inversions of first arrival-times and spectral amplitudes, mainly inferred from the analysis of local earthquake records. When using the ray theory for modeling arrival times and wave amplitudes, the resolution on the medium parameters critically depends on the coverage and the criss-crossing of rays, which may strongly vary in space and usually degrades at depth. In most cases, the best available model of the propagation medium is one-dimensional, which can satisfactorily represent the low frequency propagation of seismic waves up to the surface. 1D models, however, may be inadequate to describe the seismic wavefield in the case of dominant 3D effects perturbing the amplitude and the phase of the seismic waves, such as the presence of basins, strong lateral heterogeneities or steep topography [3].

In several active and well monitored areas, 3D models are available with a space resolution varying from few hundreds of meters to few kilometers, allowing for a more reliable computation of the Green's functions. It is worth to note that the inaccuracy in modeling the propagation effects introduces uncertainties in the rupture description and in the estimation of the source parameters. Such an uncertainty is difficult to be assessed since the true propagation model is unknown and the errors on the velocity model and on the location and morphology of the interfaces non linearly affect the Green's tractions computation. Several studies have shown that source models derived from the waveform inversion are strongly sensitive to the wave propagation [4,5,6]. Even when inverting synthetic waveforms the data misfit may rapidly degrade, when replacing the 3D true model with its best 1D approximation, and the estimations of the relative slip amplitudes, the rupture velocity and the rise time, e.g., the time required by the slip to reach its final value, are strongly sensitive to the considered Earth model. This effect becomes more pronounced at shorter wavelengths [6]. The poor knowledge of the propagation effects hence prevents from a complete match between real and simulated waveforms even if we arbitrarily increase the number of degrees of freedom of the kinematic rupture model.

Several techniques are widely used for the computation of the Green's tractions in axisymmetric media, providing the visco-elastic dynamic response of a 1D medium. For horizontally layered media, the reflectivity method [7,8] solves the wave equation in the frequency-wavenumber domain by propagator matrix techniques, which account for the continuity of displacements and tractions at the model interfaces. Reflectivity can be efficiently coupled with the discrete wavenumber decomposition of the Green's functions [9]. Such an approach was implemented in the code AXITRA (http://www-lgit.obs.ujf-grenoble.fr/~coutant/axitra.tar.gz) [10] and it was recently improved by allowing to locate both source and receiver at the same depth [11]. With minor variations, the same ideas were also implemented in the frequency-wavenumber integration code (FK) of Zhu and Rivera [12]. The discrete wavenumber finite element integration [13] becomes computationally competitive and possibly faster than the above methods when the velocity varies smoothly with depth. To accurately describe a velocity gradient, in fact, reflectivity techniques have to decompose the model into very thin layers, with thickness smaller than the wavelengths of interest. The finite element integration method is efficiently implemented for the evaluation of strong motion data from a finite fault in the COMPSYN code [14].

Several full 3D computational methods and codes are available for the evaluation of the Green's tractions in complex geological structures such as the *finite-differences method* [15,16], the *spectral element methods* [17,18]

and the *discrete Galerkin methods* [19]. Many of these softwares are available at the EU-SPICE project library (http://www.spice-rtn.org/library/software.1.html). As compared to 1D solutions, the computation of a 3D seismic wavefield still remains expensive and in many cases prohibitive for building up a Green's functions archive, usable during the inversion procedure. The computation can be improved using the reciprocity of Green's functions and, preferably, solving the representation integral in the form (3). If NS is the number of sources and NR the number of receivers, the direct computation of Green's tractions requires NS runs for a fixed focal mechanism and 2NS runs, when the slip direction is not assumed *a-priori*. Using the reciprocity we can directly evaluate the tractions on the fault as the superposition of the signals provided by three orthogonal impulse sources located at each receiver, with 3NR runs. Since the number of receivers is generally small (few to several tens of stations) as compared to the number of sources (several hundreds to many thousands) the reciprocity is very often preferred to reduce the computational cost of the Green's tractions evaluation.

Frequency-wavenumber methods, although restricted to 1D wave propagation models, are also suitable for the computation of the static field, which can be used to model GPS and SAR data. The capability of 3D techniques in representing the very-low frequency spectrum down to the static field depends instead on the size of the numerical bulk and on the ability of the absorbing boundary conditions/layers to not pollute the internal solution in this frequency band.

Teleseismic Green's functions can be obtained using full 3D propagation methods in a realistic Earth model, which includes the crust, the oceans and the topography. Spectral element codes, such as SPECFEM3D-Globe (http://www.geodynamics.org/cig/software/specfem3d-globe) are available and widely used for this goal. A complete wavefield modeling is generally required for the interpretation of surface waves, which are sensitive to the shallow structure of the Earth and its lateral variations, and of body waves at regional distances, because the ray paths inside the crust and the upper mantle can be complex. When dealing with very long periods, the Green's functions can be approximated by normal modes summation. On the other hand, direct body waves recorded between 30° and 90° travel in the almost homogeneous lower mantle and the ray theory approximation is generally adequate to describe the P and SH amplitudes. The most energetic arrivals are related to the direct waves and the first reflections from the free-surface (pP and sP for P waves and pS and sS for SH waves) [20].

Under the hypothesis of linear wave propagation, the Green's functions may also be replaced by the records of small earthquakes occurring on the same fault with the same focal mechanism and the same stress drop,

commonly referred to as Empirical Green's functions (EGFs) [21]. For a large earthquake, a database of EGFs could be represented by the records of the aftershocks following the main event. The EGF behaves as a localized moment source which can replace the numerical Green's function in the integral (6), appropriately normalized by the seismic moment [22,23]. This holds for frequencies smaller than the corner frequency of the small event, beyond which the details of the rupture process of the small event become visible. Since the EGFs describe in an accurate way the propagation effects even at high frequencies, their use may broaden in principle the investigation band of the seismic source. As an example, we show in Figure 3 the comparison between observed data and synthetics obtained by a kinematic inversion of the 2008, Japan, Tokachi-Oki earthquake (M 8.1) in which the propagation has been modeled by the EGFs. In such a case the goodness of the fit is high for most of the stations up to 1 Hz.

However, several drawbacks limit the applicability of the EGFs and a special care needs to be taken when they are used for the simulation of strong motion records. Generally, accelerometric networks do not record the EGFs with the same accuracy of the mainshock, but the ambient noise very often pollutes the low-frequency band of the small event because of instrumental sensitivity and dynamic range. To improve the quality of the aftershock records, strong motion stations should be complemented with short or intermediate period sensors. A particular care should be devoted to the determination of the focal mechanism and the earthquake location, that are not an easy task for small earthquakes. Finally, EGFs may not sample uniformly the fault plane. Aftershocks are likely to occur on fault regions that have not slept during the mainshock rupture. In fault areas where the largest slip occurred, the stress level mostly dropped down to the dynamic value during the main rupture and the probability that the same patch slips again during an aftershock rupture is low. As a consequence, the number of available EGFs is smaller than what required by the synthetic simulation and their location on the fault plane may not be optimal for retrieving the details of the main rupture. Hence, specific interpolation algorithms are needed to resample the EGFs on the fault plane and to compute the representation integral up to the frequencies of interest [24,25].

2.3. Source time functions

The representation integral (1-7) relates the slip function on the fault to the displacement observed at the Earth surface. The slip $\delta\mathbf{u}$ may be an arbitrary function of position and time, that has to satisfy only three conditions. The slip modulus must vanish at time zero (no dislocation occurs

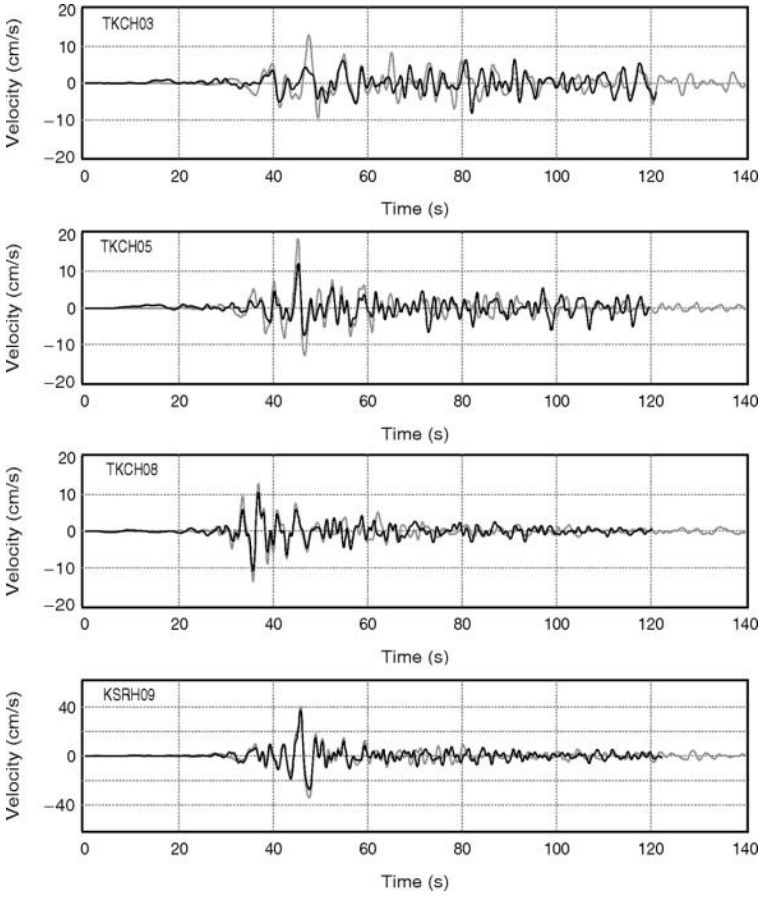


Figure 3. Comparison of observed ground velocities (gray traces) and synthetics (black traces) for the transverse component of the ground velocities at four stations which recorded the 2008, Japan, Tokachi-Oki earthquake (M 8.1). Synthetics are generated from a slip model obtained by a kinematic inversion of the strong motion data up to 1Hz coupled with Empirical Green's functions. The figure is extracted from [23], Copyright © 2011 by Seismological Society of America.

before the rupture process starts) $\delta u(\vec{\xi}, t = 0) = 0$, the modulus of the slip rate tends to zero as the time grows (the rupture cannot continue indefinitely) $\delta \dot{u}(\vec{\xi}, t \to \infty) = 0$, and the time at which each point on the fault starts to slip must satisfy a causality criterion, e.g. the rupture is not allowed to propagate

faster than the largest wave velocity in the medium (P wave speed for in-plane ruptures and S wave speed for anti-plane ruptures). However, a first limitation in the estimation of the slip function comes from the impossibility to retrieve its characteristics at arbitrary small scales due to the limited number of observations and to the uncertainties in the data, source and propagation models. As a further simplification, we assume that the slip is factorized into separated contributions:

$$\delta\mathbf{u}(\vec{\xi},t) = A(\vec{\xi})S(t - T_R(\vec{\xi});\tau(\vec{\xi}))\mathbf{r}(\vec{\xi}) \tag{9}$$

The function A represents the final amplitude of the slip: the product of A by the shear modulus, integrated on the fault plane, provides the seismic moment of the earthquake. The unit vector indicates the local direction of the slip vector, referred to as the rake. Finally, the function S accounts for the finite time during which the slip reaches its final value (the rise time τ) and describes the slip evolution with time at a specific point of the fault. S is delayed by the rupture time $T_R(\vec{\xi})$, namely the time needed by a rupture nucleating at the hypocenter to reach the point $\vec{\xi}$. Since we separated the contribution of the slip amplitude A, the function S is normalized so that its final value is the unity.

The time derivative of S, the slip velocity time function, is generally known as the source time function (STF) [1].

The representation (9) can be equivalently written in the frequency domain as:

$$\delta\mathbf{u}(\vec{\xi},\omega) = A(\vec{\xi})\hat{S}(\omega,\tau(\vec{\xi}))e^{-i\omega T_R(\vec{\xi})}\mathbf{r}(\vec{\xi}) \tag{10}$$

where the rupture time appears as a phase delay and $\hat{S}$ is the Fourier transform of the slip time function. In the kinematic approach, the functional form of the STF is generally prescribed and assumed to be the same for all the points on the fault plane. Such a strong assumption limits the number of parameters to be retrieved and helps in stabilizing the slip inversion but it is not necessarily consistent with the results issuing from the dynamic simulations of seismic ruptures. For instance, when modeling a heterogeneous rupture velocity with dynamic simulations, the peak, the duration and the shape of the STF becomes strongly sensitive to the acceleration and deceleration of the rupture [26].

The analysis of seismic data supports the hypothesis that most of the earthquake radiation is emitted in a rather short time as compared to the rupture duration, suggesting a pulse-like mode for the rupture propagation

[27]. As a consequence, the STFs are expected to be significantly different from zero in a limited time interval. Several functional forms for the STF have been proposed; here we recall the boxcar [28], the triangle [29], the exponential [30], the power law [6] and the Yoffe [26,31,32] functions among many others. We summarize their analytical expressions in Table 1 and plot their shape in Figure 4, left panel (for the power law we set the exponent p=1). Moreover, in the middle panel of Figure 4, we plot the amplitude spectrum of the STFs and in the right panel, its time integral, which represents the slip evolution with time. With the exception of the exponential function, whose spectrum slowly decreases with the frequency, all the STFs show an almost flat spectrum up to a given corner frequency, whose reciprocal is proportional to the duration of the STF. At frequencies larger than the corner frequency, the spectral decay looks different from one STF to another. The holes of the boxcar and triangle STF spectra are related to the zeros of the sinus function and are twice more frequent for the boxcar as compared to the triangle. The high frequency decay of the power law practically reduced to an ω^{-2} decay for the exponent value p=1, the same followed by the envelope of the triangle spectrum. The Yoffe function, finally, describes the slip velocity solution of a steady-state rupture pulse [33,34]. Although consistent with the rupture dynamics, the it shows a pronounced singularity at its beginning resulting into a larger high frequency

Table 1. Analytical expressions of several source time functions. The functions are normalized such that its integral is one over the duration. H is the Heaviside step function, while the indicator function $I(t,\tau)=H(t)-H(\tau-t)$ represents a time function which is one in the range $[0,\tau]$ and zero elsewhere. The constant $C_L(p)$ depends on p and its value is $C_L(p)=\sum_{n=0}^{5-p}\binom{5-p}{n}\frac{(-1)^n}{n+p+1}$.

Source Time functi	Analytical expression
Boxcar	$\frac{1}{\tau}I(t,\tau)$
Triangle	$\frac{4}{\tau^2}\left[tI\left(t,\frac{\tau}{2}\right)+(\tau-t)I\left(t-\frac{\tau}{2},\frac{\tau}{2}\right)\right]$
Exponential	$\frac{2}{\tau}e^{-\frac{2t}{\tau}}H(t)$
Power law	$\frac{1}{\tau C_L(p)}\left(\frac{t}{\tau}\right)^p\left(1-\frac{t}{\tau}\right)^{5-p}I(t,\tau)$
Yoffe	$\frac{2}{\pi\tau}\sqrt{\frac{\tau-t}{t}}I(t,\tau)$

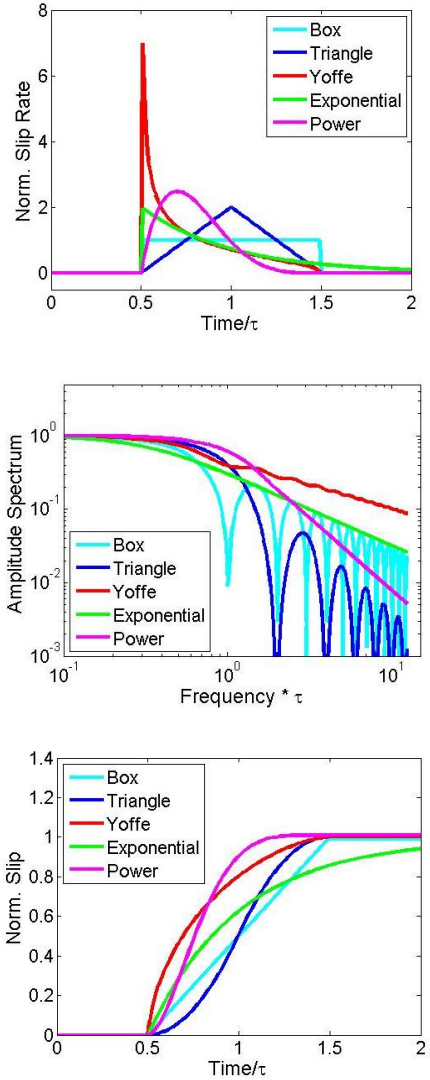


Figure 4. Left panel: Several source time functions as a function of time, centered in t=τ. Mid panel: Amplitude spectra of the STFs represented in the left panel. Apart from the exponential function, all spectra are flat up to a corner frequency which is comparable with the duration of the STF. Right panel: Slip source time functions obtained by integration of the STFs.

content in the spectrum (its spectrum decays slower with frequency than the other spectra). Such a function+ hence produces larger amplitudes in the high frequency strong motion, which could be reduced by the convolution with a triangle function [32].

Many inversions estimated the average rise time associated with M<7 to be of the order of 1 second or smaller, although this value is affected by large uncertainties. This time scale is generally smaller than the minimum period at which the kinematic rupture model is estimated from the inversion of strong motion data. Excluding the exponential function, all the other proposed STFs behave as a low-pass filter with the cut-off frequency linearly related to the STF duration. The choice of a specific STF is therefore not expected to significantly influence the inversion results. Most of the variations may affect the wavelengths whose associated frequencies are close to the cut-off frequency of the STF. Unfortunately, such a conclusion does not hold for strong motion simulations, where we are interested into a correct modeling of the high-frequency content and its decay with distance. In such a case the choice of the STF may become critical.

Some authors have proposed to relax the assumption of a fixed STF and to use multiple time windows in the attempt of discretizing a complex STF with elementary master pulses. As an example, some authors [35] superimposed a series of overlapping triangles to retrieve the shape of the STFs from the waveform inversion. Afterwards, the results of the kinematic inversion obtained with a single STF were compared to the multiple time window method [36]. In the latter case the number of free parameters significantly increases allowing for a larger flexibility in reconstructing the complexity of the STF. On the other hand, however, the inverse problem tends to become strongly ill-posed. As a conclusion, rarely the use of a multi-window approach produces better results than an *a-priori* assumed STF in retrieving the seismic moment and the average velocity of earthquake ruptures [37].

2.4. Numerical solution of the representation integral

To numerically compute the representation integral for the simulation of the displacement at the Earth surface, the fault is discretized in a grid, on which the slip function and the Green's tractions are projected. The space step of this grid is related to the minimum wavelength that the seismic data are able to resolve. To reduce the spurious oscillations from numerical dispersion and to have a correct space reconstruction of all the source contributions down to a minimum wavelength λ^*, we need a space step which is 5-15 times smaller than λ^* [6,38]. This is because the apparent wave speed depends on the specific source-receiver configuration and it is significantly

reduced when the rupture propagates away from the receiver (anti-directivity effect).

Generally the fault surface is subdivided in small areas (subfaults) inside which the source parameters and the Green's tractions are assumed constant and evaluated at the center of the subfault. The representation integral is then discretely replaced by a finite summation over the single subfaults:

$$\mathbf{u}(\mathbf{x},t) = \sum_{n=1}^{N} A_n \mathbf{r}_n \cdot \int_{-\infty}^{+\infty} S(t' - T_{Rn}, \tau_n) \mathbf{T}_n^G(\mathbf{x}, t - t') dt' \tag{11}$$

The total displacement is hence obtained as summation of the displacements provided by the single subfaults, each of them contributing with its own moment and its own STF. In the frequency domain, the integral can be solved for each frequency through the formula:

$$\mathbf{u}(\mathbf{x},\omega) = \sum_{n=1}^{N} A_n \hat{S}(\omega, \tau_n) e^{-i\omega T_{Rn}} \mathbf{r}_n \cdot \mathbf{T}_n^G(\mathbf{x}, \omega) \tag{12}$$

In the formulas (11-12) the subscript n indicates the value referred to the n-th subfault and N is the total number of subfaults. Since we assume that the kinematic fields and the tractions are constant inside each subfault, they are not continuous across contiguous subfaults. Small subfault sizes are hence required to make this jump negligible and not perturbing the synthetic waveforms. Better interpolations can be realized by making the slip and the Green's tractions (and eventually their derivatives) continuous along the fault. A finite-element discretization of the fault, for example, insures the continuity by defining the physical quantities at the nodes of the discrete grid and by building an interpolation rule to evaluate them everywhere on the fault. Such a representation has also the advantage that it can also easily handle grids whose size changes along the fault. At fixed frequency, the variability of the Green's tractions is related to the local shear wave velocity, while changes in the slip function are related to the rupture velocity, which is also dependent on the local S wave speed. As the frequency increases, the wavelength decreases: when choosing a constant grid size, the latter is associated to the maximum frequency that the data can resolve and to the smallest shear velocity value, which is generally attributed to the shallower part of the fault. Depending on the fault width and dip, the shear wave velocity can significantly vary (up to a factor of two) between the shallow and the deep regions of the same fault. In such cases we should use a finer mesh in the shallow fault region and a coarser mesh at depth. Eventually this gridding scheme needs to be refined as the frequency increases allowing for a

faster forward problem computation than taking the grid size fixed with the frequency. A finite element discretization can be obtained by a fault triangulation [39] or a quadrangulation [6] interpolating scheme. We briefly describe the triangulation method in the frequency domain, but the same concepts with minor differences can be applied for the quadrangular finite element integration. It can also easily extend to the time domain, where a local convolution is additionally required. After the definition of a set of nodes, whose inter-distance is related to the local shear wave speed, a finite element decomposition $E_i \;_{i=1,N_E}$ can be obtained through a Delaunay triangulation. The representation integral (8) can be then decomposed onto a summation over the single triangles:

$$\int_{\Sigma} \delta\mathbf{u}\cdot\mathbf{T}^{G} d\Sigma = \sum_{i=1}^{N_E} \int_{E_i} \delta\mathbf{u}\cdot\mathbf{T}^{G} dE_i \tag{13}$$

For each triangle E_i, we define the linear mapping $\vec{\xi}(\vec{\eta})$ transforming the reference right triangle into E_i (see Figure 5:

$$\vec{\xi}(\vec{\eta}) = \sum_{a=1}^{3} \vec{\xi}_a N_a(\vec{\eta}) \tag{14}$$

where $\vec{\xi}_a$ are the coordinates of the a-th vertex in the physical domain and N_a the interpolation linear function (shape function) on the reference element:

$$N_1(\vec{\eta}) = 1-\eta_1-\eta_2; \quad N_2(\vec{\eta}) = \eta_1; \quad N_3(\vec{\eta}) = \eta_2 \tag{15}$$

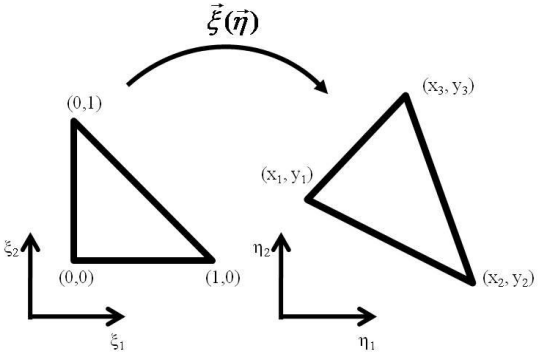


Figure 5. Linear mapping between the reference right triangle and the physical triangle.

Each function N_a has the property that its value is 1 in the vertex $\vec{\eta}_a$ and zero in the other vertices, linearly decreasing inside the reference element [40]. We then move the integral from the physical element E_i onto the reference triangle E:

$$\int_{E_i} A(\vec{\xi}) e^{-i\omega T_R(\vec{\xi})} \hat{S}(\omega, \tau(\vec{\xi})) \mathbf{r}(\vec{\xi}) \cdot \mathbf{T}(\mathbf{x}, \omega; \vec{\xi}) d\vec{\xi} = \int_{E} A(\vec{\eta}) e^{-i\omega T_R(\vec{\eta})} \hat{S}(\omega, \tau(\vec{\eta})) \mathbf{r}(\vec{\eta}) \cdot \mathbf{T}(\mathbf{x}, \omega; \vec{\eta}) J_{\vec{\xi}}(\vec{\eta}) d\vec{\eta} \tag{16}$$

$J_{\vec{\xi}}(\vec{\eta})$ the Jacobian of the transformation $\vec{\xi}(\vec{\eta})$ and since this latter is linear, $J_{\vec{\xi}}(\vec{\eta})$ is constant and twice the area of the physical triangle (the Jacobian is the ratio between the areas of the physical and the master triangles).

Let us define the function $\psi(\vec{\eta}) = A(\vec{\eta}) e^{-i\omega T_R(\vec{\eta})} \hat{S}(\omega, \tau(\vec{\eta})) \mathbf{r}(\vec{\eta}) \cdot \mathbf{T}(\mathbf{x}, \omega; \vec{\eta})$, for the sake of simplicity. We use the same shape functions to linearly interpolate the function inside the reference triangle:

$$\psi(\vec{\eta}) = \sum_{a=1}^{3} \psi_a N_a(\vec{\eta}) \tag{17}$$

With the above approximation, the integral yields:

$$\int_{E_i} \psi(\vec{\eta}) J(\vec{\eta}) d\vec{\eta} = \frac{Area(E_i)}{3} \sum_{a=1}^{3} \psi_a \tag{18}$$

Inside each triangle, the integral is the average of the function evaluated at the vertices, multiplied by the area of the triangle. Hence, a finite-element integration becomes equivalent to a triangular subfault model, in which the product of the slip by the Green tractions is averaged over the values that the function assumes at the vertices. Finally the integral on each triangle can be gathered and assembled to build up an integration rule which is based on the nodes instead of the subfsults, as it is standard for finite element methods. We can write it into two separated vectors: the point-by-point contribution of the slip function (which depends on the specific source model) and the product of the Green's tractions by the areas of the triangles, which are fixed and evaluated once at the beginning of the inversion. The displacement can be therefore written as the scalar product of these two vectors, allowing to speed up the computation.

3. Inverse modeling

3.1. *Slip parametrization*

When inverting the near-fault and/or teleseismic data for the retrieval of the earthquake source characteristics, we assume that the elastic/anelastic properties of the propagation medium and consequently the Green's tractions are known. The inverse problem therefore consists in the determination of the slip function. With the representation formula (9), the kinematic evolution of the rupture can be reconstructed through the estimation of the final slip, the rupture times, the rise time and the slip direction. As discussed above, the shape of the STF can be assumed *a-priori* or estimated, at the cost of increasing the number of the unknowns of the inverse problem. Owing to the small number of observation sites, the limited data bandwidth, and the uncertainties in the data and modeling, the source characteristics are unlikely to be retrieved at very small space scales. In addition, it is appropriate to reduce the number of the unknowns to a small number to stabilize the inversion and to increase the redundancy in the data by making the inverse problem over-determined. The set of unknowns selected to represent the slip function on the fault plane is referred to as the *slip parametrization* and the unknown final slip, rupture time, rise time and slip direction are the *parameters* to be estimated by the inversion.

The most intuitive way to limit the number of parameters is to subdivide the fault plane in large areas where the parameters are assumed constant. These areas are still denoted as *subfaults* although they should not be confused with the subfault discetization used for the computation of the representation integral. The size of the subfault is related to the minimum wavelength resolved by the data and the parameters grid in the inverse problem is generally one order of magnitude coarser than the discrete grid used in the computation of the forward problem. The i-th subfault is characterized by n_x parameters m_{ij}, j=1,..n_x and the total number of parameters is $n_p = n_s n_x$ where n_s is the number of subfaults. With this choice for the parametrization the value of a given parameter in a point of the fault is trivially the value that this parameter assumes inside the subfault to which the point belongs.

Another possibility is to define the parameters on n_c control points $\vec{\xi}_1, \vec{\xi}_2, ..., \vec{\xi}_{n_c}$, and then interpolate them on the fault plane. In such a case, the value of a parameter at a fault point can be expressed as a function of the values of the same parameter at the control nodes:

$$m_j(\vec{\xi}) = I(\vec{\xi}; m_{1j}, m_{2j}, ..., m_{n_c j}) \tag{19}$$

where I is the interpolating function and m_{ij} the value of the j-th parameter at the i-th control node. It is worth to note that no correlation is assumed among parameters, i.e., the value of the m_j parameter in any point of the fault does not depend on the values of the other parameters m_k, with $k \neq j$. As an example, the function I can be represented by a bilinear Lagrange interpolator [6] or by a bicubic spline function [41]. This latter function provides a smoother representation by also imposing the continuity of the first and the second cross-derivatives. As a drawback, when strong variations occur between two nearby points, such a representation can lead to undesired large oscillations. Finally, the unknowns can be decomposed over a set of functions which represent a discrete version of a space filter [39], allowing to rule out all the wavelengths that do not contribute to the frequency spectrum determined by the waveform data. In this case the parameter m_j is represented by a combination of the selected functions

$$m_j(\vec{\xi}) = \sum_{i=1}^{n_f} m_{ij} \psi_i(\vec{\xi}) \tag{20}$$

and the unknowns to be determined are the coefficients of the summation (20). The effect of different parametrizations on the same slip map is shown in Figure 6.

For any possible choice for the parametrization, we collect all the parameters to be determined in a single vector, that we indicate here as **m**. Among the parameters which directly appear in the representation (9), the rupture times are the most problematic to be achieved from the inversion, because it is difficult to impose the causality condition between contiguous regions of the fault [30]. For this reason the rupture times are often replaced by the rupture velocity as a parameter in the inversion. Rupture times are assumed to be related to the rupture speed through the eikonal differential equation $|\nabla T_R| = \frac{1}{v_R}$. Rupture velocity is generally preferred to rupture times because the positivity constraint intrinsically insures the causal rupture propagation. In addition, based on steady-state dynamic models we may constrain the ratio between the rupture velocity and the S wave speed as a function of the propagation mode to be consistent with the theory, and therefore limiting the exploration range of this parameter. However, since the slip function does not explicitly depends on rupture velocity, the rupture times have to computed by numerical integration of the eikonal equation, which can be efficiently performed by the finite – difference algorithm of Podvin and Lecomte [42].

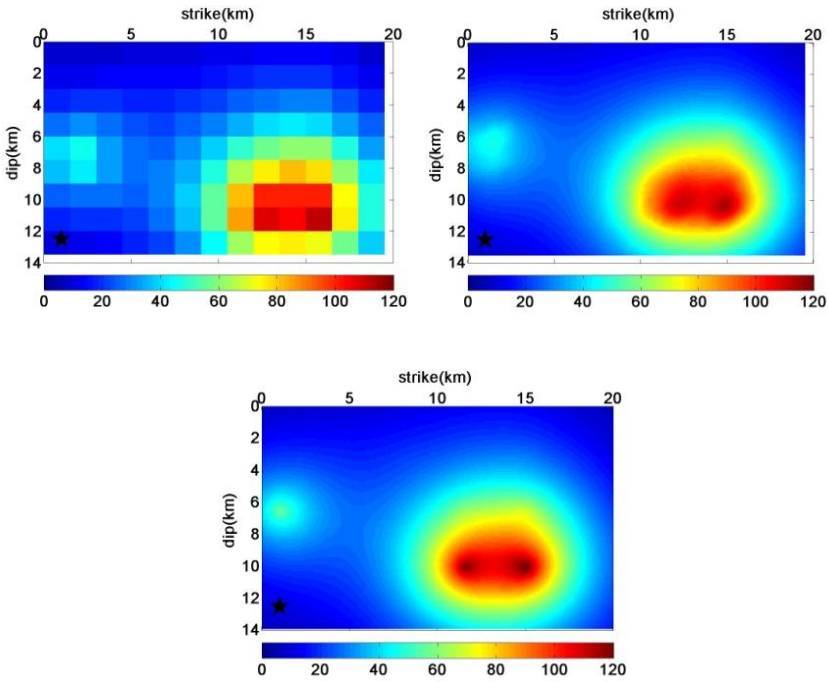


Figure 6. The same slip map represented by different parametrizations. On the left, a cell-style discretization, in the middle, a bi-cubic spline interpolation based on control nodes, on the right a superposition of overlapping Gaussian functions. The two smooth images are very similar, with differences related to the shape of the singularities.

Very often the grid size of the slip parametrization is related to the minimum wavelength that the data can resolve. Such a link is justified for the slip amplitude, which is linearly related to the observed ground motion. This is not necessarily the case for the rise time and the rupture velocity, which non-linearly depend on the data. Some authors hence have proposed the use of separate grids for the estimation of the slip amplitude, the rupture velocity and the rise time [43]. Specifically, while maintaining fixed the parametrization for the slip amplitude, the inversion starts by retrieving uniform values of the non linear parameters. In the following steps, the grid associated with these parameters is more and more refined as long as the data can justify such a complexity increase in the parametrization.

3.2. Inversion procedure

We have described the rupture kinematics using a rather small number of parameters, which have been grouped into the vector **m**. The goal of the inversion is to give an estimation of **m**, through a set of observations at the Earth surface, that we also gather into a data vector **d**. The representation integral is the physical model which links the parameter space to the data space through the non-linear application **g**, yielding the equation

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) \tag{21}$$

The function **g** is the forward problem operator, deriving from the discretization of the representation integral, as described in the section 2.4. It is a non linear function of the kinematic properties of the fault but, when the rupture times, the rise time and the rake are fixed at all the points of the fault, it linearly depends on the slip amplitude. Let us indicate with $\mathbf{d}^{obs}$ the data recorded at the surface and with $\mathbf{d}^{theo}$ the synthetic estimation obtained by the forward computation **g**(**m**). It should be pointed out that in the case of waveform data, the components of vector **d** are, for each station, the amplitudes of the digitized signal as a function of time in a given time window along the seismogram or equivalently the Fourier spectrum. The solution of the inverse problem is the set of parameters $\mathbf{m}^{+}$ such that $\mathbf{d}^{obs} = \mathbf{g}(\mathbf{m}^{+})$. Exact solutions however are unlikely to occur, due to uncertainties in the data and in the source and propagation modeling. An estimation of the true model is expected to be given by the vector **m*** such that the distance between $\mathbf{d}^{obs}$ and **g**(**m***) is minimum in the sense of a defined norm.

A first attempt to find a solution of equation (21) can be obtained by a linearization of the equation around an initial model $\mathbf{m}_0$. Using the Taylor expansion, we can replace **g**(**m**) with its first order approximation $\mathbf{g}(\mathbf{m}_0)+\mathbf{G}(\mathbf{m}-\mathbf{m}_0)$, which holds in the neighborhood of the initial model for a small perturbation $\Delta\mathbf{m}= \mathbf{m}- \mathbf{m}_0$. The corresponding linearized problem writes:

$$\mathbf{G}\Delta\mathbf{m} = \mathbf{d}^{obs} - \mathbf{g}(\mathbf{m}_0) \tag{22}$$

where the elements of the matrix **G** are the partial derivatives of the function **g** with respect to the source parameters, evaluated at the reference model $\mathbf{m}_0$:

$$G_{ij} = \left(\frac{\partial g_i}{\partial m_j} \right)_{\mathbf{m}=\mathbf{m}_0} \tag{23}$$

Depending on the specific choice of the parametrization, the partial derivatives with respect to the slip amplitude and rise time can be evaluated analytically. In most of cases, however, partial derivatives are numerically computed through their finite-difference approximation. Since the problem is over-determined, a solution for $\Delta\mathbf{m}$ can be obtained with the use of the generalized least-squares method:

$$\Delta\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T(\mathbf{d}^{obs} - \mathbf{g}(\mathbf{m}_0)) \tag{24}$$

where $\mathbf{G}^T$ is the transpose of the matrix $\mathbf{G}$. Since we deal with a small number of parameters (in the range of 10 to 10^3), the generalized inverse can be computed explicitly, through, for instance, the Cholesky decomposition (LU) or the singular value decomposition (SVD), where the components with small eigenvalues can be appropriately weighted and damped [44], accelerating the convergence of the algorithm and separating the null subspace of the parameter space [45]. The solution of the linear system (24) leads to the evaluation of the first perturbation, correcting the initial model to $\mathbf{m}_1 = \mathbf{m}_0 + \Delta\mathbf{m}$. This point can be considered as a new starting point, around which a new perturbation can be computed. The procedure can be iterated until some convergence criterion is met (e.g., the norm of the perturbation becomes smaller of an assigned threshold). The equations relating the parameters to the data can be properly weighted by left multiplying the matrix $\mathbf{G}$ by a diagonal matrix of weights, accounting for the uncertainty in the Green's functions computation or reducing the importance of large amplitudes in the data [46]. The preconditioning, however, does not change the nature of the linear system (22).

The inversion of a kinematic rupture model is known to be unstable, that is when adding a small perturbation to the data, the solution can be significantly different from the one obtained with unperturbed data. For this reason additional constraints need to be added to the linear problem (22). First, the slip amplitude, the rise time and the rupture velocity are positive quantities, defining an additional series of inequalities to be added to the linear problem. Such constraints can be included through the use of a penalization function, which sharply increases as the inequality is violated [47], through the substitution of the parameters with their logarithm [48,49], or artificially setting to zero the parameters when they assume negative values during the inversion process. This latter approach is efficiently implemented in the NNLS (Non Negative Least Squares) algorithm [50]. Second, an additional smoothing constraint is usually included to limit strong variations of the parameters between nearby subfaults or points. To maintain this condition linear, the smoothing is defined through a differencing operator

D, which equates to zero the difference between parameters from contiguous elements. The complete system becomes

$$\begin{pmatrix} \mathbf{G} \\ \varepsilon\mathbf{D} \end{pmatrix} \mathbf{m} = \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix} \tag{25}$$

where ε represents the weight of the smoothness as compared to the data fit. The choice of the smoothing value is critical: the larger the smoothness, the blurred the final solution and the worse the misfit between data and synthetics. Hence the smoothing value comes out from a compromise between the stability of the solution and the degradation of the fitness function. Usually the smoothing contribution ranges between 1% and 10% as compared to the misfit value. Several authors [51,52] propose the use of a statistical criterion, such as the Akaike Bayesian information criterion (ABIC) [53], to define the optimal choice for the smoothing coefficient.

The main drawback of the linearized approach is that the solution is critically dependent on the initial reference model. The risk to fall in a local minimum, which is far from the global one, is high and it is expected to increase as the number of parameters becomes large. Such a problem can be partially overcome, by testing a large amount of initial models and by comparing the final misfits values reached by the algorithm. Such an approach, although interesting, may become computationally expensive because of the numerical computation of $\nabla_{\mathbf{m}}\mathbf{g}$.

Instead of searching for solutions of a linear problem, gradient-free global exploration methods directly compare the observations $\mathbf{d}^{\text{obs}}$ with the theoretical estimation $\mathbf{d}^{\text{theo}}$. They explore the parameter space, searching for the solution $\mathbf{m}^*$ (or the solutions) which minimizes the distance between data and synthetic predictions $S(\mathbf{m}) = \left\| \mathbf{d}^{obs} - \mathbf{d}^{theo}(\mathbf{m}) \right\|$, defined in the sense of some norm. Such a function is referred to as the *cost*, *misfit*, *fitness* or *objective* function. The use of global methods have recently become attractive because the computational resources allow for the exploration of a huge amount of models by the fast computation of the forward problem and the parallelization.

When the number of parameters is very small (up to five), the exploration of the parameter space can be systematically carried out on regular grids or by a Montecarlo sampler. As the dimension of the parameter space increases, the exploration needs to be appropriately driven, avoiding the sampling of areas where the misfit is large and densifying the research in regions of smaller misfit. We will describe here three techniques: the simulated annealing, the genetic algorithm and the neighborhood algorithm,

which are commonly used in the investigation of the kinematic source parameters, without giving a preference to a specific method. We will discuss advantages and drawbacks of the techniques instead.

The simulated annealing (SA) algorithm numerically reproduces the physical process of the quasi-static cooling of a high-temperature solid down to the ambient temperature [54]. When the cooling occurs very slowly with time, the solid is expected to reach the minimum energy state, escaping possible metastable states that can be reached during the process. By analogy, the energy of the system is here represented by the cost function S(**m**), the global minimum of which we want to search for. The algorithm starts from an initial random state $\mathbf{m}_0$ and moves to other models through driven random walks until the temperature reaches its final value. At the k-th iteration, if the system is in the state $\mathbf{m}_k$, it moves to a new state $\mathbf{m}_{k+1}$, if $S(\mathbf{m}_{k+1}) < S(\mathbf{m}_k)$. If the latter condition is not satisfied, the algorithm accepts worse solutions with a probability

$$P(T) = \text{const}\, e^{-\frac{S(\mathbf{m}_{k+1}) - S(\mathbf{m}_k)}{T}} \tag{26}$$

which depends on the absolute temperature and the difference between the two cost functions. At the beginning of the schedule, when the temperature is high, the search looks like a random exploration, while at the end the algorithm moves only downwards, falling in the minimum closest to the last states. When the temperature decreases very slowly, the probability of escaping local minima is large but also the computational cost becomes expensive; hence there is a trade-off between the convergence rate and the probability to fall in a local minimum. The most used variant is the heat-bath SA [55], which is largely used in kinematic source inversion [6,56,57]. With this scheme, only one parameter, one component of **m**, at a time is randomly changed during a given iteration, while the others are fixed. The main drawback of SA is that the success of the algorithm depends on the choice of several input parameters: the initial temperature, the temperature decrease rate and the generator of neighbor models. These parameters generally demand an initial tuning and may be problem dependent, such that several explorations are required to find the optimal compromise between computational costs and global exploration.

The genetic algorithm (GA) emulates the evolution process in maintaining and combining the best models inside a given population, where the quality of the models is defined according to the cost function S(**m**) [41,43,58]. Starting from a random population, where models are codified through a genotype, the species tends to preserve good fitness models and

replace bad fitness individuals. In building a new population, children may inherit the same genotype of their parents or their genotype can be a recombination (crossover) of the genotypes of two parents, according to their fitness function. Such a procedure tends to homogeneize the population, simulating the convergence toward a small set of minima. Random mutation of genotypes, instead, allows to continue the global investigation of the parameter space and possibly to escape from local minima. As for the SA, GA requires the setting of the several parameters (size of the population, crossover and mutation probabilities, number of iterations), which is realized through an ad-hoc tuning, generally tackled in synthetic tests emulating the source-receiver geometry and the frequency content of data. Moreover, GA has recently raised a large criticism since it has not a theoretical base (it is not sampling a probability distribution). Hence the convergence is not insured and the exploration cannot be used for the computation of marginal probabilities in the error estimation.

Finally, the Neighborhood Algorithm (NA) geometrically explores the parameter space, through a tessellation based on Voronoi cells [59]. The initial set of points is randomly chosen, then the exploration is intensified in regions where the misfit function is smaller, resulting into a cell refinement around deep local minima in the parameter space. The algorithm maintains a global character at the cost of a slow convergence. Instead of converging to a single minimum, it points to several regions where the global minimum can be located. Such an algorithm, hence, is well suited for ill-posed problems, such as the kinematic inversion, where the solution may not be unique. The algorithm requires the setting of only two parameters, which control the sampling process (number of initial points and number of re-sampled cells). After individuating the regions where the misfit function is smaller than elsewhere, the algorithm can be accelerated by coupling the exploration with a local derivative-free search method, such as the downhill simplex [60] (for application to the kinematic source inversion see [61,62]).

A global sampler is based on the minimization of a fitness function $S(\mathbf{m})$, whose functional form may become critical in the reconstruction of the rupture process. There are not so many criteria which are used to compare data and synthetics in the kinematic inversion of the seismic source. The easiest way to compare data and synthetics is the Euclidean distance. For linear problems, the use of least-squares techniques and implicitly of the L^2 norm comes from the hypothesis of Gaussian uncertainties. The corresponding fitness function is the square L^2 norm of the difference between observations and predictions

$$S(\mathbf{m}) = \left\| \mathbf{d} - \mathbf{g}(\mathbf{m}) \right\|_{L^2}^2 = \sum_{i=1}^{n_d} d_i - g_i(\mathbf{m})^{\ 2} \tag{27}$$

In the above representation, n_d is the dimension of the data space. Such a choice has several advantages. First, least-squares methods lead to easier computation as compared to other techniques. Second, the Parseval theorem warrants that the L^2 norm in the time domain is the same as in the frequency domain (apart from a constant factor depending on the definition of the Fourier transform). Often this latter should be computationally preferred, since the dimension of a time-series is of the order of 10^3-10^4, while the dimension of a band-limited Fourier series is on average two orders of magnitude smaller. The main drawback of the L^2 norm is its lack of robustness, since the norm is sensitive to a small number of outliers. Additionally, the use of such a norm in the inversions mainly reproduces the data portions with large amplitudes, and penalizes the stations where the overall amplitude is smaller than elsewhere (e.g. anti-directive stations). The L^1 norm should limit the weight of errors in the data, increasing the contribution of low amplitude data [61,63]. Its expression is

$$S(\mathbf{m}) = \left\| \mathbf{d} - \mathbf{g}(\mathbf{m}) \right\|_{L^1} = \sum_{i=1}^{n_d} \left| d_i - g_i(\mathbf{m}) \right| \tag{28}$$

Its application in time or frequency domains may lead to different results because the two quantities are not directly linked. Some authors propose a linear combination of the two norms L^1+L^2, to take advantage form both norms [56,64].

Very often, arrival times of synthetic seismic phases are not aligned with the ones from real data, owing to uncertainties in the structure model and in the hypocenter location. Cross-correlation of data may help in eliminating the shift between real and synthetic waveforms. A misfit function based on the correlation coefficient has been proposed in the framework of 1D velocity models [65] and then widely applied to source modeling [6,66]. This function tends to represent the shape of the signal but it is also sensitive to the amplitudes, albeit with a reduced importance as compared to the L^2 norm. We write here its expression in the frequency domain:

$$S(\mathbf{m}) = \sum_{i=1}^{n_s} \left[1 - \frac{\sum_{j=n_{\omega_{\min}}}^{n_{\omega_{\max}}} d_{ij} g_{ij}^*(\mathbf{m}) + d_{ij}^* g_{ij}(\mathbf{m})}{\sum_{j=n_{\omega_{\min}}}^{n_{\omega_{\max}}} d_{ij} g_{ij}^*(\mathbf{m}) + \sum_{j=n_{\omega_{\min}}}^{n_{\omega_{\max}}} d_{ij}^* g_{ij}(\mathbf{m})} \right] \tag{29}$$

In the above formula, the summation of the contribution of the single frequencies was separated from the contribution of the stations (n_s is the total number of stations) [6]. Hence the indices i and j in the data and synthetics are related to the i-th station and the j-th frequency respectively.

3.3. Errors and resolution

Data and theory are not exact and the associated uncertainties affect the model parameters (slip amplitude, rupture velocity and rise times). The propagation of the uncertainties from data and modeling on the source parameters leads to the definition of the errors associated with the retrieved model. If the error bars for a specific parameter are significantly smaller than our prior knowledge, we also say that such a parameter is resolved by the inversion [48]. Hence resolution is a slight different concept as compared to the standard errors associated with a given estimation. Specifically to the kinematic inversions, a feature in the slip model is *resolved* if it is really required to justify the data. Hence, to interpret a slip solution coming from the inversion of earthquake data, we have to know which slip patches, which variations in the rupture velocity and rise time are resolved on the fault plane and which is the error that affects such estimations.

The evaluation of the errors and the assessment of the resolution are not an easy task in non-linear inverse problems and in the case of the slip inversion, it is made harder by the difficulties in assessing the error on the Green's functions due to the uncertainties in the velocity structure. In the case of a linearized problem, however, a closed expression can be written for both the error and the resolution, if we assume Gaussian uncertainties in the data and in the theory. Let us indicate with $\mathbf{C}_\mathrm{d}$ and $\mathbf{C}_\mathrm{T}$ the covariance matrices associated with data and theory, respectively. Then, the generalized solution of the linear problem, in the sense of the weighted least-squares, is $\Delta\mathbf{m} = (\mathbf{G}^T\mathbf{C}_D^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{C}_D^{-1}(\mathbf{d}^{obs} - \mathbf{g}(\mathbf{m}_0))$, where the matrix $\mathbf{C}_\mathrm{D} = \mathbf{C}_\mathrm{d} + \mathbf{C}_\mathrm{T}$ [48]. Such an expression, which does not account for an a-priori statistical information on the parameters, corrects formula (24) and reduces to it when $\mathbf{C}_\mathrm{D}$ is diagonal with constant elements. The operator

$$\mathbf{C}_m = (\mathbf{G}^T\mathbf{C}_D^{-1}\mathbf{G})^{-1} \tag{30}$$

is the covariance matrix in the parameter space. The diagonal components of this matrix represent the variance associated with the parameters and its square root is the standard error that can be ascribed to the single estimations. Off-diagonal elements cannot be directly inspected. The normalized values

$$\rho_{ij} = \frac{(C_m)_{ij}}{\sqrt{(C_m)_{ii}(C_m)_{jj}}} \quad i \neq j \tag{31}$$

represent instead the correlation between couples of parameters. ρ_{ij} ranges between -1 and 1, where large absolute values are associated to highly correlated or anti-correlated parameters, depending on the sign of ρ_{ij}, while values close to zero indicate almost independent parameters.

For a given model $\mathbf{m}^*$, let us consider now the forward theory producing synthetic data $\mathbf{d}^*=\mathbf{G}'\mathbf{m}^*$. Then, let us invert these data to find a solution $\mathbf{m}$. The operator $\mathbf{R}$, which relates the original model to the inverted one ($\mathbf{m}=\mathbf{R}\mathbf{m}^*$) is referred to as the resolution operator. In the case of least-squares with no a-priori information in the parameter space the resolution operator can be written as $\mathbf{R} = (\mathbf{G}^T\mathbf{C}_D^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{C}_D^{-1}\mathbf{G}'$. If the same operator is used for the forward and inverse problem ($\mathbf{G}'=\mathbf{G}$) and $\mathbf{C}_m$ admits an inverse, $\mathbf{R}$ reduces to the identity. However, for the regularized linear problem (25), this condition does not hold because the smoothing is artificially added to the inverse problem.

When the problem has been tacked with a non-linear approach, the solution can be always developed around the minimum to find out the local covariance operator. If the cost function is represented by the weighted L^2 norm $S(\mathbf{m}) = \mathbf{d}-\mathbf{g}(\mathbf{m})^T C_D^{-1} \mathbf{d}-\mathbf{g}(\mathbf{m})$, the Hessian of the function S at the minimum, i.e. local the curvature, approximates the inverse of the covariance operator in the parameter space:

$$\left(\frac{\partial^2 S}{\partial m_i \partial m_j}\right)_{\mathbf{m}=\mathbf{m}^*} \approx \mathbf{G}^T\mathbf{C}_D^{-1}\mathbf{G} \tag{32}$$

The estimation of the curvature can be also performed by discretely evaluating the partial derivatives of the function S with a finite-differences approach.

The computation of errors and resolution requires finally the definition of the rows of the matrix $\mathbf{C}_D$. Errors in the data are related to the *noise*, which represents in a generic sense the contribution to the record that is not modeled by the theory. The main sources of noise are the ambient, which usually produces a very low amplitude disturbance as compared to the strong motion signal, the instrumental baseline and the data processing. When using displacement data coming from a double integration of accelerograms, baselines from instrumental tilting can lead to a noise amplitude at low frequency as large as the signal. When removing the baseline or appropriately

high-pass filtering the data, the noise can be as small as 1% or less as compared to the signal in the whole frequency band of investigation. On the other hand, the error associated with the theory is more difficult to be estimated. It is mainly related to the uncertainties in the Green's functions, to the chosen analytical form of the STF, to the smoothing used to stabilize the inversion and to the parametrization. In the framework of the comparison between 1D and 3D velocity models in the retrieval of large wavelength source properties, preliminary synthetic tests were performed to assess the effect of the velocity model on the data as well as on the final slip model [4]. Although the overall comparison between waveforms could be good as well as the shape and amplitude of first arrivals, the amplitude of later reflected /converted phases may be quite different, indicating a degradation in waveform fitting as the time from the first arrival increases. To check the quality of the Green's functions, a good test is to compare the synthetic waveforms with the records from the aftershocks, that can be assimilated to EGFs. Such a technique was initially applied for comparing several velocity models available for the Landers area [36]. In such a case the authors found an average variance reduction from the best model around 25% even at very low frequencies (f<0.25 Hz), indicating that the choice of specific velocity models is very critical for the rupture investigation. The comparison between synthetic Green's functions and aftershock records can be further exploited to find the optimal 1D model that minimizes the misfit between the waveforms [67]. To minimize the error in the estimation of the Green functions, some authors propose the use of station dependent 1D velocity models [68]. However, if these models are different also at depth, we should be aware that they do not insure the same seismic moment at all the stations. Another possibility to investigate the weight of the uncertainties in both theory and data is to use the variance reduction associated with the best solution of the inverse kinematic problem [41].

When dealing with a non-linear problem and possibly non Gaussian statistics, one could use the exploration as a sampling of the parameter space. If the exploration is based on random Monte-Carlo techniques, such as for simulated annealing and neighborhood algorithms, a probability density function can be defined from the misfit as $\mathrm{P}(\mathbf{m})=\mathrm{const}\, e^{-S(\mathbf{m})}$. In such a case, the normalization constant can be computed by constraining the integral over the whole parameter space to be 1. After normalization, the mean value and the covariance matrix can be computed as:

$$\begin{aligned} \langle \mathbf{m} \rangle &= \int_M \mathbf{m} P(\mathbf{m}) d\mathbf{m} \\ (\mathbf{C}_m)_{ij} &= \int_M m_i m_j P(\mathbf{m}) d\mathbf{m} - < m >_i < m >_j \end{aligned} \tag{33}$$

The integrals can be then computed using the Monte Carlo method, providing that the sampler stores the values of the cost function in all the explored points of the parameter space.

Instead of searching for the unique solution of the inverse problem, the sampling algorithms may also provide several solutions which have almost the same misfit. These solutions may be found in the same final population or as the result of several explorations started from different initial models/populations. As a better kinematic representation of the source, some authors [69] propose to perform a statistical analysis on the final models and select the mean model as the best solution, the standard deviations as the errors associated with the parameters. The statistical analysis can be performed through all the models investigated during a single exploration, with the solution chosen as the weighted mean of the explored models and the weights inversely proportional to the cost function [57]. However, we should be aware that the misfit of the average solution can be significantly larger than the one associated with the minimum of the objective function.

In most cases the model resolution is investigated through synthetic tests, with the same geometrical configuration of the fault and source-receiver distribution [43,56,70]. The goal of the test is to verify if a retrieved slip patch, an anomalous value of the rupture velocity etc., resulting from the inversion, are a method artifact or are justified by the data. The test is built by computing synthetic data with an assigned rupture model. After the addition of white or correlated noise in a small percentage (5-10%), synthetics are inverted to retrieve the rupture history, which is directly compared with the original one. As a result, we are able to understand if the presence, the shape and the amplitude of the anomaly is really required by the data or if it was caused by a smearing effect due to the acquisition layout.

3.4. Back-projection, beamforming, time-reversal techniques

Broad-band records from large earthquakes are a complex combination of body and surface waves, near field and static contributions, the correct modeling of which requires an accurate Green's functions computation. In several cases, however, inspection of waveforms may allow to simplify the estimation of the propagation contribution, either because the single phases are well separated in time, as for teleseismic or regional records, or some phases are dominant along the seismogram, as the direct S waves are in the far field condition for local records [71]. The arrival time and the amplitude of a selected body wave can be modeled in the ray-theory asymptotic approximation and time-reversal techniques can be applied to back-project

such an information on the fault plane to retrieve the space-time history of the seismic rupture.

When seismic waves are recorded far from the source by dense networks, array beamforming methods can be used to measure the back-azimuth and the velocity of the incoming waves [72]. Such a technique can be efficiently applied to the P-wave train, which can be clearly identified on the seismograms and separated from later phases. To further reduce the contribution of later arrivals, direct P waves can be aligned by waveform cross-correlation and then stacked. Relative amplitudes are hence back-propagated on the fault to infer the slip as a function of space and time. Such methods allow for the estimation of the large scale source parameter variations (rupture duration, rupture velocity, seismic moment) and may individuate the regions on the fault plane that emitted the largest amount of radiation. Back-projection techniques can be applied when the high frequency contribution radiated by the source is still recorded at distances larger than the fault dimensions. Such a condition holds for large magnitude events recorded at teleseismic distances [72,73] and moderate magnitude events at regional distances [74]. In Figure 7 (right panel) we show the back-projected kinematic model for the L'Aquila earthquake. It is obtained by a beamforming technique using the data recorded by ISNet, the Irpinia Seismic Network, installed in Southern Italy. Back-projection of the P wave signals (Figure 7, left panel) into the source region shows in this case a southward directivity and an average rupture velocity of 3 km/s.

At local scales, the direct S waves dominate the far–field records when the source-receiver distance is larger than few (3 to 5) wavelengths. Such a condition can be hence applied to records from stations located at distances larger than a dozen kilometers from the fault and for frequencies higher than 0.1 Hz. For a single body-wave phase, the far-field representation theorem can be further simplified. Assuming zero rise time (the slip instantaneously reaches its final value), the displacement at a given receiver located at **x** can be obtained as:

$$u_i(\mathbf{x},t) = \mu \int_{L(t)} K^c \frac{R_{ij}^c(\mathbf{x},\vec{\xi})}{r_p(\mathbf{x},\vec{\xi})} \delta u_j(\vec{\xi}) C(v_r, c, \phi) dl(\vec{\xi}) \tag{34}$$

The quantity $K^c = \dfrac{F^c}{4\pi\sqrt{\rho(\vec{\xi})\rho(\mathbf{x})c(\vec{\xi})c^5(\mathbf{x})}}$ accounts for the source and receiver impedances and for the amplitude changes that a wave undergoes during its path because of the free surface and the interfaces (ρ is the density,

c the wave velocity, F^c the total interface amplitude factor). R_{ij}^c is the radiation pattern, r_p the geometrical spreading and C is the Doppler coefficient at the source

$$C = \frac{v_r c}{c - v_r \cos\phi} \tag{35}$$

being ϕ the take-off angle of the ray connecting the point on the fault to the receiver. Such a relationship holds for any body wave phase (P and S) with propagation velocity c. Finally L is a curve on the fault plane satisfying the condition that the sum of rupture and propagation times is constant:

$$T_R(\vec{\xi}) + T_c(\mathbf{x}, \vec{\xi}) = \text{const} \tag{36}$$

Such a curve is referred to as the isochrone [75,76]. The information coming from points lying along the same isochrone simultaneously reaches the receiver located at the surface where it may constructively or destructively interfere. In a homogeneous medium such curves are forth-order ellipses, whose eccentricity depends on the ratio between the rupture and the wave velocities and on the relative location of the receiver with respect to the fault. For directive stations and S waves, the isochrones are more elongated than for lateral stations and for P waves [77].

The main idea of the back-projection is to distribute the amplitude recorded at a given time and at a given receiver as slip along the corresponding isochrone on the fault plane [43]. When no a-priori information is available about the location of high-slip patches, the slip is assumed to be uniformly partitioned along the isochrone. Hence, with only one recording station, one is not able to distinguish which points along the same isochrone provided the effective dislocation. When stacking amplitudes on isochrones from several stations, which sample the fault plane from different angles, the regions which provided effective slip can emerge from the background. However, because of the limited number and the uneven distribution of the stations, the final slip image still remains defocused around the main slip patches. A *restarting* procedure can be applied, according to which the back-projection is iterated and the n-th slip map is built assuming as an *a priori* information the slip map obtained at the previous iteration. The restarting can be arrested when the improvement in the objective function comparing data and synthetics becomes incremental or degrades. When applied to real data, such a procedure still helps in reducing the smearing of the slip, but it is generally efficient for a small number of iterations.

However, back-projection of displacement amplitudes requires an ad-hoc data processing because baselines strongly degrade the quality of the signal and the radiation pattern may lead to instabilities when it changes its sign along the same isochrone [43]. Several authors propose the use of the displacement or velocity envelopes [78,79] which are positive defined and less sensitive to the radiation pattern. The isochrones can be drawn on the fault plane as far as the rupture velocity is known. Before starting a kinematic inversion instead, the rupture velocity is generally unknown. To bypass the assumption of a defined rupture velocity map, we can include the back-projection method into a more general technique which searches for a rupture velocity map, in a non-linear approach, as described in the Section 3.2. As soon as a rupture velocity map is available, the corresponding slip map is obtained by back-projection. From the comparison of data and synthetics, an objective function can be computed for each couple rupture velocity-back projected slip, and the one which minimizes the misfit is chosen as the solution to the problem. In addition, some authors suggest to separate the discretization of the slip, whose space scale is related to the maximum frequency investigated in the data from the discretization of the rupture velocity [43]. For this latter, they propose to start from a constant value and then add variations at smaller and smaller scales. Such a procedure can be arrested when a further reduction of the discretization grid does not contribute to a significant reduction of the objective function.

With the development of computation resources, an interesting back-projection technique based on time reversal imaging is going to be developed for the reconstruction of the rupture process. Time-reversal is widely used in acoustics and exploration seismology for source location and structure imaging by noise and/or signal cross-correlation. Since the wave equation is symmetric with respect to time, the wavefield recorded at the stations can be time-reverted to collapse again into the source. When seismograms are injected back into a numerical model of the Earth, they propagate and constructively interfere at the source to retrieve the space-time evolution of the rupture. Such a method has been applied to reconstruct the rupture history of the great Sumatra earthquake, using teleseimic data at very low frequency (T>200s) [80]. However, the time-reversal field does not exactly behave as if we rewound the movie of the direct wave propagation. After converging into the source, the wave energy does not disappear from the numerical model, returning to the rest condition that we had before the rupture started. On the contrary, after the focalization of the wavefield at the source, the wave energy it spreads out again preventing from a complete reconstruction of the source process.

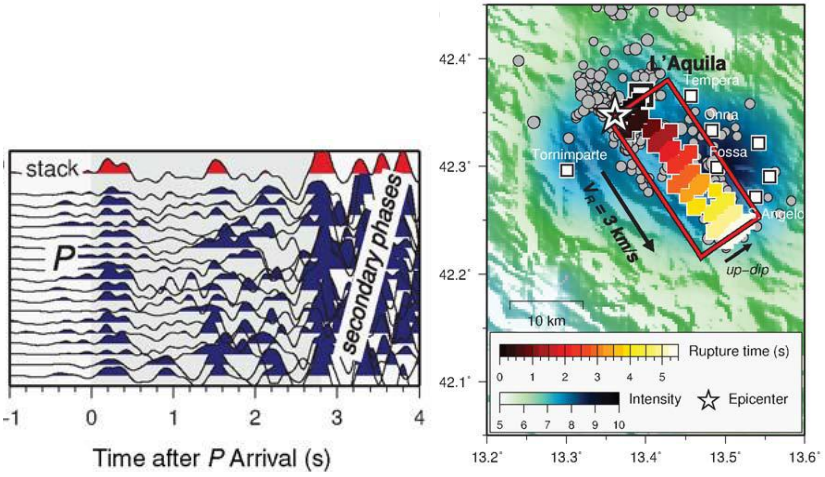


Figure 7. On the left, vertical-component acceleration records of the 2009 L'Aquila earthquake, recorded at the stations of ISNet (Irpinia Seismic Network), a network installed in Southern Italy at about 250 km of distance from the epicenter. The traces are aligned along the section by waveform cross correlation of the P wave first arrival. On the right, reconstruction of the rupture kinematic associated with L'Aquila earthquake, obtained by a back-propagation of the stack function. The figure is extracted from [74], Copyright © 2011 by Seismological Society of America.

4. Data

4.1. Strong motion data

Accelerometric data recorded in the fault vicinity are suitable for the retrieval of the kinematic properties of the rupture up to high frequencies (1-2 Hz). Attenuation relationships indicate that peak ground acceleration (PGA) does not significantly increase with magnitude, with major changes ascribed to the source-station geometry (rupture mechanism and directivity) and site effects [81]. It is worth to note that the largest values of PGA ever recorded by strong motion stations in the near-source are of the order of 3-4 g (2008, M 6.9, Iwate-Miyagi earthquake, station IWTH25, PGA=4.3g; 2011, M 6.3,Christchurch earthquake, station HVSC, PGA=1.7g; 2011, M 9.0, Tohoku earthquake, station MYG004, PGA = 3g). Hence, accelerometric data do not saturate also in the case of very large magnitude events (M>8.5), providing an effective picture of waves radiated by the earthquake. Moreover, accelerometers have a flat response from DC up to high frequencies,

indicating that we are able in principle to recover the correct ground motion in a broad frequency range, at least in the band where the signal to noise ratio is significantly larger than one. However, as previously discussed, the ability in describing the rupture kinematics strongly depends on the knowledge of the propagation model, which becomes more and more imperfect as the frequency of seismic waves increases. The upper frequency at which kinematic inversions are performed do not generally overpass 0.5 Hz, up to which 1D velocity models can be assumed as a reliable description of the propagation medium. Displacement data are hence more appropriate to represent the low frequency motion associated with the earthquake. Since the far-field spectrum of the displacement is flat up to the corner frequency and decreases beyond it, the kinematic inversions are expected to well reproduce the seismic moment, to locate the high-slip regions on the fault plane and to resolve the average properties of the rupture velocity.

However, a major limitation occurs when the displacements are obtained by double integration of acceleration records. Small steps, offsets and perturbations in the acceleration records are amplified by the integration and are transformed in long-period baselines which significantly distort the trace, providing unrealistic static displacements and incorrect waveforms. Several sources could produce baselines in the strong motion data, as the hysteresis of the sensor or the tilt of the ground associated with the arrival of the seismic waves at the site, which induces a rotation of the inner mass [82,83]. Since the physical mechanisms beyond the offsets are not known and the rotations are not recorded within the strong motion, a general cure for suppressing the baselines is not available and ad-hoc procedures have to be adopted to recover correct displacements at low frequencies. The simplest way to reduce the effect of the baselines is to high-pass filter the displacement data with cut-off frequency generally in the range 0.05-0.1 Hz. Such a procedure could not affect the correct estimation of the seismic moment and its distribution on the fault plane for earthquakes with magnitude $M<7.5$, because the corner frequency is larger than the cutoff frequency due to the filter. It could provide an incorrect image of the slip distribution for large magnitude earthquakes ($M>8$) instead. Several procedures have been suggested for the direct removal of the baselines, starting from the arrival time of the strong motion. [84] suggested to fit the acceleration baseline with polynomials, [85] removed the baseline from the velocity record by fitting data portions with straight lines, [86] subtracted to the acceleration the derivative of a quadratic form that fits the velocity record. It is worth to note here that all of the procedures include some subjective choices (the order of the polynomial functions, the selection of data portions for the fit, etc.), generally verified *a-posteriori* by inspection of the resulting displacement. At the end of the

procedure, hence, the final displacement can look realistic but it is not necessary correct. As an example, we show in Figure 8 (from [86]) the displacement obtained by double integration of an acceleration record from the 1999 Chi-Chi, Taiwan, earthquake, where the effect of the baseline dominates the signal at very low-frequency. On the same Figure, the signals obtained by applying different baseline correction schemes are superimposed. When removing straight lines from velocity data portion, the final displacement becomes strongly sensitive to the end point of the straight line. It is hence a better choice to constrain, when possible, the static final displacement with GPS or INSAR measurements. As a last observation, the displacement records are a low-frequency representation of the waves emitted by the rupture. Hence, the low frequency hypocenter (centroid) could not coincide with the high-frequency origin point of the earthquake. This difference, summed to the

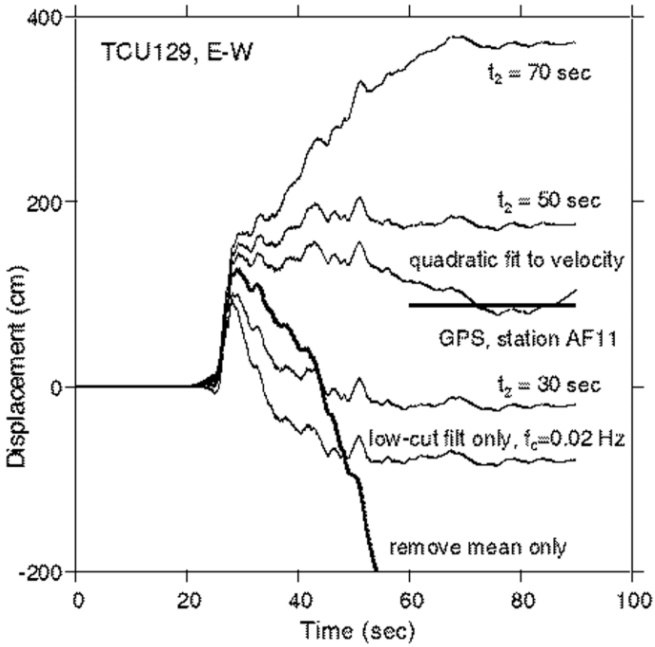


Figure 8. Displacements obtained by double integration of the acceleration record of the EW component at TCU129 from the 1999 Chi-Chi, Taiwan, earthquake. The solid black line is the result of the integration after removing the pre-event mean level. The different traces are obtained by applying several baseline corrections. The figure is extracted from [86].

incorrect velocity structure, may affect the comparison of data with synthetics generated for a rupture propagating from the high-frequency hypocenter. Hence, it is better to use cross-correlation measurements instead of direct norms to evaluate the misfit or to time-shift the real data before performing the comparison.

4.2. Teleseismic data

The Earth behaves as a natural low-pass filter and the high frequencies generated by a seismic rupture are strongly attenuated at regional and teleseismic distances. For this reason, source modelers generally investigate teleseismic data at very low frequencies (<0.1 Hz). Depending on the earthquake magnitude and the recording distance, the useful spectrum of the seismic signal ranges between 1 mHz and 0.5 Hz. The choice of the frequency band is based on the signal to noise ratio. Moreover, when analyzing displacement records, the spectrum should contain both the flat part, associated with the seismic moment, and the corner frequency, beyond which the phases generated by the different regions of the fault interfere. Since at teleseismic distances the seismic phases may appear separated on the records, it is convenient to extract the most energetic contributions by windowing the trace around the selected phases and separately interpret each of them. This is the case of direct P and SH waves, within their first reflections at the free surface in the epicentral distance range between 30° and 90°. At such distances, direct waves mostly travel in the deeper part of the mantle, which is almost homogeneous [87]. At smaller angles the heterogeneous structure of the crust and the upper mantle and the lateral variations may strongly affect the ray path and hence the shape and the amplitude of the direct phases [20]. Also the surface waves can be used for the characterization of the rupture kinematics. Because their speed is close to the rupture velocity, they are expected to constrain the average speed of the rupture [88]. However, synthetic Green's functions associated with surface waves are sensitive to the crust and upper mantle structure and their lateral variation. Hence, at periods smaller than 500-800 s, the Green's functions need to be computed in a 3D realistic Earth model, which should contain both the crust and topography. At larger periods, the Green's functions can be determined by normal mode summation. Synthetic Green's functions can also be substituted by EGFs, which can be deconvolved from the mainshock (phase by phase) to retrieve the relative source time functions [89].

Because data are investigated in the low frequency range and the amplitude spectrum of the displacement can be easily related to the

macroscopic properties of the seismic rupture, it is preferable to invert displacement records. These latter are obtained from velocity seismograms, by extraction of the single phases and integration [89].

4.3. GPS data

The Global Positioning System (GPS) is a constellation of satellites that enables the receivers on the Earth surface to position themselves. With the deployment of a large number GPS sites, allowing for an accurate determination of the satellite orbits in a well defined terrestrial reference frame [90], the employment of double frequency receivers and the use of advanced post-processing software, GPS measurements have reached the precision of millimeters to centimeters, becoming suitable for the imaging of the earthquake source. Most of GPS data provide static displacements associated with the coseismic slip occurred during an earthquake. Static displacements decay with the distance faster than seismic waves (as the inverse of the square source-receiver distance). Hence static displacements recorded very close to the fault have a different resolution and enlighten smaller regions of the fault plane as compared to the dynamic waves. Moreover, the difficulty in yielding the static displacement from strong motion owing to the drifts (see section 4.1), makes GPS data complementary to accelerometric data. To accurately determine the displacement offset associated with the earthquake, time series are corrected by removing the average position of the sensor before and after the seismic event.

In the last decade a large number of high rate continuous GPS receivers (1 Hz, 5Hz, 10 Hz) have been installed, allowing to record not only the static field associated with an earthquake, but also the low frequency spectrum produced by seismic waves. To obtain reliable data, it is important to process the GPS time series epoch by epoch, correcting its position with respect to a network of stations [91]. In such a case, the measurements are strongly sensitive to the geometry and the number of the satellites in view and the accuracy can be different from one component to another one [92]. If for a moderate magnitude earthquake it is difficult to go beyond few oscillations, such instruments may provide new insights into the moment release of large to mega-thrust earthquakes. Several successful applications have shown the potential of the continuous GPS data in reproducing the ground motion and retrieving the low-frequency time-space evolution of the slip associated with seismic ruptures. As an example we report the cases of Denali [92], San Simeon [93], Tokachi-oki [94] and Chile [95] earthquakes.

4.4. Other observations

SAR data

Differential Interferometry on SAR (DinSAR) is the best satellite remote sensing approach to investigate the surface displacement caused by a moderate to large earthquake. The main idea of the DinSAR is the computation of the phase difference between two images, the former acquired before the earthquake and the latter after its occurrence. The differential image is an interferogram formed by a series of fringes which can be more or less coherently tracked on the ground. The fringes can be unwrapped to obtain the surface static displacement in the direction of the satellite. In the most favorable cases, the resolution can be as low as few centimeters (and even smaller than 1 cm for the X band) when the degree of the phase correlation between the two processed scenes is high. Decorrelation may occur because of large baselines between the two passages, important changes at the surface and atmospheric disturbance. Its application remains hence confined to inland earthquakes or offshore events close to the coast. Since its first application to the 1992, M 7.1, Landers earthquake [96], remote sensing by differential interferometry has been widely used to track the surface displacement associated with major earthquakes also in regions which are not well monitored by GPS and strong motion sensors. The main drawback of SAR is related to the delay of the passage of the satellite since the occurrence of the earthquake. The seismic moment retrieved by the inversion of SAR data could be therefore higher than the seismological moment because the afterslip associated to the post-seismic reassessment could have summed up to the coseismic deformation.

Surface data

Sometimes the rupture reaches the free surface and the local slip can be efficiently measured on the field through a geological survey. To reduce the effect of the postseismic deformation, however, field measurements should be performed soon after the earthquake occurrence. In the characterization of the source kinematics we should take particular care in the use of such data. *De facto*, surface observations may not be representative of the large-wavelength shallow slip, because the field observations could be influenced by the geological structure nearby the surface. However its order of magnitude could be used as a constraint for the slip in the shallower regions of the fault.

Tsunami data

Offshore earthquakes rupturing close to the Earth surface cause a deformation of the seafloor. Static vertical displacements induce a motion of the water column due to the gravity which triggers long-wavelength gravity waves propagating away from the source as a tsunami. The rupture propagation occurs at a time scale (seconds to hundreds of seconds) which is significantly faster than the motion of water waves. The deformation caused by an earthquake at the bottom of the ocean can be assumed to occur instantaneously and then used as an initial condition for the tsunami propagation. Tsunami waves are then recorded by buoys located offshore or close to the coast and the water wave amplitudes can be inverted to retrieve the macroscopic features of the rupture which generated them. Such data could result very precious in constraining the extension and the geometry of the rupture and the amount of the slip at long wavelengths associated with mega-thrust earthquakes, such as for the 2004, M 9.0, Sumatra [97,98] and the 2010, M 8.8, Chile earthquake [99].

5. Examples

5.1. Tottori earthquake

The 2000 October 6, Western Tottori earthquake (Mw 6.8) originated at 04:30:18.07 UTC, 35.27N, 133.35E, and propagated through the digital networks K-net and Kik-net, with 20 stations located at less than 50 km from the epicenter. The depth of the hypocenter was estimated to range between 11 and 15 km [100,101,102]. According to the aftershock relocation, the rupture propagated bilaterally, along an almost vertical plane at about 145N with a bending to 165N in the last kilometers northwards [101]. The moment tensor (CMT) indicates a pure strike-slip solution. To skip the geometrical complexity of the rupture, the fault was simplified to a rectangle with length 30 km, width 20 km, and hypocenter location at 13.5 km of depth and angular parameters defined as follows: strike = 150°, dip = 90° and rake = 0°. Since there was no evidence of surface slip observed in the main shock area, the top of the fault was set to 1 km [103]. A simplified 1-D model was adopted for wave propagation [100]. The specific parameters for all of the layers are detailed in Table 1. Some other 1-D models that differ from this one mainly in the shallower layering, do not have a significant influence in the computation of the Green's functions. [101] did not observe relevant variations for the aftershock relocation, when testing different 1-D propagation models. This allows for a robust computation of the traveltime

between the stations and the subfaults. The final source-station geometry with the accelerometric data is sketched in Figure 9.

In this example, the inversion of seismic data was performed with a back-projection technique of the direct S waves. A non-linear exploration was carried out to the search for the best representation of the rupture velocity by the use of the genetic algorithm [43].

In the selection of the stations, the closest ones having a distance from the fault less than 10 km, were discarded to guarantee the far-field condition. The set of records used in the inversion consisted of 12 stations, located in the epicentral distance range of 20–50 km. As required by the backprojection, the displacement records were directly obtained by double integration from accelerations, and the low-frequency trend was removed through a baseline correction, using a fourth-order polynomial interpolation on the noise before and after the signal [86,104,105].

This procedure allowed for the retrieval of the displacements with proper sign. In the comparison between synthetics and data, we added a band-pass filter. The lower frequency was chosen to be 0.1 Hz, in order to have at least two to three wavelengths between any subfault and any station (far-field condition). The maximum frequency was set to 0.5 Hz as the limit at which the S signal remained coherent with time. Beyond that frequency, the pollution of additional reflected and diffracted phases from the 3-D propagation becomes significant. To infer the upper threshold, we looked at the stability of the motion on the S train in the data by polarigrams. The selection of the S wave in the seismograms was done by limiting the record

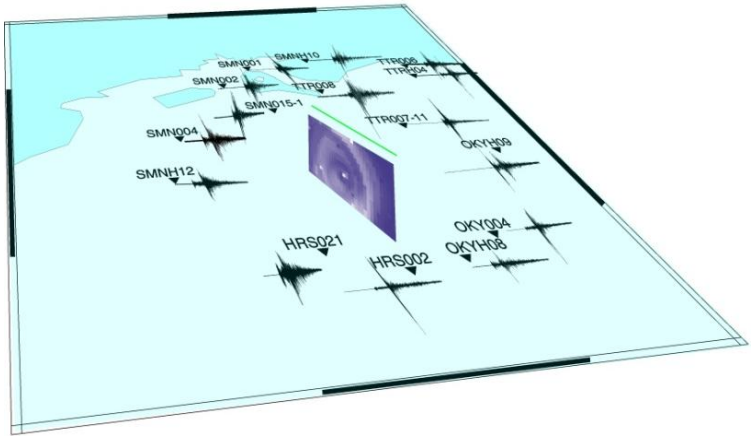


Figure 9. Station location and acceleration data for the Tottori earthquake.

window by the theoretical first arrival S time, as obtained from the 1-D propagation model. This time was checked against the direct picking of the S wave, when possible. In Figure 10 we compare polarigrams for the data and the synthetics, the latter obtained with a constant slip and a constant rupture velocity of 2.4 km/s. We show some examples for the stations HRS021, TTRH04 and SMN001, which are located southwards, laterally and northwards with respect to the fault plane. For all the synthetics, the orientation does not change as the rupture moves away from the hypocenter, at least in the first part of the record. Hence, the S-wave train is expected on average to follow an almost constant direction with time for any station. The

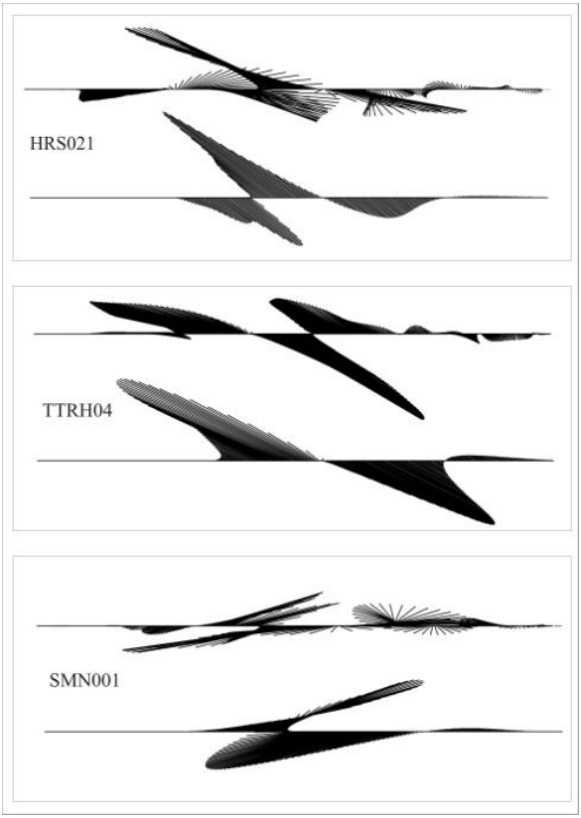


Figure 10. Comparison between the polarigrams from data (upper trace) and synthetics (lower trace) for three selected stations located northwards, laterally and southwards with respect to the fault.

lateral station TTRH04 shows a good agreement with the expected direction. At the station HRS021, the direction of the motion on the horizontal component fits the theoretical one everywhere, except at the beginning of the signal, where the amplitudes may correspond to a foregoing phase. For the inversion, the fault was discretized into squares of size 1 km, providing 30×20 subfaults. The rupture velocity was discretized using nine control points that are equally spaced along the strike and the dip directions. A bilinear Lagrange interpolation of the rupture velocity was performed for the points inside the control nodes. We used a weighted L^1 norm as cost function, with the weights associated to the maximum amplitude measured on each component of the displacement records. For any station, the two horizontal components were considered as independent traces to be backprojected. The slip map obtained from the inversion (Figure 12) has been compared directly with the solution provided by [100]. Two different patches of slip are retrieved. The major one starts above the hypocenter, close to the surface and goes down southwards until to reach the bottom of the fault. The second one is located northwards at a depth between 10 and 18 km. The reliability of the images can be analyzed with the help of the spike tests, associated to the final solution (Figure 13). As far as we look at the defocusing of the discrete pixels in the major patch, the backprojected images are mainly correlated along the dip direction, downwards. Hence we can argue that the slip close to the surface is real. At depths larger than the hypocenter location, the lack of resolution does not allow to infer how deep is the anomaly. Moreover the focusing of energy at the bottom of the fault is probably an artifact, because the isochrones of both patches intersect in that region. If we raise up the bottom of the fault, such a focusing effect is largely reduced. On the other hand, the width of the other anomaly is larger and its location is harder to be defined. High slip concentration close to the surface is a common feature of all the slip maps obtained by inversion of strong motion data. However its non symmetric character is not shared by all the models, some of which do have a symmetric pattern upward the hypocenter [102,106]. The same asymmetry was instead obtained by other authors [57,100]. A southern predominance of the slip can be also confirmed by the aftershock location. The mean value of the rupture velocity is 2.2 km/s, with almost regular fronts (Figure 12). Finally, the comparison between the data and the synthetics (Figure 11) shows good agreement at the lateral stations (SMN003, SMN004, SMNH02, SMNH12, TTRH04 and TTR006). Larger amplitudes in the synthetics for the directive stations OKY004 and OKYH08 could be indicative of an artifact, that is, the extension of the slip downwards at the bottom of the fault. Some inconsistencies are also present for the northern

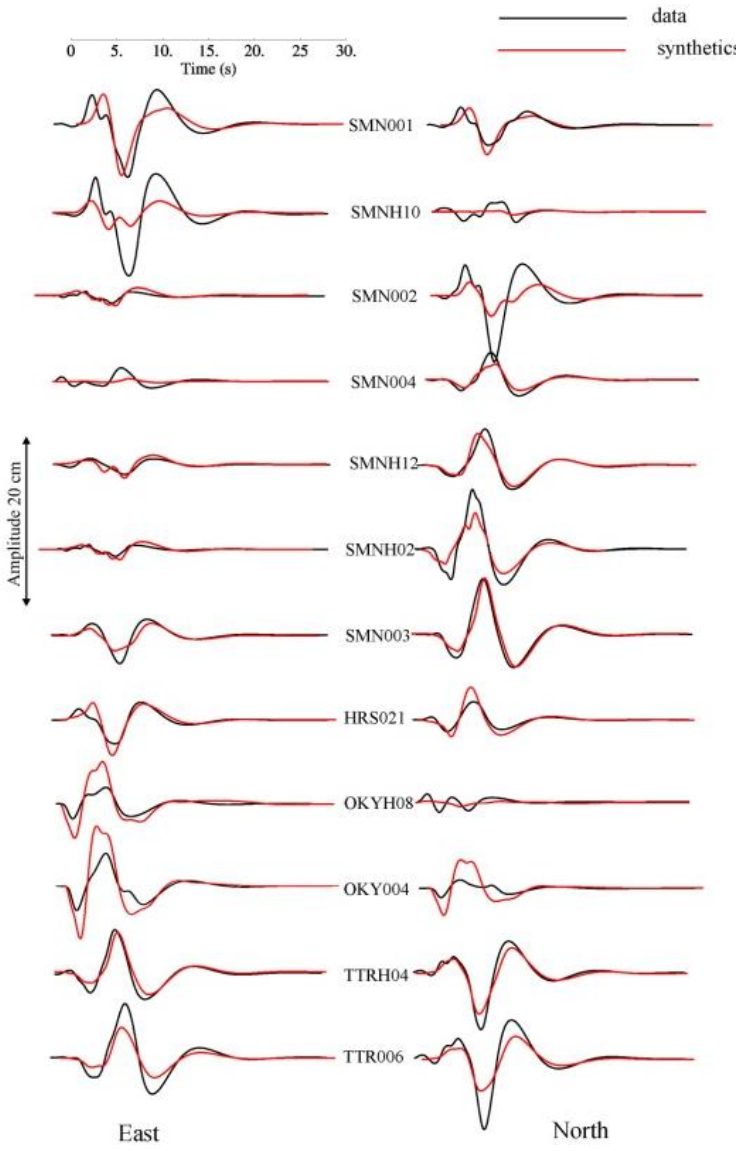


Figure 11. Comparison between the synthetics (red traces) and real data (black traces) for the horizontal components of 12 stations that recorded the 2000, Tottori earthquake.

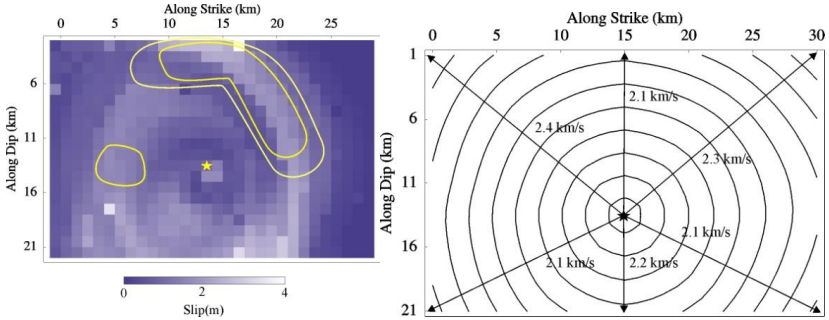


Figure 12. On the left slip map obtained by back- propagation of displacement amplitudes. With yellow contours, slip map obtained by [100]. Most of the slip is located southwards close to the surface. Lack of vertical resolution avoids to define the vertical extension of the slip anomaly. A second smaller anomaly is located deeper northwards of the hypocenter. On the right, rupture velocity values and rupture times contours. The rupture is slightly faster in the upper portion of the fault.

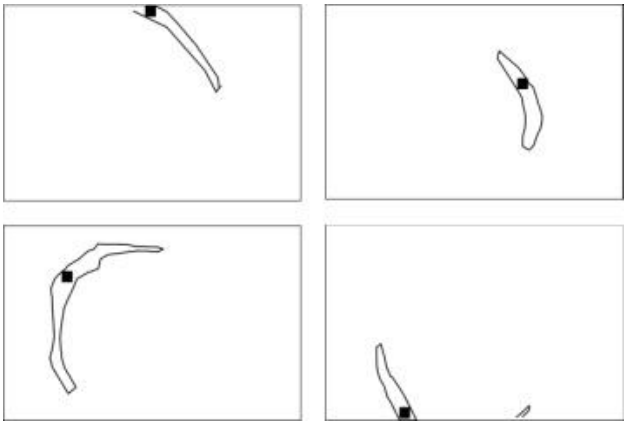


Figure 13. Resolution kernels for 4 grid points. Since the fault is vertical there is a lack of resolution from body waves in the dip direction. Hence slip images are blurred along the vertical direction and blurring depends on the smoothness weight as compared to the misfit in standard kinematic linearized inversions.

stations SMNH10 and SMN002, but not for SMN001, for which the displacement amplitudes are reasonably retrieved. Low-frequency amplification could occur at those stations, due to site effects associated with the marine sedimentary coverage [102].

5.2. Loma Prieta earthquake

The Loma Prieta earthquake (M 6.7) occurred at 00:04:15.2 UTC on 18 October 1989, at latitude 37.036N, longitude 121.883W, and 17 km depth [107], southwest of the southern Santa Cruz Mountains section of the San Andreas fault. Aftershocks occurred on both sides of the fault relative to the hypocenter for a distance of approximately 50 km, suggesting that the rupture propagated bilaterally along the strike and unilaterally up-dip [108]. Focal mechanism associated with this earthquake indicates that slip occurred on a fault approximately dipping 70°, striking 130°, with rake 140° [109]. Here the inversion was performed by a simultaneous search for the slip and the rupture velocity in a multiscale approach within a ray theory approximation [41]. A non linear exploration was performed by the use of the genetic algorithm.

In the modeling the fault extends from 15 km northwest to 20 km southeast of the hypocenter. The dipping fault plane intersects the hypocenter at a depth of 18 km. The modeled extension of the rupture zone in the up-dip direction from the hypocenter is 14 km, corresponding to depths from 18 km to slightly less than 5 km. The velocity model (Table 2) used to compute the Green's functions was determined analyzing the Loma Prieta aftershocks [107]. A constant quality factor $Q_S = 300$ [110,111] was used to parameterize the Azimi's attenuation function. The Loma Prieta earthquake occurred in a region densely instrumented with strong motion accelerometers. The locations of stations is shown in Figure 14. The instrument-corrected acceleration data were integrated to get the ground velocity and bandpass filtered with a four-pole zero-phase Butterworth filter with corner frequencies at 0.5 and 5.0 Hz. The low-frequency limit was constrained by the requirement that the receivers have to be a few wavelengths away from the nearest segment of the fault. Receivers located within about 40-km epicentral distance have been considered because our modeling is limited to the direct S waves. In this distance range complex propagation effects (dominant secondary arrivals, multipathing) are minimized [111,112,113]. A window delimiting the direct S arrival was selected from the horizontal components of the ground velocity by polarization analysis on both filtered velocity and displacement records.

The representation integral (equation 1) was numerically solved by discretizing the fault plane in 50x50 m^2 subfaults to avoid undesired spatial aliasing effects [114]. The rupture velocity and the final slip distributions on the fault were determined at this fine grid by bicubic spline interpolation of the values obtained by the data inversion at the control points. The initial run of the inversion procedure was performed starting from homogeneous models

Table 2. 1D velocity model used for the computation of the Green functions in the inversion of strong motion data for the Tottori earthquake.

Layer top (km)	c_p(km/s)	c_s(km/s)	ρ (g/ cm^3)
0	5.5	3.18	2.6
1.9	6.0	3.5	2.7
16	6.6	3.81	2.8
38	8.0	4.62	2.9

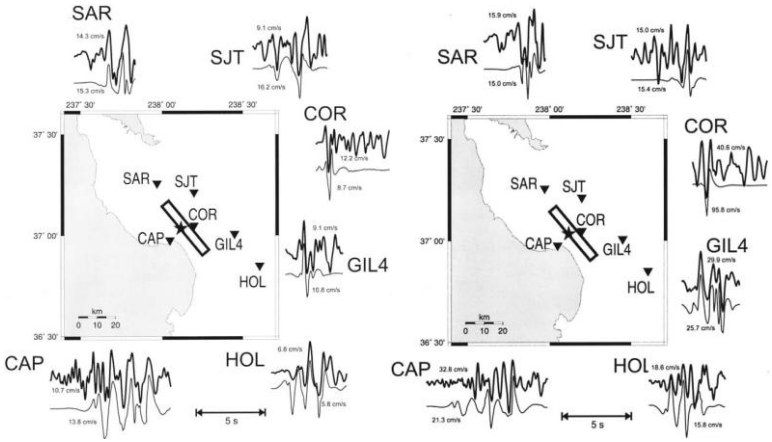


Figure 14. Source-station geometry for the Loma Prieta earthquake. Superimposed on the figure, the comparison between real data (solid black) and synthetics (solid gray) for the north-south (left panel) and east-west (right-panel) components.

Table 3. 1D velocity model used for the computation of the Green functions in the inversion of strong motion data for the Loma Prieta earthquake. slightly different shallow models are used for the two sides of the fault.

Layer top (km)	c_p(km/s)	c_s(km/s)	ρ (g/cm^3)
Northeast of the fault			
0	3.34	1.93	2.5
1.1	5.01	2.89	2.7
9.1	6.26	3.61	2.7
24.5	6.95	4.01	2.8
Southeast of the fault			
0	3.42	1.97	2.5
1.1	4.58	2.64	2.7
9.1	6.26	3.61	2.7
24.5	6.95	4.01	2.8

both for the rupture velocity (vr=2.0 km/s) and for the final slip (1.7 m, according to an average estimate of the scalar seismic moment) with five control points in the strike direction and three control points in the dip direction. As a consequence, with 5 x 3 control points we had 30 parameters to be determined through the data inversion. In this first run the slip was allowed to vary between 0 and 10 m, whereas the rupture velocity was allowed to vary between 2.3 km/s and 3.3 km/s. Individuals in the initial model population were selected randomly around the uniform starting models. In the following runs, the starting models were chosen around the model obtained in the previous search stage and the inversion was performed allowing for smaller and smaller variations around them as the number of control points increased. The preferred model parameterization according to the Akaike Information Criterium was 15 x 8. This grid corresponds to a spatial sampling of the fault of 2.5 km in the strike direction and 2.0 km in the dip direction. Top and bottom panels of Figure 15 show, respectively, the final slip and rupture-time distributions on the fault plane for this model. The mean value of the final slip on the fault for this grid is 1.4 m, which corresponds to a scalar seismic moment of 2.1 10^{19} N m. The distribution of the final slip on the fault is characterized by two main asperities located southeast and northwest of the hypocenter. Moreover, we found another high-slip region near the top middle-western edge of the fault which is probably driven by the high amplitude associated with the west–east component of the seismogram recorded at the COR receiver. The mean rupture velocity value for the best-fit model is 2.8 km/s, slightly higher than the mean value obtained by [108]. From the distribution of the rupture times on the fault it is evident that the rupture advanced faster toward the southeast than in the opposite direction. Finally, the total duration of the rupture is about 9 s.

In Figure 14 the ground-velocity records in the 0.5- to 5-Hz frequency band are compared with the synthetics computed for the best-fit rupture model at all the considered stations. The fit between synthetic and real seismograms is, in general, quite satisfactory although some discrepancies are found in the peak amplitude values. However, synthetic amplitudes are within a factor 1.5 of the observed ones. The north–south synthetics (Figure 14- left panel) reproduce well the observed complex shape and duration of ground motion at different azimuths and distances (see, for instance, stations SAR, CAP, and GIL4).

The overall fit of the west–east components of ground motion (Figure 14- right panel) is worse than for the north–south components, mainly because of the mismatch of stations SJT and COR, whereas for the other stations the synthetics match the spatial variation of the low frequency shape

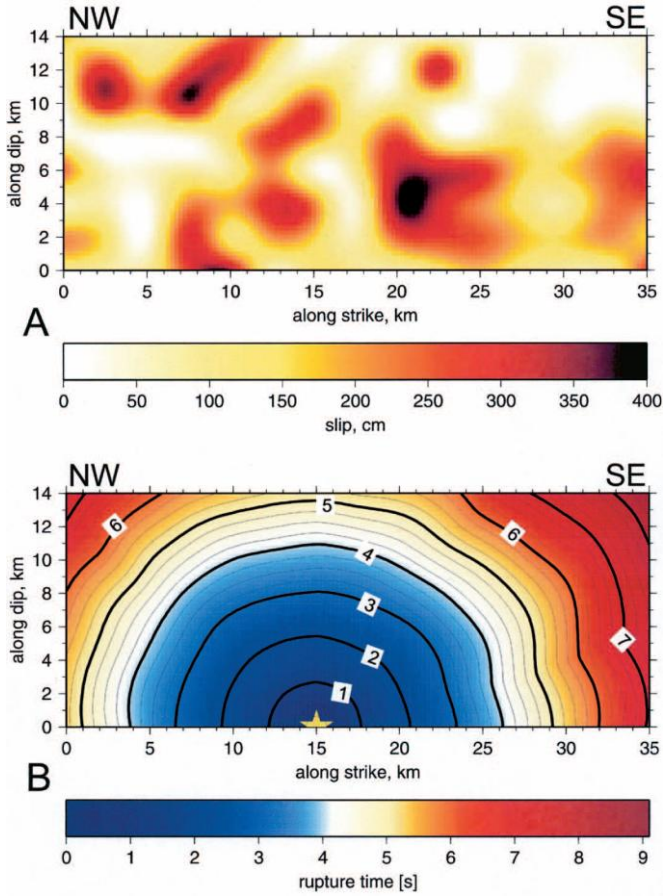


Figure 15. Top Panel: slip map as retrieved by the inversion of strong motion data for the Loma Prieta earthquake. Bottom Panel: Rupture times for the same earthquake.

and duration of observed records fairly well. In particular, the remarkable fit of the west–east component of station GIL4 must be pointed out.

6. Conclusions and perspectives

Seismic rupture kinematic models describe the space-time evolution of the relative motion of the two sides of the fault (the slip function) from the inversion of the seismic data recorded at the Earth surface. Although they

represent a simplified picture of the seismic rupture which does not account neither for the stress state on the fault, nor for the energy balance at the rupture front, they provide insights into the understanding of the physical processes governing the generation, the propagation and the short wavelength radiation associated to the earthquake rupture. Kinematic source models are obtained as the solution of an inverse problem, whose associated operator is the representation integral (eq. 1-8). The inversion of the seismic observations to retrieve the characteristics of the slip function on the fault surface is known to be an ill-posed inverse problem, with non-unique solutions and, in many cases, a not continuous dependence on the data. As a consequence, when analyzing the data from the same earthquake, rather inconsistent results arise from the kinematic inversions performed by the different groups, due to specific choices in the data selection and processing, in the model parametrization, in the definition of the objective function, in the inversion procedure itself and in the computation of the Green's functions. As an example, the several slip maps obtained for the Landers earthquake [30,36,115,116,117] show important variations in the position, shape and amplitude of the slip patches and differences in the average rupture velocity as large as 20% (results from inspection of the database of finite source rupture models SRCMOD by P.M. Mai, http://www.seismo.ethz.ch/static/srcmod/Homepage.html). Moreover, the earthquake source blind test performed during the EU-SPICE project (http://www.spice-rtn.org) showed that even when inverting synthetic noise-free strong motion data in a known velocity structure, the retrieved slip models do not completely match the true model, although that the main low-frequency features in the data appear well described by the different models. For this case, when cross-correlating the inverted models with the original one, some of the models have a cross-correlation value which is not better than a random but correlated slip model [118]. The results from the blind test hence indicated that subjective choices during the inversion procedure strongly affect the solution of the inverse problem, calling for a deeper understanding of the link between dynamic and kinematic parameters which would help in better constraining the inversion, for a proper data selection and processing and for a refined estimation of the uncertainties and the resolution on the final model. A further investigation of the robustness and limitations of the kinematic inversion could be performed by a suite of inversions of synthetic waveforms obtained as output of dynamic rupture simulations, as proposed by the SIV project (http://eqsource.webfactional. com/wiki/).

Uncertainties in the Green's functions computation can be reduced by the massive use of EGF records. The development of specific interpolation strategies and the use of hybrid methods to couple the low-frequency

numerical simulations with the high-frequency EGFs [119] may help in overcoming the present limitations associated to the use of the EGFs, mainly related to the non-uniform sampling of the fault plane by the aftershocks and to availability of an adequate number of stations for which both mainshock and aftershock are recorded with a large signal-to-noise ratio in wide frequency band. Another complementary and quite robust approach in case of moderate size earthquakes, is the use of the source time functions as input waveforms for the inversion of source kinematic parameters. The source time functions are obtained by deconvolution of the mainshock record by an EGF [e.g.,120,121].

The resolution of kinematic models can be improved by a joint inversion of seismic (strong motion, regional/teleseismic recordings), geodetic (GPS, InSAR) and eventually tsunami data, with the limitation that the different observed quantities have a different wavelength and frequency resolution on the physical processes under study. This promising approach has been widely applied for the kinematic inversion of moderate and large earthquakes [e.g., 89,116,122]. However, the relative weight of the different datasets is still somewhat arbitrary and it can strongly influence the final kinematic solution. Additional constraints on the kinematic models may come from the integration of different seismic approaches and observables, such as the use of back-projections to map the average behavior of the rupture velocity, and of P and S wave polarizations from teleseismic or near-source records to get further constraints on source location, propagation and mechanism.

As a complementary approach, instead of searching for a single *unique* solution of the kinematic inversion, one could explore the different solutions obtained by several Monte-Carlo explorations and extract the common stable features to build up the resolved part of a kinematic source model. At one extreme, instead of using the data to retrieve one single model, the data could be used to falsify admissible models [123]. Following this approach, one could test a suite of possible kinematic models and retain the ones which explain the observations according to the associated uncertainties in the forward theory and data. Very recently several studies have investigated the statistical properties of the kinematic models obtained both by the inversion of seismic/geodetic data and as a result of the dynamic simulations [124,125]. The use of statistically admissible kinematic models coupled with an accurate computation of the Green's functions and the development of the computational resources may lead to the application of this approach in the near future.

Finally, only very recently seismic and geodetic data were inverted to directly obtain a dynamic description of the rupture [62,103]. A whole dynamic inversion would require the knowledge of the initial stress, the

yielding stress and the constitutive law that describes the evolution of the traction on the fault as a function of the slip (or the slip velocity). To provide reliable simulations, a certain degree of heterogeneity needs to be imposed on the fault, to capture the different phases of the dynamic rupture, from the nucleation to the propagation, the radiation and the arrest. The large number of degrees of freedom and the correlation among parameters, however, lead to a strongly ill-posed problem and require a huge amount of simulations, to hope to fall in the "minimum" of the selected objective function. The few attempts to obtain a dynamic model directly from the observations, used strong *a-priori* constraints on the dynamic parameters and on the shape of the constitutive function with the aim of reducing the number of the parameters, addressing the correlation among them and providing a (smooth) degree of heterogeneity for the stress on the fault. Nevertheless these inversions came out with a reasonable estimation of the fracture energy and a picture of the stress evolution on the fault plane consistent with the kinematic models.

7. References

1. Aki, K., and Richards, P.G. 1980, Quantitative Seismology, Freeman and Co., New York.
2. Burridge, R., and Knopoff., L. 1964,Bull. Seism. Soc. Am., 54, 1875-1888.
3. Hartzell, S.H. 1989, J. Geophys. Res., 94, 7515-7534.
4. Graves, R.W., and Wald, D.J. 2001, J. Geophys. Res., 106, 8745-8766.
5. Wald, D.J., and Graves, R.W. 2001, J. Geophys. Res., 106, 8767-8788.
6. Liu, P., and Archuleta, R.J. 2004, J. Geophys. Res., 109, B02318, doi:10.1029/2003JB002625.
7. Kennett, B.L.N., and Kerry, N.J. 1979, Geophys. J. Royal Astr. Soc., 57, 557-583.
8. Müller, G. 1985, J. Geophys., 58, 153-174.
9. Bouchon, M. 1981, Bull. Seism. Soc. Am., 71, 959-977.
10. Coutant, O. 1989, Programme de simulation numérique AXITRA, *Rapport LGIT*, Université Joseph Fourier, Grenoble, France.
11. Favreau, P., Mangeney, A., Lucas, A., Crosta, G., and Bouchut. F. 2010, Geophys. Res. Lett., 37, L15305, doi:10.1029/2010GL043512.
12. Zhu, L., and Rivera, L. 2001, Geophys. J Int., 148, 619-627.
13. Olson, A.H., Orcutt, J.A., and Frazier, G.A. 1984, Geophys. J. Royal Astr. Soc., 77, 421-460, doi: 10.1111/j.1365-246X.1984.tb01942.x.
14. Spudich, P., and Xu, L.S. 2003, Software for calculating earthquake ground motions from finite faults in vertically varying media, International Handbook of Earthquake & engineering seismology (Edi. W.H.K. Lee, H. Kanamori, P.C. Jennings, C. Kisslinger), Academic Press, Part B, 1857-1875.
15. Graves, R.W. 1996, Bull. Seism. Soc. Am., 86, (4), 1091-1106.

16. Moczo, P., Lucka, M., Kristek, J., and Kristekova, M. 1999, Bull. Seism. Soc. Am., 89, (1), 69-79.
17. Komatitcsh, D., and Vilotte, J.-P. 1998, Bull. Seism. Soc. Am., 88, (2), 368-392.
18. Chaljub, E., Komatitsch, D., Vilotte, J.-P., Capdeville, Y., Valette, B., and Festa, G. 2007, Spectral element analysis in seismology, in *Advances in Wave Propagation in Heterogeneous Media*, 48, 365-419, eds Wu, R., and Maupin, V., Advances in Geophysics, Elsevier - Academic Press.
19. Dumbser, M., and Käser, M. 2006, Geophys. J. Int., 167, (1), 319-336, doi:10.1111/j.1365-246X.2006.03120.x.
20. Nabelek, J. 1984, Determination of earthquake fault parameters from inversion of body waves, Ph.D. thesis, 361 pp., Mass. Inst. of Technol., Cambridge, Massachusetts.
21. Irikura, K. 1983, Bull. Disas. Prev. Res. Inst., 33, 63-104.
22. Hutchings, L. 1991, Bull. Seism. Soc. Am., 81, 88-121.
23. Nozu, A., and Irikura, K. 2008, Bull. Seism. Soc. Am., 98, 180-197, doi: 10.1785/0120060183.
24. Hutchings, L., and Wu, F. 1990, J. Geophys. Res., 95, 1187-1214.
25. Hutchings, L. 1994**,** Bull. Seism. Soc. Am., 84, (4), 1028-1050.
26. Piatanesi, A., Tinti, E., Cocco, M., and Fukuyama, E. 2004, Geophys. Res. Lett., 31, L04609, doi:10.1029/2003GL019225.
27. Heaton, T.H. 1990, Phys. Earth Planet. Inter., 64, 1-20.
28. Haskell, N.A. 1964, Bull. Seism. Soc. Am., 54, 1811-1841.
29. Sato, T., and Hirasawa, T. 1973, J. Phys. Earth, 21, 415-431.
30. Cotton, F., and Campillo, M. 1995, J. Geophys. Res., 100, 3961-3976.
31. Nielsen, S., and Madariaga, R. 2003, Bull. Seism. Soc. Am., 93, 2375-2388.
32. Tinti, E., Fukuyama, E., Piatanesi, A., and Cocco, M. 2006, Bull. Seism. Soc. Am., 95, (4), 1211-1223.
33. Yoffe, E. 1951, Phil. Mag., 42, 739-750.
34. Freund, L.B. 1979, J. Geophys. Res., 84, 2199-2209.
35. Hartzell, S.H., and Heaton, T.H. 1983, Bull. Seism. Soc. Am., 73, 1553-1583.
36. Cohee, B.P., and Beroza, G.C. 1994, Ann. Geophys., 37, 1515-1538.
37. Guatteri, M., and Spudich, P. 2000, Bull. Seism. Soc. Am., 90, 98-116.
38. Moczo, P., Kristek, J., and Halada, L. 2000, Bull. Seism. Soc. Am., 90, (3), 587-603.
39. Lucca, E. 2011, Kinematic description of the rupture from strong motion data: strategies for a robust inversion, Ph.D. thesis, 173 pp., Università di Bologna "Alma Mater Studiorum", Bologna, Italy.
40. Hughes, T.J.R. 1987, The finite Element method, Prentice-Hall Inc., New Jersey.
41. Emolo, A., and Zollo, A. 2005, Bull. Seismol. Soc. Am., 95, (3), 981-994, doi:10.1785/0120030193.
42. Podvin, P., and Lecomte, I. 1991, Geophys. J. Int., 105, 793-804.
43. Festa, G., and Zollo, A. 2006, Geophys. J. Int., 166, 745-756, doi:10.1111/j.1365-246X.2006.03045.x.
44. Olson, A.H. 1987, Phys. Earth Planet. Inter., 47, 333-345.
45. Gallovič, F., and Zahradník, J. 2011, J. Geophys. Res., 116, B02309.

46. Beroza, G.C., and Spudich, P. 1988, J. Geophys. Res., 93, 6275-6296.
47. Luenberger, D.G. 1984, Linear and non linear programming, pp. 277-304, Addison-Wesley, Reading, Massachusetts.
48. Tarantola, A. 2004, Inverse problem theory and methods for model parameter estimation, SIAM.
49. Gallovič, F., Zahradník, J., Krizova, D., Plicka, V., Sokos, E., Serpetsidaki, A., and Tselentis, G.-A. 2009, Geophys. Res. Lett., 36, L21310, doi:10.1029/2009GL040283.
50. Lawson, C.L., and Hanson, R.J. 1974, Solving Least Square Problems, Prentice-Hall, Inc., New Jersey.
51. Ide, S., and Takeo, M. 1997, J. Geophys. Res., 102, (B12), 27,379-27,391.
52. Sekiguchi, H., Irikura, K., and Iwata, T. 2000, Bull. Seism. Soc. Am., 90, 117-133.
53. Akaike, H. 1980, Likelihood and the Bayes procedure (with discussion), Bayesian Statistics, 143-165, University Press, Valencia, Spain.
54. Kirkpatrick, S., Gerlatt, C.D., and Vecchi, M.P. 1983, Science, 220, 671-680.
55. Rothman, D. 1986, Geophysics, 51, 332-346.
56. Ji, C., Wald, D.J., and Helmberger, D.V. 2002, Bull. Seism. Soc. Am., 92, (4), 1192-1207.
57. Piatanesi, A., Cirella, A., Spudich, P., and Cocco, M. 2007, J. Geophys. Res., 112, B07314, doi:10.1029/2006JB004821.
58. Goldberg, D.E. 1989, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, New York.
59. Sambridge, M. 1999, Geophys. J. Int., 138, 479-494.
60. Nelder, J.A., and Mead, R. 1965, Comput. J., 7, 308-313.
61. Vallée, M., Bouchon, M., and Schwartz, S.Y. 2003, J. Geophys. Res., 108, B4, 2203.
62. Di Carli, S., Holden, C., Peyrat, S., and Madariaga, R. 2010, J. Geophys. Res., 115, B12328, doi:10.1029/2009JB006358.
63. Vallée, M., and Bouchon, M. 2004, Geophys. J. Int., 156, 615-630.
64. Zhao, L.-S., and Helmberger, D.V. 1994, Bull. Seism. Soc. Am., 84, (1), 91-104.
65. Sen, M., and Stoffa, P.L. 1991, Geophysics, 56, 1624-1638.
66. Hartzell, S.H., Liu, P., and Mendoza, C. 1996, J. Geophys. Res., 101, 20, 091- 20,108.
67. Asano, K., Iwata, T., and Irikura, K. 2005, Bull. Seism. Soc. Am., 95, 1701-1715.
68. Asano, K., and Iwata, T. 2009, Bull. Seism. Soc. Am., 99, 123-140.
69. Liu, P., Custodio, S., and Archuleta, R.J. 2006, Bull. Seism. Soc. Am., 96, (4B), S143 – S158, doi:10.1785/0120050826.
70. Page, M.T., Custodio, S., Archuleta, R.J., and Carlson, J.M. 2009, J. Geophys. Res., 114, B01314, doi:10.1029/2007JB005449.
71. Farra, V., Bernard, P., and Madariaga, R. 1986, Fast near source evaluation of strong motion for complex source models, in *Earthquake Source Mechanics*, American Geophysical Monograph 37, S. Das, J. Boatwright, and C. H. Scholz (Editors), 121-130.

72. Ishii, M., Shearer, P.M., Houston, H., and Vidale, J.E. 2005, Nature, 435, 933-936.
73. Krüger, F., and Ohrnberger, M. 2005, Nature, 435, 937-939, doi 10.1038/nature03696.
74. Maercklin, N., Zollo, A., Orefice, A., Festa, G., Emolo, A., De Matteis, R., Delouis, B., and Bobbio, A. 2011, Bull. Seism. Soc. Am., 101, (1), 354-365.
75. Bernard, P., and Madariaga, R. 1984, Bull. Seism. Soc. Am., 74**,** 539-557.
76. Spudich, P., and Frazer, L.N. 1984, Bull. Seism. Soc. Am., 74**,** 2061-2082.
77. Festa, G., Zollo, A., and Lancieri, M. 2008, Geophys. Res. Lett., 35, L22307, doi:10.1029/2008GL035576.
78. Pulido, N., Aoi, S., and Fujiwara, H. 2008, Earth Planets Space, 60, 1-6.
79. Jakka, R.S., Cochran, E.S., and Lawrence, J.F. 2010, Geophys. J. Int., 182, 1058-1072. doi: 10.1111/j.1365-246X.2010.04670.x.
80. Larmat, C., Montagner, J.-P., Fink, M., Capdeville, Y., Tourin, A., and Clévédé, E. 2006, Geophys. Res. Lett., 33, L19312, doi:10.1029/2006GL026336.
81. Cotton, F., Pousse, G., Bonilla, F., and Scherbaum, F. 2008, Bull. Seism. Soc. Am., 98, 2244-2261.
82. Trifunac, M.D., and Todorovska, M.I. 2001, Soil Dyn. Earth. Eng., 21, 275-286.
83. Boore, D.M., Stephens, C.D., and Joyner, W.B. 2002, Bull. Seism. Soc. Am., 92, 1543-1560.
84. Bogdanov, V.E., and Graizer, V.M. 1976, The determination of the residual displacement of the ground from the seismogram. Reports of the Academy of Sciences of the USSR, 229, 59-62.
85. Iwan, W.D., Moser, M.A., and Peng, C.-Y. 1985, Bull. Seism. Soc. Am.,75, 1225-1246.
86. Boore, D.M. 2001, Bull. Seism. Soc. Am., 91, 1199-1211.
87. Burdick, L.J., and Helmberger, D.V. 1978, J. Geophys. Res., 83, 1699-1712.
88. Ammon, C.J., Ji, C., Thio, H.-K., Robinson, D., Ni, S., Hjorleifsdottir, V., Kanamori, H., Lay, T., Das, S., Helmberger, D.V., Ichinose, G., Polet, J., and Wald, D. 2005, Science, 308, (5725), 1133-1139, doi:10.1126/science.1112260.
89. Delouis, B., Nocquet, J.-M., and Vallée, M. 2010, Geophys. Res. Lett., 37, L17305, doi:10.1029/2010GL043899.
90. Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B., and Boucher, C. 2007, J. Geophys. Res., 112, B09401, doi:10.1029/2007JB004949.
91. Larson, K., Bilich, A., and Axelrad, P. 2007, J. Geophys. Res., 112, B05422, doi:10.1029/2006JB004367.
92. Larson, K., Bodin, P., and Gomberg, J. 2003, Science, 300, 1421–1424, doi:10.1126/science.1084531.
93. Ji, C., Larson, K.M., Tan, Y., Hudnut, K., and Choi, K. 2004, Geophys. Res. Lett., 31, (17), L17608, doi:10.1029/2004GL020448.
94. Miyazaki, S., Larson, K., Choi, K., Hikima, K., Koketsu, K., Bodin, P., Haase, J., Emore, G., and Yamagiwa, A. 2004, Geophys. Res. Lett., 31, (21), L21603, doi:10.1029/2004GL021457.
95. Vigny,C., Socquet, A., Peyrat, S., Ruegg, J.-C., Métois, M., Madariaga, R., Morvan, S., Lancieri, M., Lacassin, R., Campos, J., Carrizo, M., Bejar-Pizarro,

M., Barrientos, S., Armijo, R., Aranda, C., Valderas-Bermejo, M.-C., Ortega, I., Bondoux, F., Baize, S., Lyon-Caen, H., Pavez, A., Vilotte, J.-P., Bevis, M., Brooks, B., Smalley, R., Parra, H., Baez, J.-C., Blanco, M., Cimbaro, S., and Kendrick, E. 2011, *Science*, 332, (6036), 1417-1421, doi:10.1126/science. 1204132.

96. Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., and Rabaute, T., 1993, Nature, 364, 138-142.
97. Tanioka, Y., Yudhicara, Y., Kususose, T., Kathiroli, S., Nishimura, Y, Iwasaki, S.I., and Satake, K. 2006, Earth Planets Space, 58, 203-209.
98. Piatanesi, A., and Lorito, S. 2007, Bull. Seism. Soc. Am., 97, S223-S231, doi:10.1785/0120050627.
99. Lorito, S., Romano, F., Atzori, S., Tong, X., Avallone, A., McCloskey, J., Cocco, M., Boschi, E., and Piatanesi, A. 2011, Nature Geosci., 4, 173-177, doi:10.1038/ ngeo1073.
100. Iwata, T., and Sekiguchi, H. 2002, Source process and near-source ground motion during the 2000 Tottori-ken seibu earthquake (in Japanese with English abstract), Proc. 11th Japan Earthq. Eng. Symp., 125-128.
101. Fukuyama, E., Ellsworth, W.L., Waldhauser, F., and Kubo, A. 2003, Bull. Seism. Soc. Am., 93, (4), 1468-1478.
102. Semmane, F., Cotton, F., and Campillo, M. 2005, J. Geophys. Res., 110, B03306, doi:10.1029/2004JB003194.
103. Peyrat, S., and Olsen, K.B. 2004, Geophys. Res. Lett., 31, L05604, doi:10.1029/ 2003GL019058.
104. Zhu, L. 2003, Geophys. Res. Lett., 30, 1740.
105. Boore, D.M., and Bommer, J.J. 2005, Soil Dyn. Earth. Eng., 25, 93-115.
106. Yagi, Y. 2001, Source rupture process of the Tottori-ken Seibu earthquake of the October 6, 2000 obtained by joint inversion of the near field and teleseismic data, Meeting of Japan Earth and Planetary Science, Tokyo, Japan.
107. Dietz, L.D., and Ellsworth, W.L. 1990, Geophys. Res. Lett., 17, 1417-1420.
108. Beroza, G.C. 1996, Rupture history of the earthquake from high frequency strong motion data, in *The Loma Prieta, California, earthquake of October 17, 1989- main shock characteristics, U.S. Geol. Surv. Profess. Pap. 1550-A*, P. Spudich (Editor), 9-32.
109. Zhang, J., and Lay, T. 1990, Geophys. Res. Lett., 17, 1195-1198.
110. Boatwright, J., Fletcher, J.B., and Fumal, T.E. 1991, Bull. Seism. Soc. Am., 81, 1754-1782.
111. Chin, B.H., and Aki, K. 1991, Bull. Seism. Soc. Am., 81, 1859-1884.
112. Somerville, P., and Yoshimura, J. 1990, Geophys. Res. Lett., 17, 1203-1206.
113. Zeng, Y., Aki, K., and Teng, T. 1993, J. Geophys.Res., 98, 11,981-11,993.
114. Zollo, A., Bobbio, A., Emolo, A., Herrero, A., and De Natale, G. 1997, J. Seism., 1, 305-319.
115. Wald, D.J., and Heaton, T.H. 1994, Bull. Seism. Soc. Am., 84, (3), 668-691.
116. Hernandez, B., Cotton, F., and Campillo, M. 1999, J. Geophys. Res., 104, (B6), 13083-13099.

117. Zeng, Y., and Anderson, J. 2000, Evaluation of numerical procedures for simulating near-fault long-period ground motions using Zeng method, Report 2000/01 to the PEER Utilities Program, available at http://peer.berkeley.edu.
118. Mai, P.M., Burjanek, J., Delouis, B., Causse, M., Festa, G., Monelli, D., Uchide, T., and Zahradnik, J. 2007, Earthquake source inversion blindtest: initial results and further development, AGU Fall Meeting, San Francisco.
119. Causse, M., Chaljub, E., Cotton, F., Cornou, C., and Bard, P.-Y. 2009, Geophys. J. Int., 179, 1627-1644, doi: 10.1111/j.1365-246X.2009.04354.x.
120. Capuano, P., Zollo, A., and Singh, S.K. 1994, Ann. Geophys., 37, (6), 1659-1677.
121. Courboulex F., Deichmann, N., and Gariel, J.C. 1999, Geophys. J. Int, 139, 152-160.
122. Cirella, A., Piatanesi, A., Cocco, M., Tinti, E., Scognamiglio, L., Michelini, A., Lomax, A., and Boschi, E. 2009, Geophys. Res. Lett., 36, L19304, doi:10.1029/2009GL039795.
123. Tarantola, A. 2006, Nature physics, 2, 492-494, doi:10.1038/nphys375.
124. Mai, P.M., and Beroza, G.C. 2002, J. Geophys. Res., 107, (B11), 2308, doi:10.1029/2001JB000588.
125. Song, S.G., and Sommerville, P. 2010, Bull. Seism. Soc. Am., 100, (2), 482-496, doi: 10.1785/0120090134.

Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

The Mechanics of Faulting: From Laboratory to Real Earthquakes, 2012: 63-91
ISBN: 978-81-308-0502-3 Editors: Andrea Bizzarri and Harsha S. Bhat

3. What real earthquakes can tell us about the fundamentals of faulting

David P. Robinson
Department of Earth Sciences, University of Oxford, South Parks Rd. Oxford, OX1 3AN, UK

Abstract. Integration of observation, theory and experiment has led to significant advances in understanding the earthquake faulting process in the past 30 years. Progress has typically been made in an iterative fashion whereby a new theory or observation has led to a re-assessment of the existing ideas and consequent re-evaluation of some concepts. This paper focuses on one such strand of work, namely the understanding of supershear ruptures. Early theoretical work suggested that supershear ruptures are possible but unlikely to occur in the natural environment. This view was essentially unchallenged until the start of the 21st Century when observations of several earthquakes that had occurred at that time combined to suggest that supershear earthquakes could occur. This observation led to several carefully constructed lab experiments that aimed to produce supershear ruptures similar to those observed by seismologists. The majority of seismologists now accept that supershear earthquakes can and do occur, and now the focus of work is to try to understand the circumstances under which these earthquakes occur and the effects that they may produce. The improving quality of data recorded by the world's seismometers combined with novel geodetic methods for investigating static displacements and a greater than average number of significant

Correspondence/Reprint request: Dr. David P. Robinson, Department of Earth Sciences, University of Oxford South Parks Rd. Oxford, OX1 3AN, UK. E-mail: davidr@earth.ox.ac.uk

earthquakes in recent years present a unique opportunity to advance the understanding of earthquake faulting, and it is likely that progress similar to the development of ideas governing supershear earthquakes will happen in the coming years.

1. Introduction

The ultimate goal of earthquake seismologists is to mitigate the risk associated with them. In order to do this, it is essential that the fundamentals of earthquake faulting are be understood. While laboratory experiments and theory may tell us of the possible modes of rock failure, we cannot be sure exactly how these scale to real-world situations unless observations of earthquakes are made. Hazard maps can be constructed based on the expected repeat times of earthquakes, but any map that relies only on the historical record of earthquakes in an area will be fatally flawed as it makes the assumption that the historical data records all potential earthquakes on a fault. Recent devastating earthquakes have highlighted how earthquakes with very long repeat times can strike in areas with low perceived risk [1,2,3]. Thus to mitigate risk, seismologists must be in a position to determine the full range of possible earthquakes in a given area, even if some types of earthquake are not observed in the historical record.

With the advent of digital seismometers in the latter part of the 20th Century, earthquake seismologists have developed a number of methods to invert seismograms to determine the rupture history of an earthquake. The methods developed for inverting seismograms will be discussed in more detail in a later section, but suffice to say, assumptions regarding the rupture process are typically made. These assumptions are primarily made to stabilise the inversion process [4], but also have the effect of reducing the number of degrees of freedom of the problem investigated; an important consideration with limited computing power. Early methods chose a set of assumptions based on theoretical considerations [5,6,7]. Although as time has progressed, some of these assumptions and conditions have been called into question by observations of earthquakes. This consequently leads to a re-evaluation of the models of fracture propagation used. The iterative approach used with earthquake observation leading to improved models of rupture propagation which in turn leads to better models of real earthquakes being constructed is one which has seen the field move forward considerably in recent years. This coupled with the unusually high number of very large earthquakes producing good data in the first decade of the 21st Century (Figure 1) means that the most comprehensive models of earthquake rupture ever produced are being constructed today. It is likely that considerable further progress will be made

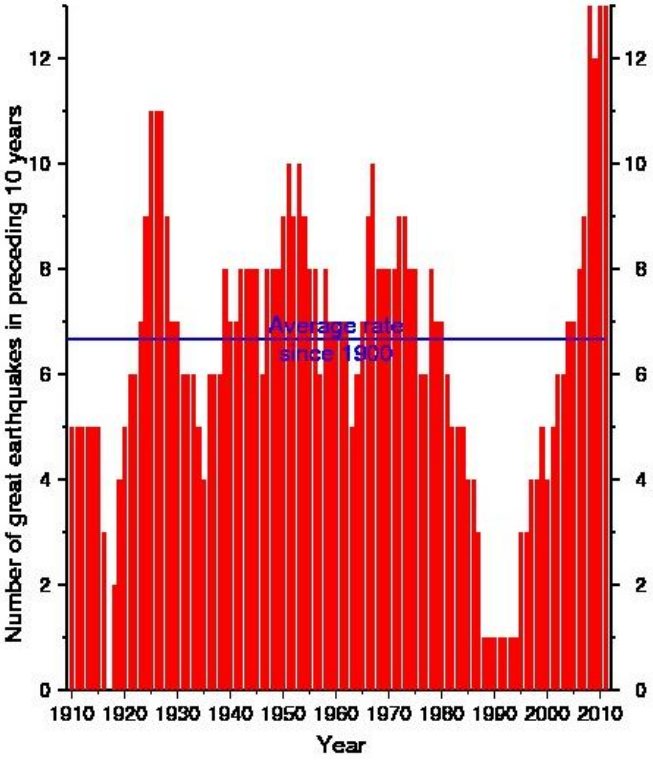


Figure 1. Number of great earthquakes (with magnitude ≥8.0) occurring in the 10 year period preceding January 1 of each year since 1910 is shown as red bars. The average rate of great earthquakes since 1900 (0.67 per year) is indicated with a blue line. Magnitudes for earthquakes are the M_w reported by Pacheo & Sykes (1992) [8] for earthquakes prior to 1977 and shallower than 50km, and M_s as reported by the Internation Seismological Centre (ISC) for deeper earthquakes in this period. The magnitude of earthquakes since the start of 1977 are values of M_w reported by the Global CMT project (*GCMT*) [9].

in the coming years with the hope that the advances made will be used by policy makers to reduce the hazard associated with earthquakes.

2. Theory of faulting

A full description of the theory of faulting encompasses not only considerations of earthquake seismologists but also material scientists and is

not necessary for the purposes of this paper, however the key points are outlined in this section. It is widely acknowledged that earthquakes propagate as shear-plane cracks under the influence of a 'rate-weakening' friction law, that is to say that the frictional resistance to slip is greatly reduced while slip is occurring [10,11]:

$$\mathrm{T}(D) = \mathrm{T}_u\left(\frac{1-D}{D_c}\right) + \sigma_f \qquad \text{for } D < D_c$$

$$\mathrm{T}(D) = \sigma_f \qquad \text{for } D > D_c \tag{1}$$

Where T(D) is the stress within a fault surface, D_c is a characteristic slip distance and $\mathrm{T_u} = \sigma_u - \sigma_f$; the yield stress minus the residual stress. For an earthquake to initiate there must be an initial stress patch that is ready to break [12,13]. It must be noted here that there is no constraint that the values of sigma should be constant over a fault surface. Indeed, it is highly likely that heterogeneity in the fault surface will produce big variations in the frictional constants that determine σ_f and σ_u. Similarly, the initial stress on a fault is a combination of tectonic loading, the residual stress field from previous earthquakes on that fault, and the induced stress from earthquakes on other nearby faults. Hence it seems highly unlikely that any fault will behave as a uniformly loaded surface with constant frictional properties in an earthquake, however, this assumption was frequently used as a simplification when constructing models of theoretical rupture propagation [12,13,14,15,16].

Although the exact details of the slip weakening law can be debated, the consequence of the relationship as expressed in equation (1) can be seen in Figure 2, namely that there is energy released by a release of stress by the propagating crack, this is termed the fracture surface energy. The rupture velocity is such that the energy absorbed at the crack tip in creating a new fracture surface is equal to the fracture surface energy [17]. This implies that there are certain rupture velocities where a propagating crack is stable. The question of permitted rupture velocities, in particular in relation to supershear ruptures, will be returned to later in this paper. One consequence of equation (1) is that a dimensionless value, S, a measure of the material strength of the fault relative to the tectonic stress can be calculated [18]:

$$S = \left(\frac{\sigma_u - \sigma_0}{\sigma_0 - \sigma_f}\right) \tag{2}$$

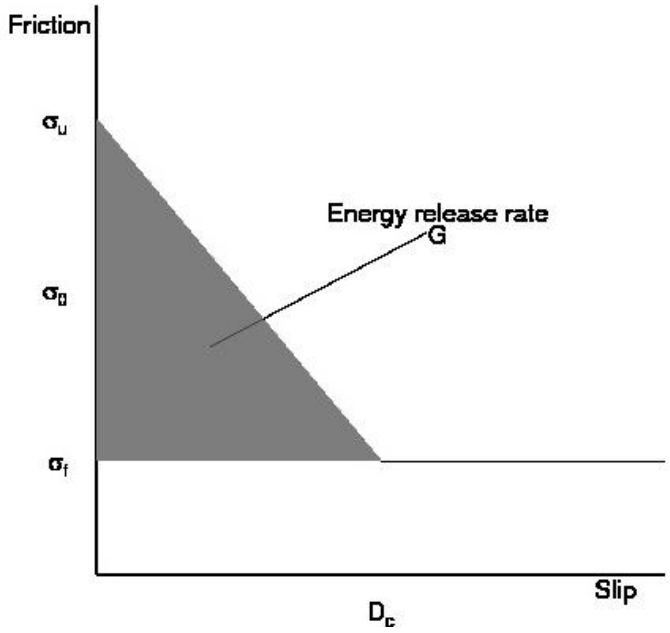


Figure 2. Shows the consequence of the slip weakening law given in equation (1). The energy radiated, G is given by the formula: $G = \int_0^{D_c} T(D)dD = \frac{1}{2} T_u D_c$ where D_c is the slip weakening distance and is shown shaded in the diagram. Modified from Madariaga et al 2000 [11].

Where σ_0 is the initial stress on a fault. Numerical simulations of the behaviour of cracks propagating on faults with varying values of S have been carried out in 2 and 3 dimensions and the results of these models are discussed in more detail later in the text, however, the main conclusion is that, under most circumstances, cracks begin at rest and accelerate up to a terminal rupture velocity which tends towards the shear wave velocity [18]. The addition of "barriers", can cause the rupture to terminate [19]. Barriers can be areas of either low effective stress either as a result of lower than normal initial stress, or higher than normal residual stress due to higher than normal friction. Relating this to the earth, barriers can either be areas of the fault with low coupling and hence a higher proportion of aseismic slip (low effective stress) or areas with high coupling (higher normal friction), hence act as a locked portion of the fault.

3. Methods of observing earthquake rupture

The combination of theoretical work on rupture propagation and the availability of digital records from earthquakes meant that by the late 1970s methods could be devised to determine the earthquake properties. This is perhaps most famously highlighted by the development of the Harvard CMT project [20] which, to this day, produces centroid locations, magnitudes and mechanisms under the moniker of the *Global CMT* (*GCMT)* [9] project. Although this is the most visible example of earthquake 'inversion', due to the frequency of seismic data it uses, it effectively shrinks the earthquake under investigation to be a point source, the centroid. By comparing the position in space and time of the earthquake hypocentre with that of the centroid, it may be possible to draw some broad conclusions about the rupture duration and any directivity in the rupture for very large earthquakes but, apart from that, there is precious little information about the details of an earthquake rupture contained in its GCMT record. The majority of methods for determining the rupture process of earthquakes use higher frequency data, that allow the earthquake to be treated as a source with finite length, width and duration and these are outlined in this section.

The 1979 Imperial Valley earthquake provided the first real opportunity for seismologists to develop tools for the inversion of earthquake data to find the rupture history of an earthquake. Essentially the problem consists of attempting to produce synthetic seismograms that resemble as closely as possible the real seismograms recorded. The earthquake fault is split into a series of discrete cells along strike and dip with cells permitted to slip at various points in time. The individual cells are small enough in comparison to the frequency of seismic data being investigated to be considered as point sources. Consequently, given the crustal structure at the source and, to a lesser extent, at the receiver, a Green's function response to a point source dislocation at any individual cell can be calculated. Thus the recording made at any station is the sum of all the Green's functions multiplied by the appropriate slip at each point in space and time taken over the entire fault, or mathematically:

$$u_k(\mathbf{x}_1, t_1) = \int_0^{t_1} dt \iint_{fault} \mathrm{K}(\mathbf{x}_1, \mathbf{x}, t_1, t) a_i(\mathbf{x}, t) dS \tag{3}$$

where i,k=1,2,3, $u_k(\mathbf{x_1},t_1)$ are the components of the displacement vector at any point, $a_i(\mathbf{x},t)$ are the components of slip on the fault and $K_{ik}(\mathbf{x_1},\mathbf{x},t_1,t)$ are the components of the impulse response of the medium at $(\mathbf{x_1},t_1)$ to a point

source dislocation at (**x,**t). The exact version of equation (3) can be approximated, following some transformations [5,4] to the form:

$$b_k \approx \sum A_{ik} x_i \tag{4}$$

where b_k is the data recorded at point k, x_i is the sliprate at point *i* and A_{ik} is the Green's functions relating the sliprate at *i* to the response at *k*. It should be noted here that *k* and *i* are points in space and time, that is to say that the time-varying position vectors in (3) have been replaced by a 1-dimensional numbering system adopted when approximating the equation. This can be rewritten as:

$$\mathbf{b} \approx \underline{A}\mathbf{x} \tag{5}$$

which is a matrix equation that is familiar to many readers. For seismic studies this equation can be shown in graphical form [6,21] and for illustrative purposes such a diagram is shown in figure 3.

The forward problem is relatively trivial; if the matrix of Greens functions is known along with the slip history of each cell on the fault, then it is fairly easy to calculate the synthetic seismograms that should be recorded. This can be used to deduce the rupture process by a process of trial and educated guess [22]. However the inverse problem, namely what is the optimal slip distribution to describe a given data set, is much more difficult. In the inverse problem, equation (5) is rewritten:

$$\underline{A}\mathbf{x} - \mathbf{b} = \mathbf{r} \tag{6}$$

And the vector, **r,** is minimised. Usually the ℓ_1, ℓ_2 or ℓ_∞ norm is minimised, all 3 being equivalent in the sense that they tend to zero simultaneously. Additionally there are further complications when the physical properties of the problem under study are considered:

1. What is the mechanism of the slip in each cell? The mechanism of faulting at a point source has a profound effect on the radiation pattern, hence if the direction of slip varies along the fault, then the matrix of Green's functions in equations (5) and (6) will also vary. Without accurate knowledge of how the mechanism of faulting varies over the length of an earthquake, an accurate model of the earthquake rupture history cannot be deduced.
2. Are there physical rules which the cells must obey? There are many examples that have been suggested by theoretical studies, such as slip

must always be non-negative (the 'no backslip' constraint) or slip is terminated by the passage of a healing front.

3. Are there constraints that apply to the rupture front? Theoretical studies have suggested that there are stable velocities that ruptures should proceed at. It may be appropriate to add this as a constraint to the problem in the same way that a healing front may be applied. A similar result can be obtained by specifying a given rise-time for each cell once it has begun to slip.
4. Are there distributions of slip that are physically more realistic? For example it may be unrealistic for there to be cells that have zero or very low slip immediately neighbouring cells with very high slip, it may be more appropriate to minimise the difference in slip between neighbouring cells.
5. Are there any other constraints to how much slip has taken place or where it has occurred? There may be information from ground or geodetic observations, which equally raise the question of how well does surface deformation translate to deformation at depth. Equally, it may be thought that the total moment of an earthquake is sufficiently well known from other studies (such as the GCMT solution) to add that as a constraint.

The choice of which assumptions to apply to a given earthquake inversion may have profound effects on the model produced. The fact is that in dicretizing the problem in moving from equation (3) to equation (4) in order to solve problems using computers means that there is intrinsically an error introduced, represented by the approximately equals sign in equations (4) and (5). This means that any method derived from this approach will only produce a solution which cannot explain all features observed in the data. There is no guarantee that the best solution mathematically closely resembles what occurred in reality. Indeed it has been shown that when inverting realistic synthetic data, unless constraints are imposed, the solutions obtained are not necessarily physically meaningful [23]. It may be correct that adding certain assumptions to models of various earthquake mean that the solutions produced are more likely to accurately model reality, however, due to the intrinsic 'approximate' nature of the problem there is a danger of a 'satisfactory' model of an earthquake being produced including a false assumption when a superior model might exist where that assumption is not applied.

The previous discussion is deliberately vague in that the principles outlined can be applied to data at any frequencies and can then be solved using an appropriate inverse method. Equally, similar equations lead to a

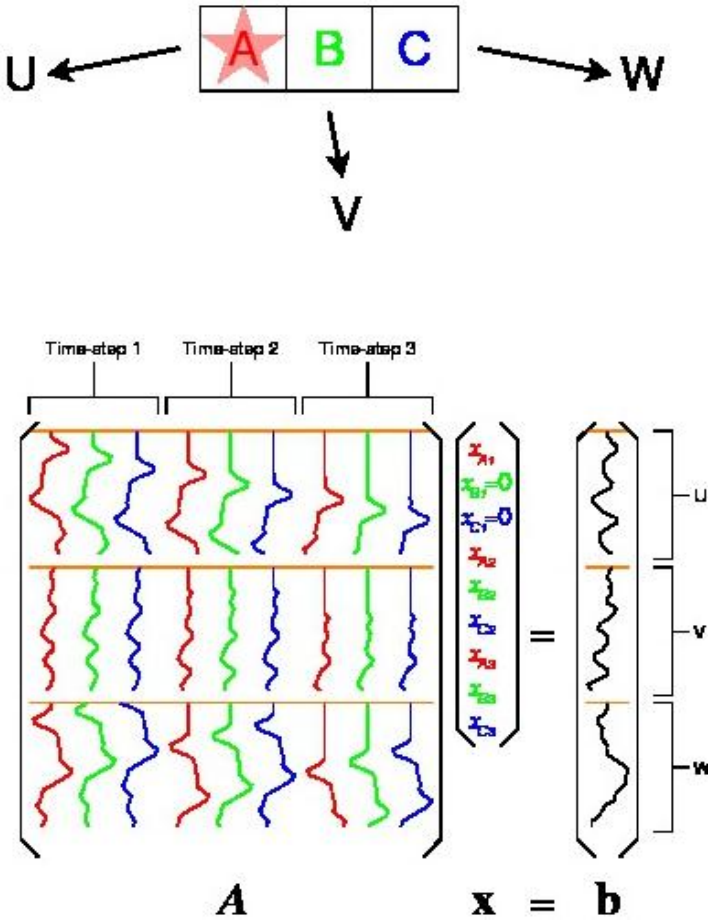


Figure 3. Simple visual representation of equation (4), for the case of an earthquake modelled by 3 cells and 3 time-steps recorded at 3 station. Slip initiates in cell A, propagates through cells B and C and is recorded at 3 seismic stations, U, V and W in the approximate directions shown at the top of the diagram. Greens functions relating the impulse response of each of the stations to slip in each of the cells is calculated and shifted by an appropriate time depending on the position of the cell in relation to the recording station and the timing of slip, and are colour coded relative to the cell diagram at the top of the figure. The sliprate vector, x is similarly colour coded with the subscript denoting the cell and time-step that slip occurs on. Due to causality, the sliprate at time 1 is zero in all cells apart from that containing the hypocentre (i.e. x_{B1}=x_{C1}=0).

similar problem, without the time dimension, in determining the final or static displacement caused slip on a finite fault that can be modelled using Synthetic Interferometic Apature Radar (InSAR), GPS or other similar data (e.g. Massonnet and Feigl (1998) [24]). Early seismological models initially focussed on locally recorded strong motion data using minimisation either of a least squared [5] or a cross-correlation error function between synthetics and data [25] each with their own inherent assumptions. Extensions were made to show that teleseismic data could equally be used [6]. Once it had been shown that teleseismic body wave could be used to produce models the door was open to potentially study any sufficiently large earthquake. This period in seismology was characterised by a low number of very large earthquakes (see figure 1) as anomalous as the current very high rate of very large earthquakes. Indeed there were no great earthquakes in the first half of the 1980s. The Andreanof Islands earthquake of 1986 presented a significant opportunity to test and further develop methods.

The 1986 Andreanof islands earthquake was studied by many authors using a variety of different methods [26,27,28,29,4]. Although the number of digital stations recording the earthquake was low, intermediate and long period body waves were successfully used to find a rupture history [26] along with higher frequency (up to 5Hz) body wave data [28] and broadband body wave data [29,4]. Despite the success at producing models for this earthquake that can explain the recorded data for this earthquake, there are significant differences between models produced by different authors. This highlights the point made previously that the exact method for inverting earthquake data and the inherent assumptions involved in the inversion method are of vital importance. Potentially extending the portion of the earthquake record under investigation could help resolve these discrepancies. Surface waves potentially extend the usable portion of the seismic record that can be modelled and have also been used to determine rupture histories of earthquakes. Surface waves recorded at local strong motion stations were used to analyse the Imperial Valley earthquake [5]. However, the use of teleseismic surface waves is complicated by the fact that the waves have travelled large distances through the highly heterogeneous crust as opposed to body waves that have the vast majority of their ray paths within the relatively homogenous mantle. Additionally attenuation of the high frequency portion of the surface wave means that teleseismic surface waves are dominated by relatively long-period (20 seconds and upwards [30]) energy consequently their ability to resolve fine details of a rupture history is severely curtailed. Nevertheless, surface waves have proved useful in imaging the ruptures of very long earthquakes [30,2].

As already mentioned, early attempts at modelling earthquakes resulted in earthquake models being produced by different authors that were very different to each other. Attempts to improve the inversion technique included the replacement of calculated Greens functions in equations (4)-(6) with Empirical Greens Functions (EGFs) from the recordings of relatively small fore- or after-shocks [31]. This approach failed to radically improve the disparity between models proposed by different authors. Consequently the approach frequently adopted today is to examine a suite of models that satisfy the available data and identify features that are common across all solutions. This approach was hinted at in Olsen and Apsel's 1982 [5] paper when they discuss the conditions for a model to be acceptable:

"*…3. If more than one solution fits the data equally well, additional information must be supplied to uniquely define which solution is being obtained.*"

Today the corollary is used, namely that in the absence of additional data, all solutions are acceptable, and only features common across solutions are considered robust. A method for perturbing a model and investigating the robustness of features within it was described more than 15 years ago [32].

4. Observations of 'ordinary' earthquakes

As previously mentioned, many of the techniques for investigating earthquakes were developed in the early 1980s, a period where there were relatively few very large earthquakes. Figure 1 shows how the frequency of great ($M_w \geq 8.0$) earthquakes has varied with time since the start of the 20th Century and has two major striking features; the high number of great earthquakes since the start of the 21st century and the low number of great earthquakes through the 1980s and 1990s. Great earthquakes are important as their large size means that a greater degree of resolution of details is possible with seismic records of a given frequency range, whereas their potential for widespread destruction means they are of interest to society and policy makers in general. The rupture area of an earthquake is related to its moment by the relationship:

$$M_0 = \mu A \bar{s} \tag{7}$$

Where M_0 is moment, μ is the modulus of rigidity of the material breaking, A is the rupture area and $\bar{s}$ is the average slip [33]. Great earthquakes have $M_0 \geq 1.25 \times 10^{21}$ Nm which implies a minimum rupture area of ~10,000 km^2 when

sensible values of μ and $\bar{s}$ are substituted into equation (7). Teleseismic body wave data is essentially devoid of frequencies higher than ~0.5 Hz due to attenuation along the ray path hence the smallest feature that can be accurately modelled in a rupture inversion is of the order 10 km in size. Teleseismic surface wave data contains even less high frequency data hence can only model at a correspondingly coarser degree. Very large earthquakes consequently provide the best opportunity to model the earthquake rupture process; it can be guaranteed that there will be digital high quality data available in a range of azimuthal directions with sufficient resolution to allow splitting the rupture area into a series of smaller sub-faults. There are several places worldwide where there is a sufficiently dense network of seismic stations to allow the investigation of smaller earthquakes using local high frequency data, comparing the models presented for great earthquakes to those produced for smaller earthquakes allows the testing of various hypotheses of earthquake scaling relations.

As previously discussed, researchers have now been modelling the rupture process of real earthquakes for some 30 years, and much has been learnt in that time. The greatest progress is usually made in the immediate aftermath of significant earthquakes. The term 'significant' is used here to mean one of two things; (i) the earthquake in question is unusual in terms of its size or human impact which results in a disproportionate amount of scientific interest being shown in it, (ii) the earthquake cannot be adequately modelled using existing techniques. The paucity of large earthquakes during the 1980s and 90s meant that any great earthquake during that period attracted a great deal of scientific interest, and techniques were developed and honed as they were applied to successive earthquakes. For example the methods applied to the 1986 Andreanof Islands earthquake [4] were improved and expanded on following the 1989 MacQuarie Ridge earthquake [32].

An example of an earthquake not being adequately modelled using existing techniques is the 2004 Sumatra-Andaman Islands earthquake. Existing body-wave methods for determining the rupture history for this earthquake could not image the latter portions of the earthquake rupture because of the earthquake's very long rupture duration [2]. Body wave solutions rely on the accurate production of Green's functions to be input into the matrix *A* in equation (6). These can be produced if the crustal structure is known to sufficiently high resolution. However, the Green's functions calculated typically rely on the assumption that the mantle is an infinite half-space. This assumption is clearly not true but is unimportant if the duration of the data being modelled is sufficiently short to ensure core-reflection phases

do not arrive in the wave train of the data being modelled. Frequently data is clipped prior to the arrival of core reflections to ensure that such unmodelled 'noise' does not interfere with the inversion process. When the rupture duration is of the order of 10 minutes as was the case for the Sumatra-Andaman Islands earthquake [34] core reflections from the early part of the rupture arrive at stations at the same time as the direct waves from latter portions of the rupture making it impossible to accurately model the body waves. A novel approach termed 'back-projection' used a Japanese array to map the position of the portion of the fault that was slipping at any time to image the rupture front as the earthquake progressed [35]. The method stacks data from a large number of stations close to each other in an array. By comparing time-shifts required to stack the arrivals correctly across the array to the theoretical travel times between points on the earthquake fault and stations in the array, it is possible to determine the origin of the direct wave at any time in the earthquake history. Although this method does not give a slip distribution directly it can be used to as a constraint in other methods and as a method of determining the rupture speed of an earthquake and has been used to determine the rupture velocity of several subsequent earthquakes [36,37,38].

5. Theory of supershear earthquakes

One of the assumptions that is fed into many of the earthquake inversions is that earthquakes rupture at a certain velocity. This assumption stems from theoretical work on the mechanics of fractures. The accepted wisdom for a long time was that ruptures accelerate from rest to a terminal velocity close to the Rayleigh speed [39]. Theoretical work on fracture dynamics actually showed that there were two possible stable rupture velocities, either close to the Rayleigh wave speed or in excess of the shear wave velocity [40]. Owing to the fact that ruptures at the shear wave velocity are not stable, it was widely assumed that this acted as a 'barrier' to the velocity of the accelerating crack tip hence no earthquake could propagate in the supershear rupture regime. It has, however, been shown that in-plane faults modelled in 2 dimensions can propagate at supershear velocities courtesy of an interesting phenomenom whereby the rupture front literally jumps from the Rayleigh wave velocity to the supershear regime [39]. In this case two rupture fronts co-exist for a short period of time before the initial Rayleigh wave rupture dies out to leave a single rupture propagating in the supershear regime. There is currently no known analytical approach to determination of the rupture propagation in 3 dimensions hence all studies to date have utilised numerical simulation. The question of supershear rupture in 3 dimensions has been

addressed [11] and it has been shown that supershear rupture can occur under certain conditions. The process in moving from Rayleigh to supershear velocities in 3 dimensions is different from 2 dimensions; instead of the rupture front 'jumping' ahead, the rupture becomes initially unstable in the in-plane direction before this instability propagates laterally along the rupture front. This means that an initially circular rupture becomes elongated along the in-plane direction eventually forming 'ears' on the edge of the rupture. This results in no discontinuity in the rupture front with consequently no region where slip is temporarily arrested between the passing of the supershear and Rayleigh rupture fronts in contrast to the two dimensional studies undertaken. Figure 4 reproduces figures 3 of Madariaga et al (2000) [11] which illustrates this.

Further work in three dimensions has shown that for this transition between Rayleigh and supershear rupture velocities occurs when S (as defined in eqn (2)) ≤ 1.19 [41], much lower than the 1.77 as determined for the two-dimensional case [42].

6. Laboratory studies

There has been considerable work in attempting to verify the theory of rupture velocity and the possibility of supershear ruptures taking place, as studies into both earthquakes and the fundamental processes involved in the fracture of materials. Real material studies have crack tip speeds that seldom exceed 40-50% of the Rayleigh velocity [43,44] as opposed to the theoretical crack velocities approaching the Rayleigh wave velocity. To explain this discrepancy between theory and laboratory studies a variety of explanations have been invoked such as high strain and micro-damage zones around the crack tip [44,45] or non-linear crack paths [45]. In plane shear intersonic crack growth was initially reported in homalite with a pre-existing fault plane [47]. This experimental set-up was designed to resemble the case where an earthquake ruptures a pre-existing fault plane in a rock unit, with the use of homalite serving to allow a high-speed camera to photograph the rupture as it progresses. Measurements of the position of the crack tip with time and the angle of the shear shock waves visible in the stress field allowed a determination of the rupture velocity, both independently confirming a steady-state velocity of $\sim\sqrt{2}\,v_s$ where v_s is the shear wave velocity of the homalite. This work was then extended to show the transition between the sub-Rayleigh and supershear rupture regimes [48]. Variation in the angle and magnitude of the far-field pressure exerted on a pre-existing fault in homalite allowed investigation of the sub- to supershear transition. For lower pressures

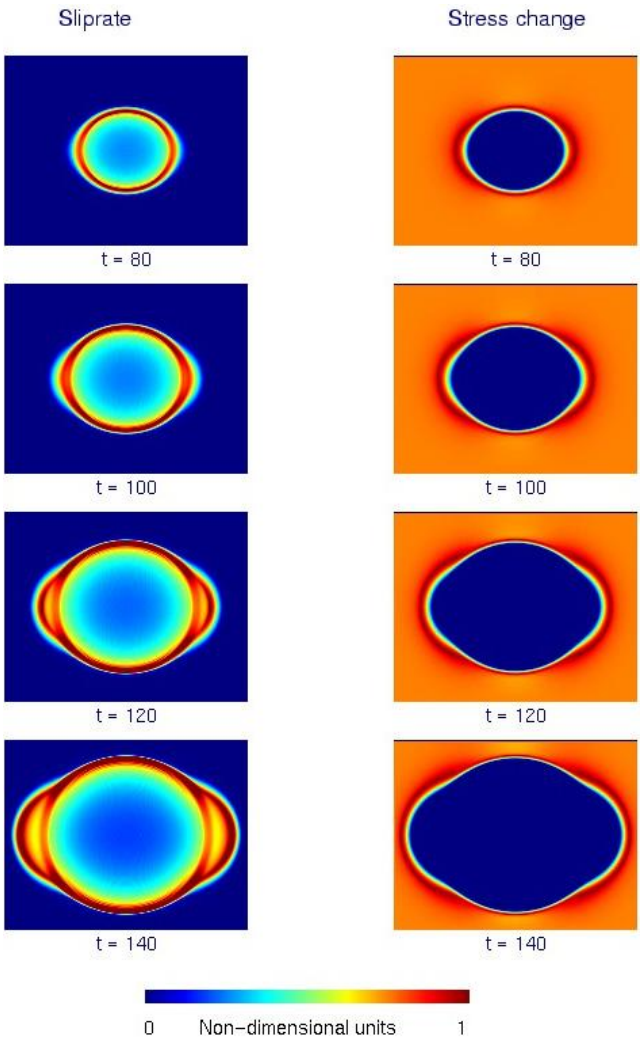


Figure 4. Rupture growth on a flat uniform fault embedded in a homogeneous elastic medium reproduced with permission from Madariaga et al (2000) [11]. Rupture starts from a finite initial asperity and then grows at subsonic speed in all directions. After a while, rupture along the inplane direction (horizontal) jumps at a speed that is faster than the shear wave velocity. Snapshots show the sliprate (left) and associated stress change (right) at four successive instants during rupture. Slightly after time =100, rupture jumps from the subshear to supershear regime in the inplane direction.

and angles, the rupture progresses at a velocity close to the Rayleigh wave whereas for higher values of pressure rupture nucleated and progressed at supershear velocities. At intermediate values, a transition was observed whereby the rupture initiated at Rayleigh wave velocities and then 'jumped' to a supershear regime. This has been likened to the theoretical Burridge-Andrews mechanism [39]. Most recently this work has been further extended by considering the effect of having material with different bulk properties on either side of the fault [49]. By using homalite and polycarbonate plates on either side of a fault orientated at a given angle to far-field pressure it is possible to see ruptures progress in different manners in each direction. Depending on the pressure and angle used (see figure 5), three distinct rupture characteristics were observed:

1. *Bilateral rupture progressing at velocities below the shear wave velocity of the slower material:* The rupture velocity was higher in the direction where the fault was angled such that the direction of slip of the lower wave speed material is positive (rupture propagating to the left in figure 5 is faster). The faster rupture propagates at a velocity close to the generalised Rayleigh wave speed of the system, the slower rupture progresses at a speed lower than the Rayleigh wave speed in the slower material.
2. *Bilateral rupture with one rupture progressing at the generalised Rayleigh wave velocity, the other rupture progressing at a velocity in excess of the shear wave velocity of both materials, close to, but below, the longitudinal wave velocity of the slower material:* The velocity of the rupture progressing in the direction of positive slip of the slower material is close to the generalised Rayleigh wave velocity as in case (1). However, the supershear rupture occurs in the direction that exhibited slower rupture velocities in case (1) (rupture propagating to the right in figure 5 is supershear).
3. *The mode of rupture transitions from case (1) to case (2):* That is to say the rupture in the direction of positive slip of the slow material stays at a constant velocity close to the generalised Rayleigh wave velocity whereas the rupture in the direction of negative slip of the slow material is initially less than the Rayleigh wave speed of the slower material but then jumps to a velocity in excess of both shear wave velocities and slightly less than the longitudinal wave velocity of the slower material. This behaviour is similar to that observed for faults in a single material with intermediate pressure and angle discussed previously [48].

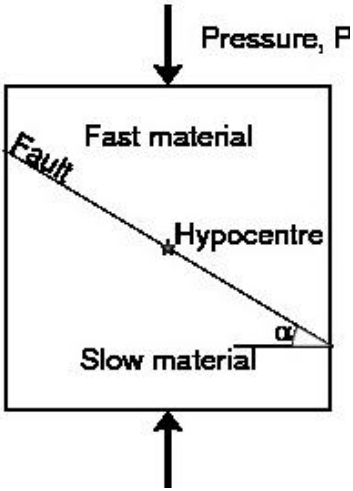


Figure 5. Experimental set-up used to image different rupture speeds in different directions along a fault between two materials with differing properties. The crack is triggered at the hypocenter using an exploding wire.

The experiments show that a wide range of possible rupture processes can occur depending on the magnitude and angle of the far-field pressure. By careful observation of earthquakes it may be possible to determine whether supershear ruptures have occurred. The combination of theoretical work with laboratory experiment has highlighted the conditions that supershear rupture can occur under, by observing if and where supershear earthquakes occur it may be possible to quantify some of the boundary forces at play.

7. Observation of supershear earthquakes

Much of the work discussed in the previous section was carried out as a result of observations of supershear earthquakes. The initial theoretical work suggested that although ruptures might be stable at supershear velocities, the fact that they were not stable at the shear wave velocity meant that many perceived this barrier would prevent ruptures from reaching supershear velocities. Observation of earthquake where models with only subshear rupture did not fully explain all the observations led to original theoretical work being revisited and to experiments being carried out to investigate the sub- to supershear transition. It is this iterative approach to understanding where an observation leads to previous work being re-examined with a

different emphasis placed on the original research that has characterised earthquake seismology in the past decade. This paper focuses mostly on the debate surrounding supershear earthquakes as this has been one of the most interesting developments in the field recently, although with the great quantity of high quality data that is now available relating to recent earthquakes, it is likely that other major advances will be made in the near future.

Although supershear earthquake ruptures had been predicted during the early days of fracture modeling [40,18,13] there were very few early observations of supershear earthquakes. The only observation of supershear rupture in the literature in the 20th Century was associated with the Imperial valley earthquake of 1979 [5,50]. However, even for this earthquake, there was not a consensus that rupture was supershear, with many solutions proposed that did not require a supershear portion of rupture [22,51,6]. This highlights the problems associated with using observations of earthquake ruptures to determine which set of assumptions is the most suitable to apply to an earthquake. The question of whether the Imperial Valley earthquake ruptured at supershear velocities for some portion of its rupture history is not unambiguously determined by the investigations of the time. Consequently, using the shear wave velocity as a maximum rupture velocity was still seen as a reasonable constraint on earthquake inversions by many researchers. The occurrence of several large strike-slip earthquakes in 1999 and the early years of the 21st Century (Figure 6) has subsequently challenged this.

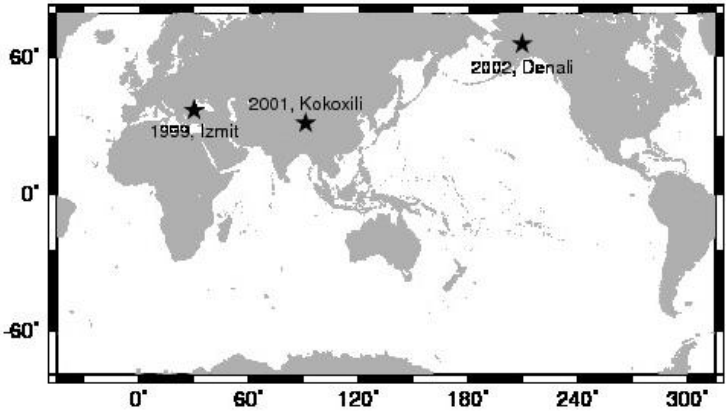


Figure 6. Map of the world showing locations of earthquakes with inferred supershear ruptures to have occurred around the turn of the Century discussed in the text.

The first of the large strike-slip earthquakes in this sequence is the 17th August 1999 Izmit earthquake. This M_w 7.6 event was the latest earthquake in the sequence of the 'unzipping' of the North Anatolian Fault [52]. Two strong motion recordings were made near the earthquake fault, one to the East of the earthquake epicentre, the other to its West [53]. These two recording were very different; the recording to the West had a 'typical' character with well separated P and S wave arrivals and a peak ground acceleration of 0.21g. To the East, the recording had no clear separation of P and S waves and a much higher peak ground acceleration of 0.41g. The conclusion drawn is that rupture to the East was supershear velocity while rupture westwards was a more 'usual' sub-shear rupture. This behaviour is the same as that seen in experiments with rupture along a fault between materials with different properties [49]. To be completely analogous with the experimental data, the seismic velocity of material to the North of the North Anatolian fault in this area must be higher than to the south of the fault. This hypothesis has yet to be fully tested.

This observation, along with the more uncertain observation of supershear rupture for the Imperial Valley earthquake [5,50] led to a re-assessment of the assumptions present in many of the inversions carried out. Strike-slip earthquakes could no longer be assumed to rupture at sub-shear velocities. Further weight was added to this argument by the 3rd November 2002, Denali, earthquake. The Denali earthquake was M_w 7.8 and initiated on the previously unrecognised Susitna Glacier thrust fault and propagated west for close to 50 km. Rupture then proceeded on the McKinley strand of the Denali fault for ~70 km before reaching the main strand of the Denali fault on which it propagated for a further 156 km before stepping across a ~20 km transition zone and propagating for a further ~70 km on the Totschunda fault [54]. The total mapped surface rupture length is 341 km [54,55]. Despite the Denali fault's remote location, earthquakes on it have the potential to have significant economic impact as the trans-Alaskan oil pipeline runs across it. Consequently, there were a number of local strong-motion local instruments recording events as the earthquake occurred. Although some early inversions produced models with sub-shear ruptures [56,57], inversions using local strong-motion data [58,59] and more sophistcated subsequent techniques [37] have shown supershear rupture occurred, transitioning from a sub- to supershear rupture regime about 120 km after earthquake initiation.

Twelve months previously the Kokoxili earthquake occurred. This earthquake broke the Kunlun fault and was to be the first earthquake with supershear rupture observed in teleseismic data. In contrast to the Denali earthquake, in this remote area there were no local strong-motion seismometers deployed. The Kokoxili earthquake (also termed the Kunlun or

Kunlunshan earthquake) occurred on November 14, 2001 rupturing with M_w 7.8, breaking a ~400 km long section of the Kunlun fault in Tibet [60]. Rupture initiated on an extensional step-over before transferring to the left-lateral Kunlun fault, rupturing the fault unilaterally eastwards [56]. This earthquake has the largest measured epicentre-centroid location shift of any strike-slip earthquake in the global CMT catalogue [56]. The first inference of supershear rupture for this earthquake came from teleseismic surface wave data [30]. As previously discussed, teleseismic surface waves can only resolve details on a broad scale due to their dominant frequencies. Nevertheless, due to the Kokoxili earthquake's very long rupture length, it was possible to split the fault into 4, 100 km long sections and show that rupture occurred at an average velocity in excess of the local supershear speed in all but the first section of the fault although the rupture velocity is not well resolved in the final section [30]. Early studies of body wave data settled on a rupture with an average velocity very close to that of the local shear wave velocity of 3.5 km/s with studies reporting 3.4 km/s [61,56] and 3.6 km/s [62]. All the early body-wave studies only reported average rupture velocity and used either predominantly [61,62] or exclusively [56] P wave records. P wave records are less sensitive to variation in rupture velocity due to their higher speed relative to S waves. A study looking at only S wave body waves was carried out and found that although the early portions of rupture were sub-shear, after approximately 120 km of rupture, the rupture velocity jumped abruptly to supershear velocities [63]. The supershear rupture continued for at least 150 km at which point the rupture front became poorly resolved. This observation was verified [37] using the method of P wave back-projection [35]. Further observations of the origin of the high frequency components of Rayleigh waves due to the earthquake using an array at regional distances show that they originated from regions of the fault where rapid transitions between sub- and supershear ruptures occurred [64]. This is expected from theoretical and numerical work on the transition [65,66,67]. Field observations reveal unusual off-fault cracking [68] in precisely the region of the high rupture velocity imaged in other studies [63,64,37]. Investigation of the theoretical stress field induced by a passing supershear rupture front indicates that the area in which the cracks are observed would have been subjected to considerable transient tensional stress if supershear rupture occurred as proposed [68]. The Denali and Kokoxili earthquakes have remarkable similarities in their geographical expression and the style in which they ruptured. Another, less obvious, similarity is that although the crustal thickness is very different in Alaska to the double thickness crust observed in Tibet, the seismogenic zone thickness is similar in both meaning that the earthquakes had similar fault widths of ~15-20 km.

The debate surrounding the rupture velocity of the Kokoxili earthquake highlights the major problem that exists is earthquake modelling mentioned in section (3); namely how good does a solution have to be to be considered 'good enough'. If the work on the Izmit earthquake had not been combined with more uncertain models from other earthquakes such as the Imperial Valley earthquake and experimental work to create a more convincing set of evidence, then the models that have fast, but critically sub-shear, rupture velocities might be considered sufficient. The growing consensus that supershear ruptures can and do occur is down to a combination of increased data quantity and quality, theoretical work on fracture mechanics and experimental work combined with several earthquakes.

8. Discussion

The spate of large supershear earthquake reported in the scientific literature of the early years of the 21st Century is intriguing. The first question that arises is whether these earthquakes are unusual, or whether supershear earthquakes occur regularly and only recently have advances in data and techniques allowed their observation. The problem with the classical rupture inversion techniques used since the pioneering work of the early 1980s [5,6] is that the seismic radiation is a function of not only the slip on the fault but also the rise-time and rupture velocity. Many inversions attempt to deduce the slip as this can then be compared to observations on the ground via surface measured slips or at depth via GPS or InSAR with the necessary caveats on timing of slip and how measurements on the free surface reflect the nature of the fault at depth. The argument that the majority of earthquakes have a rupture velocity of approximately 80% of the local shear wave velocity [69] has been used as justification in many studies constraining or fixing the value of rupture velocity. It has been argued that, where possible, the rupture velocity should be separately inverted for [64] prior to determination of the slip distribution. There are methods available for the direct determination of rupture velocity that use the presence of both local [70] and teleseismic [35] arrays. The recent earthquakes that exhibit supershear ruptures were only identified following very careful analysis. Frequently early analyses produced models with no supershear rupture portion, and it was only once all the available data was analysed that supershear rupture became evident. The concern is that more minor earthquakes are not generally subjected to the raft of modelling techniques and hence models might have been produced that would not stand up to a high degree of scrutiny.

The advances made in earthquake modelling over the past 30 years have been as a result of two major events. Firstly, there was the technological breakthrough of digital seismometers and the development of computing to allow the discretisation and investigation of the inverse problem. This led to the raft of inversion techniques developed in the early 1980s and refined in the following years. The second major advance is occurring at the present time. The unusually large number of large earthquakes coupled with the increase in number and quality of seismic stations means that there in an unprecedented degree of seismic data available to investigate. This combined with improvements in measuring the static displacements due to earthquakes and in the interseismic period has led to the development of new techniques and the questioning of many previously produced seismic models. This paper has focussed on supershear earthquakes; the story which surrounds the evolution of thinking on the subject has mirrored the general progress in the field of seismology, an evolution which looks set to continue in the coming years.

9. Future work

Although it is inherently impossible to predict what advances will be made in any field of research, it is possible to predict what method is likely to give the best route to proceed down. Theoretical work on supershear ruptures is progressing and it is likely that the next step will be to incorporate the possibility of supershear ruptures into existing earthquake hazard assessments. This will require considerable work to accurately determine the likely ground response to the passage of a supershear rupture as well as an accurate assessment of the necessary conditions to allow supershear rupture to occur. It has already been noted that all of the supershear ruptures that have so far been observed occurred on remarkably straight sections of fault [71]. Experimental lab work has suggested that the angle and magnitude of the controlling stress field plays an important role in the sub- to supershear rupture transition [47,48] and further work scaling this observation up to earthquakes is required. Once the probability of a supershear rupture occurring on a given fault is estimated and this information incorporated into likely ground acceleration models, it may be that building codes in some regions of the world will require updating.

Early work by seismologists focussed on explaining short segments of the earthquake record, be it direct P, S, or surface waves. Modern techniques frequently utilise much more of the spectrum. The problem in doing this is that Greens' functions that form the basis of the matrix equation (equation (6)) become increasingly complex the more phases are considered.

Inaccuracies in the Greens' function will lead to errors in any inversion produced in a non-linear fashion. The use of small earthquakes as 'Empirical Greens Functions' (EGFs) of larger earthquakes has long been discussed. By definition, a small earthquake in the same location with the same mechanism as a larger earthquake will contain all of the same phases as the trace relating to the larger earthquake, hence errors in the Greens function matrix can theoretically be reduced to near zero. This idea has been around for some time [72], and applied to some of the early seismic source analysis [73]. The approach has been developed by a number of authors [74,75,76,77,31,78,79,80,81] in the intervening years but has mainly been restricted to investigations of surface waves. Body wave recordings are sufficiently complex that several EGF are required at different depths on the subduction interface to model a typical great subduction zone earthquake. The chances of a subduction zone earthquake having sufficient aftershocks with the correct mechanism and distribution over the earthquake fault plane is low, however, with the improving standard of data recording and as the total number of earthquakes recorded increases with time, the chances of finding suitable EGFs in the historical catalogue increases. It is likely that EGFs will soon be used to model the full trace of a large earthquake including body and surface waves although how much more this will tell us about the earthquake rupture process than traditional techniques remains to be seen.

Historically there have been several examples of earthquakes with mapped surface ruptures that indicate rupture on multiple fault planes, the 1927 Tango, Japan, earthquake being one famous example [82] and the Landers earthquake of 1992 being a more recent example [83]. When mapping of the surface rupture is possible it is easy to use this as a constraint when inverting for the seismic history of the earthquake and allow slip to occur on multiple appropriate faults. However, the majority of earthquakes do not have mapped surface ruptures. Much progress has been made in recent years on using different techniques to determine the fault geometry by one method (for example Interferometric Synthetic Appature Radar, InSAR) and then use that geometry in the inversion of seismic data. Equally it has been demonstrated that careful analysis of high quality seismic data can resolve simultaneous rupture on multiple faults in certain circumstances in the absence of other geodetic data [84,85,38]. It is in the combination of multiple data sets to produce a composite solution [38] that the immediate future lies. The technique of back projection of P waves [35] dovetails nicely with traditional kinematic inversion techniques, in that it provides a constraint to the earthquake rupture front, and it is expected that more studies combining back-projection with inversion of body and surface waves are likely in the coming years. Recent work has also focussed on combining geodetic

measurements with seismic data in a simultaneous inversion [86,87], building on earlier work [88]. The problem with this is that many geodetic measurements may include movement on a fault or faults that occurred after the seismic slip. For example, InSAR data may include a portion of afterslip on a fault in the immediate aftermath of an earthquake. It is likely that the magnitude of this afterslip can be significant. It may be that the slip determined due to geodetic inversions acts as an upper bound to a subsequent seismic inversion (i.e. seismic slip in any cell can be 0-100% of the geodetic slip in that cell), exactly how to paramaterise this appropriately while simultaneously inverting for geodetic and seismic slip is a question that is yet to be answered. If these questions can be fully answered and then the answers applied to the datasets that currently exist and are continually being created as more earthquakes occur then they will doubtless shed great light on the fundamental mechanisms by which earthquakes occur.

10. Acknowledgements

This work has been carried with funding from Natural Environment Research Council grant NE/C518806/1.

11. References

1. T. Lay, H. Kanamori, C. J. Ammon, M. Nettles, S. N. Ward, R. C. Aster, S. L. Beck, M. R. Brudzinski, R. Butler, H. R. DeShon, G. Ekström, K. Satake, S. Sipkin, 'The great Sumatra-Andaman earthquake of 26 December 2004', *Science*, 308, 1127-1133 (2005).
2. C. J. Ammon, C. Ji, H-K. Thio, D. P. Robinson, S. Ni, V.Hjorleifsdottir, H. Kanamori, T. Lay, S. Das, D. Helmberger, G. Ichinose, J. Polet, D. Wald, 'Rupture Process of the 2004 Sumatra-Andaman earthquake', *Science*, 308, 1133-1139 (2005).
3. J. Liu-Zeng, Z. Zhang, L. Wen, P. Tapponnier, J. Sun, X. Xing, G. Hu, Q. Xu, L. Zeng, L. Ding, K. W. Hudnut, J. Van der Woerd, Co-seismic ruptures of the 12 May 2008, M_s 8.0 Wenchuan earthquake, Sichuan: East-west crustal shortening on oblique, parallel thrusts along the eastern edge of Tibet', *Earth Planet. Sci. Lett.*, 286 (3-4), 355-370 (2009).
4. S. Das, B. V. Kostrov, 'Inversion for seismic slip rate history and distribution with stabilising constraints: Application to the 1986 Andreanof Islands earthquake', *J. Geophys. Res.*, 95, 6899-6913 (1990).
5. A. H. Olsen, R. J. Aspel, 'Finite faults and inverse theory with applications to the 1979 Imperial Valley earthquake', *Bull. Seism. Soc. Am.*, 72 (6A), 1969-2001 (1982).

6. S. H. Hartzell, T. H. Heaton, 'Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake', *Bull. Seism. Soc. Am.,* 73 (6A), 1553-1583 (1983).
7. M. Kikuchi, H. Kanamori, 'Inversion of complex body waves', *Bull. Seism. Soc. Am.,* 72 (2), 491-506 (1982).
8. J. F. Pacheo, L. R. Sykes, 'Seismic moment catalog of large shallow earthquakes, 1900 to 1989', *Sull. Seismol. Soc. Am.,* 82, 1306-1349 (1992).
9. Global CMT solutions can be accessed at: http://www.globalcmt.org.
10. Y. Ida, 'Cohesive force across the tip of a longitudinal-shear crack and Griffith's specific surface energy', *J. Geophys. Res.,* 77, 3796-3805 (1972).
11. R. Madariaga, S. Peyrat, K. B. Olsen, 'Rupture dynamics in 3D: a review', in *Problems in Geophysics for the New Millenium*, eds. E. Boschi, G. Ekström, A. Morelli; Instituto Nazionale di Geofisica e Vulcanologia, Roma (2000).
12. D. J. Andrews, 'Rupture propagation with finite stress in antiplane strain', *J. Geophys. Res.,* 81, 3575-3582 (1976).
13. D. J. Andrews, 'Rupture velocity of plane strain shear cracks', *J. Geophys Res.,* 81, 5679-5687 (1976).
14. S. Das 'Three-dimensional spontaneous rupture propagation and implications for the earthquake source mechanism' *Geophys. J. R. Astron. Soc,* 67, 375-393 (1981).
15. S. M. Day, 'Three-dimensional simulation of spontaneous rupture: The effect of nonuniform prestress', *Bull. Seism. Soc. Am.,* 72, 1881-1902 (1982).
16. S. Das, B. V. Kostrov, 'Breaking of a single asperity: Rupture process and seismic radiation', *J. Geophys. Res.,* 88 (B5), 4277-4288 (1983).
17. B. V. Kostrov, 'On the crack propagation with variable velocity', *J. Appl. Math. Mech.,* 30, 1241-1248 (1966).
18. S. Das, K. Aki, 'A numerical study of two-dimensional spontaneous rupture propagation' *Geophys. J. R. Astron. Soc,* 50, 643-668 (1977).
19. S. Das, K. Aki, 'Fault plane with barriers: A versatile earthquake model', *J. Geophys Res*., 82, 5658-5670 (1977).
20. A. M. Dziewonski, J. H. Woodhouse, 'Studies of the seismic source using normal-mode theory', in *Earthquakes: Observation, Theory and Interpretation*, eds H. Kanamori, E. Boschi, *Proc. Int. Sch. Phys. Enrico Fermi*, 85, 45-137 (1983).
21. C. Henry, 'Teleseismic studies of large submarine earthquakes', *DPhil. Thesis, University of Oxford*, (2002).
22. S. H. Hartzell, D. V. Helmberger, ,Strong motion modeling of the Imperial Valley earthquake of 1979', *Bull. Seism. Soc. Am.,* 72 (2), 571-596 (1982).
23. D. J. Doornbos, 'Seismic moment tensors and kinematic source parameters' *Geophys. J. R. Astron. Soc*., 69 (1) 235-251 (1982).
24. D. Massonnet, K. L. Feigl, 'Radar interferometry and its application to changes in the Earth's surface', *Rev. Geophys.,* 36, 441-500 (1998).
25. R. J. Le Bras 'Methods of multiparameter inversion of seismic data using the acoustic and elastic Born approximations', *PhD Thesis, California Institute of Technology*. (1985).

26. P. Y. Hwang, H. Kanamori, 'Of the May 7, 1986 Andreanof Islands earthquake source parameters', *Geophys.Res. Lett.,* 13, 1426-1429 (1986).
27. G. A. Ekström, 'A broad band method of earthquake analysis', *PhD. Thesis, Harvard University* (1987).
28. T. M Boyd, J. L. Nábělek, 'Rupture process of the Andreanof Islands earthquake of May 7, 1986', *Bull. Seism. Soc. Am.,* 78, 1653-1673 (1988).
29. H. Houston, E. R. Engdahl, 'A comparison of the spatio-temporal distribution of moment release for the 1986 Andreanof Islands earthquake', *Geophys. Res. Lett.,* 16, 1421-1424 (1989).
30. M. Bouchon, M. Vallée, 'Observation of long supershear rupture during the magnitude 8.1 Kunlunshan earthquake', *Science,* 301, 824-826 (2003).
31. A. A. Velasco, C. J. Ammon, T. Lay, 'Empirical green function deconvolution of broadband surface waves: Rupture directivity of the 1992 Landers, California (M_w = 7.3), earthquake', *Bull. Seism. Soc. Am.,* 84, 735-750 (1994).
32. S. Das, B. V. Kostrov, 'Diversity of solutions of the problem of earthquake faulting inversion. Application to SH waves for the great 1989 Macquarie Ridge earthquake', *Phys. Earth Planet. Inter.,* 85, 293-318 (1994).
33. C. Scholz, *'The mechanics of earthquakes and faulting',* Cambridge University Press, New York (1990).
34. R. Bilham, 'A flying start, then slow slip', *Science,* 308, 1126-1127 (2005).
35. M. Ishii, P. M. Shearer, H. Houston, J. E. Vidale, 'Extent, duration and speed of the 2004 Sumatra–Andaman earthquake imaged by the Hi-Net array', *Nature,* 435, 933-936 (2005).
36. K. T. Walker, M. Ishii, P. M. Shearer, 'Rupture details of the 28 March 2005 Sumatra M_w 8.6 earthquake imaged with teleseismic P waves', *Geophys Res Lett.,* 32, L24303, doi: 10.1029/2005GL024395 (2005).
37. K. T. Walker, P. M. Shearer, 'Illuminating the near-sonic rupture velocities of the intracontinental Kokoxili M_w 7.8 and Denali fault M_w 7.9 strike-slip earthquakes with global P wave back projection imaging', *J. Geophys. Res.,* 114, B02304, doi: 10.1029/2008JB005738 (2009).
38. T. Lay, C. J. Ammon, H. Kanamori, L. Rivera, K. D. Koper, A. R. Hutko, 'The 2009 Samoa-Tonga great earthquake triggered doublet', *Nature,* 466, 964-968 (2010).
39. L. B. Freund, '*Dynamic fracture mechanics*', Cambridge University Press, Cambridge, UK (1990).
40. R. Burridge, 'Admissible speeds for plane-strain shear cracks with friction but lacking cohesion', *Geophys. J. R. Astr. Soc.,* 35, 439-455 (1973).
41. E. M. Dunham, 'Conditions governing the occurrence of supershear ruptures under slip-weakening friction', *j. Geophys. Res.,* 112 B07302, doi: 10.1029/2006JB0044717 (2007).
42. D. J. Andrews, 'Dynamic plane-strain shear rupture with a slip-weakening friction law calculated by a boundary integral method', *Bull. Seism. Soc. Am.,* 75(1), 1-21 (1985).
43. B. Cotterell, 'Velocity effects in fracture propagation.' *Appl. Mater. Res.,* 4, 227-232 (1965).

44. K. Ravi-chandar, W. G. Knauss, 'An experimental investigation into dynamic fracture: III. On steady-state crack propagation and crack branching', *Int J. Fract.,* 26, 141-154 (1984).
45. K. B. Broberg, 'On the speed of a brittle crack', *J. appl. Mech.,* 31, 546-547 (1964).
46. H. Gao, 'Surface roughening and branching instabilities in dynamic fracture', *J. Mech. Phys. Solids.,* 1, 457-486 (1993).
47. A. J. Rosakis, O. Samudrala, D. Coker, 'Cracks faster than the shear wave speed', *Science,* 284, 1337-1340 (1999).
48. K. Xia, A. J. Rosakis, H. Kanamori, 'Laboratory earthquakes: The sub-Rayleigh-to-supershear rupture transition', *Science,* 303, 1859-1861 (2004).
49. K. Xia, A. J. Rosakis, H. Kanamori, J. R. Rice, 'Laboratory earthquakes along inhomogeneous faults: Directionality and supershear', *Science,* 308, 681-684 (2005).
50. R. J. Archuleta, 'A faulting model for the 1979 Imperial Valley earthquake', *J. Geophys. Res.,* 89 (B6), 4559-4585 (1984).
51. R. J. Archuleta, 'Analysis of near-source static and dynamic measurements from the 1979 Imperial Valley earthquake', *Bull. Seism. Soc. Am.,* 72 (6), 1927-1956 (1982).
52. N. N. Ambraseys, J. A. Jackson, 'Seismicity of the Sea of Marmara (Turkey) since 1500', *Geophys. J. Int.,* 141 (3), F1-F6 (2000).
53. M. Bouchon, M.-P. Bouin, H. Karabulut, M. Nafi Toksöz, M. Dietrich., A. J. Rosakis, 'How fast is rupture during an earthquake? New insights from the 1999 Turkey earthquakes', *Geophys. Res. Lett.,* 28 (14), 2723-2726 (2001).
54. P. J. Haeussler, D. P. Schwartz, T. E. Dawson, H. D. Stenner, J. J. Lienkaemper, B. Sherrod, F. R. Cinti, P. Montone, P. A. Craw, A. J. Crone, S. F. Personius, 'Surface rupture and slip distribution of the Denali and Totschunda faults in the 3 November 2002 M 7.9 earthquake, Alaska', *Bull. Seismol. Soc. Am.,* 94 (6B), S23-52 (2004).
55. D. Eberhart-Phillips, P. J. Haeussler, J. T. Freymueller, A. D. Frankel, C. M. Rubin, P. Craw, N. A. Ratchkovski, G. Anderson, G. A. Carver, A. J. Crone, T. E. Dawson, H. Fletcher, R. Hansen, E. L. Harp, R. A. Harris, D. P. Hill, S. Hreinsdóttir, R. W. Jibson, L. M. Jones, R. Kayen, D. K. Keefer, C. F. Larsen, S. C. Moran, S. F. Personius, G. Plafker, B. Sherrod, K. Sieh, N. Sitar, W. K. Wallace, 'The 2002 Denali Fault Earthquake, Alaska: A Large Magnitude, Slip-Partitioned Event', *Science,* 300, 1113-1118 (2003).
56. A. A. Ozacar, S. L. Beck, 'The 2002 Denali fault and the 2001 Kunlun fault earthquakes: Complex rupture processes of two large strike-slip earthquakes', *Sull. Seismol. Soc. Am.,* 94, S278-S292 (2004).
57. D. S. Dreger, D. D. Oglesby, R. Harris, N. Ratchkovski, R. Hansen, 'Kinematic and dynamic rupture models of the November 3, 2002 M_w 7.9 Denali, Alaska, earthquake', *Geophys. Res. Lett.,* 31, L04605, doi: 10.1029/2003GL018333 (2004).

58. E. M. Dunham, R. J. Archuleta, 'Evidence for a supershear transient during the 2002 Denali fault earthquake', *Bull. Seismol. Soc. Am.,* 94(6B), S256-S268 (2004).
59. A. Frankel, 'Rupture process of the M 7.9 Denali fault, Alaska, earthquake: Subevents, directivity, and scaling of high-frequency ground motions', *Bull. Seismol. Soc. Am.,* 94(6B), S234-S255 (2004).
60. X. Xu, W. Chen, W. Ma, G. Yu, G. Chen, 'Surface Rupture of the Kunlunshan Earthquake (M_s 8.1), Northern Tibetan Plateau, China', *Seismol. Res. Lett.,* 73, 884-892 (2002).
61. A. Lin, M. Kikuchi, B. Fu, 'Rupture segmentation and process of the 2001 M_w 7.8 central Kunlun, China, earthquake', *Bull. Seismol. Soc. Am.,* 93, 2477-2492 (2003).
62. M. Antolik, R. E. Abercrombie, G. Ekström, 'The 14 November 2001 Kokoxili (Kunlunshan), Tibet, earthquake: Rupture transfer through a large extensional step-over', *Bull. Seismol. Soc. Am.,* 94, 1173-1194 (2004).
63. D. P. Robinson, C. Brough, S. Das, 'The M_w 7.8, 2001 Kunlunshan earthquake: Extreme rupture speed variability and effect of fault geometry', *J. Geophys. Res.,* 111, B08303, doi: 10.1029/2005JB004137 (2006).
64. M. Vallée, M Landès, N. M. Shapiro, Y. Klinger, 'The 14 November 2001 Kokoxili (Tibet) eartyhquake: High-frequency seismic radiation originating from the transitions between sub-Rayleigh and supeshear ruputre velocity regimes', *J. Geophys. Res.,* 113, B07305, doi: 10.1029/2007JB005520 (2008).
65. R. Madariaga, 'High-frequency radiation from crack (stress drop) models of earthquake faulting', *Geophys. J. R. Astrom. Soc.,* 51, 625-651 (1977).
66. M. Campillo, 'Numerical evaluation of the near-field high-frequency radiation from quasidynamic circular faults', *Bull. Seismol. Soc. Am.,* 73, 723-734 (1983).
67. T. Sato, 'Seismic radiation from circular cracks growing at variable rupture velocity', *Bull. Seismol. Soc. Am.,* 84, 1199-1215 (1994).
68. H. S. Bhat, R. Dmowska, G. C. P. King, Y. Klinger, J. R. Rice, 'Off-fault damage patterns due to supershear ruptures with application to the 2001 M_w 8.1 Kokoxili (Kunlun) Tibet earthquake', *J. Geophys Res.,* 112, B06301, doi: 10.1029/2006JB04425 (2007).
69. T. H. Heaton, 'Evidence for and implications of self-healing pulses of slip in earthquake rupture', *Phys. Earth. Planet. Inter.,* 64, 1-20 (1990).
70. P. Spudich, E. Cranswick, 'Direct observation of rupture propagation during the 1979 Imperial Valley earthquake using a short baseline accelerometer array', *Bull. Seismol. Soc. Am.,* 74, 2083-2114 (1984).
71. D. P. Robinson, S. Das, M. P. Searle, 'Earthquake fault superhighways', *Tectonophysics,* 493 (3-4), 236-243 (2010).
72. K. Aki, 'Scaling law of seismic spectrum', *J. Geophys. Res.,* 72, 1217-1231 (1967).
73. S. H. Hartzell, 'Earthquake aftershocks as Green's functions', *Geophys. Res. Lett.,* 5, 1-4 (1978).
74. C. S. Mueller, 'Source pulse enhancement by deconvolution of an empirical Green's function', *Geophys. Res. Lett.,* 12, 33-36 (1985).

75. E. Fukuyama, K. Irikura, 'Rupture process of the 1983 Japan Sea (Akita-Oki) earthquake using a waveform inversion method', *Bull. Seismol. Soc. Am.,* 76, 1623-1640 (1986).
76. J. Mori, A. Frankel, 'Source parameters for small events associated with the 1986 North Palm Springs, California, determined using empirical Green's functions', *Bull. Seismol. Soc. Am.,* 80, 278-295 (1990).
77. C. J. Ammon, A. A. Velasco, T. Lay, 'Rapid estimation of rupture directivity: Application to the 1992 Landers (M_s = 7.4) and Cape Mendocino (M_s = 7.2), California earthquakes', *Geophys. Res. Lett.,* 20, 97-100. (1993).
78. P. F. Ihmlé, 'Frequency dependent relocations of the 1992 Nicaragua slow earthquake: An empirical Green's function approach', *Geophys. J. Int.,* 127, 75-85 (1996).
79. F. Courboulex, J. Virieux, A. Deschamps, D. Gilbert, A. Zollo, 'Source investigation of a small event using empirical Green's functions and simulated annealing', *Geophys. J. Int.,* 125, 768-780 (1996).
80. M. Bertero, D. Bindi, P. Boccacci, M. Cattaneo, C. Eva, V. Lanza, 'Application of the projected Landweber method to the estimation of the source time function in seismology', *Inverse Problems,* 13, 465-486 (1997).
81. M. Vallée, 'Stabilising the empirical Green function analysis: Development of the projected Landweber method', *Bull. Seimol. Soc. Am.,* 94, 394-409 (2004).
82. C. F. Richter, '*Elementary Seismology',* W. H. Freeman and Company, San Francisco (1958).
83. K. Sieh, L. Jones, E. Hauksson, K. Hudnut, D. Eberhart-Phillips, T. H. Heaton, S. Hough, K. Hutton, H. Kanamori, A. Lilje, S. Lindvall, S. F. McGill, J. Mori, C. Rubin, J. A. Spotila, J. Stock, H.-K. Thio, J. Treiman, B. Wernicke, J. Zachariasen, 'Near-field investigations of the Landers earthquake sequence, April to July 1992', *Science,* 260, 171-176 (1993).
84. D. P. Robinson, C. Henry, S. Das, J. H. Woodhouse, 'Simultaneous Rupture Along Two Conjugate Planes of the Wharton Basin Earthquake', *Science,* 292, 1145-1148 (2001).
85. R. E. Abercrombie, M. Antolik, G. Ekström, 'The June 2000 M_w 7.9 earthquakes south of Sumatra: Deformation in the India–Australia plate', *J. Geophys. Res.,* 108, doi:10.1029/2001JB000674 (2003).
86. A. Kaverina, D. Dreger, E. Price, 'The Combined Inversion of Seismic and Geodetic Data for the Source Process of the 16 October 1999 M_w 7.1 Hector Mine, California, Earthquake', *Bull. Seismol. Soc. Am.,* 92(4), 1266-1280 (2002).
87. R. W. Graves, D. J. Wald, 'Resolution analysis of finite fault source inversion using one- and three-dimensional Green's functions 1. Strong motions', *J. Geophys. Res.,* 106 (B5), 8745-8766 (2001).
88. S. Yoshida, K. Koketsu, 'Simultaneous inversion of waveform and geodetic data for the rupture process of the 1984 Naganoken–Seibu, Japan, earthquake', *Geophys. J. Int.,* 103 (2), 355-362 (1990).

Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

The Mechanics of Faulting: From Laboratory to Real Earthquakes, 2012: 93-124
ISBN: 978-81-308-0502-3 Editors: Andrea Bizzarri and Harsha S. Bhat

4. Numerical algorithms for earthquake rupture dynamic modeling

Luis A. Dalguer
Swiss Seismological Service, ETH Zurich, CH-8092, Switzerland

Abstract. Numerical models of dynamic fault rupture provide a convenient framework to investigate the physical processes involved in the fault rupture during earthquake and the corresponding ground motion. This kind of model usually idealizes the earthquake rupture as a dynamically running shear crack on a frictional interface embedded in a linearly elastic continuum. This idealization has proven to be a useful foundation for analyzing natural earthquakes. The problem basically incorporates conservation laws of continuum mechanics, constitutive behavior of rocks under interface sliding, and state of stress in the crust. The fault kinematics (slip), is determined dynamically as part of the solution itself, by solving the elastodynamic equation coupled to frictional siding. Here we describe the numerical implementation of this problem in finite difference solvers, but easily can be adapted to the different classes of finite element methods. Two approaches of fault representation are formulated, first the so called traction at split-node (TSN) scheme in which explicitly incorporates the fault discontinuity at velocity (and/or displacement) nodes, and second the inelastic-zone scheme, so called stress glut (SG) method, in which approximate the fault-rupture conditions through inelastic increments to the stress components. Finally we develop numerical tests to shortly evaluate the numerical models as well as to analyze some rupture phenomena.

Correspondence/Reprint request: Dr. Luis A. Dalguer, Swiss Seismological Service, ETH Zurich, CH-8092, Switzerland. Email: dalguer@sed.ethz.ch

Introduction

The study of earthquake rupture using dynamic models has the potential for important contributions to understanding different aspects related to the earthquake mechanism and near source ground motion. The idealization that earthquake ruptures in a shear crack embedded in a linearly elastic continuum, propagating spontaneously under pre-defined conditions of initial stresses, and sliding under a constitutive friction law, is a useful model for analyzing natural earthquake (e.g., [1,2,3,4,5,67,8,9,10,11,12,13]). This model leads to nonlinear, mixed boundary value problems. The nonlinearity occurs because the respective domains of the kinematic and dynamic boundary conditions are time dependent, and these domains have to be determined dynamically as part of the solution itself. The theoretical study of this problem class is usually possible only with computationally intensive numerical methods that solve the elastodynamic equations of motion in the continuum, coupling them to additional equations governing frictional sliding on the boundary representing the fault surface.

Suitable numerical solution techniques for the spontaneous rupture problem can be built into elastodynamic methods based upon, for example, finite difference (FD), finite element (FE), spectral element (SE), Discontinuous Galerking (DG) or boundary integral (BI) methods. Each of these numerical methods can be implemented on any of several different grid types, and the elastodynamic equations solved to any specified order of accuracy. However, recent work by [14,15,16] has shown, at least in the case of the most widely used FD-based methods, that solution accuracy is controlled principally by the numerical formulation of the jump conditions on the fault discontinuity. In that study, as stated in [16], neither grid type nor order of spatial differencing in the grid is found to have a significant effect on spontaneous-rupture solution accuracy, but the method of approximation of the jump conditions has a very large effect. It is likely that a similar conclusion will hold for other solution methods such as the different classes of FE [16].

Here we compile some parts of our series of papers [14,15,16] to describe and evaluate the applications of two of the well know fault representation methods: 1) the so called traction-at-split-node (TSN) methods, and 2) the ‘‘inelastic-zone’’ stress glut (SG) method.

The TSN Methods represent the fault discontinuity by explicitly incorporating discontinuity terms at velocity and/or displacement nodes in the grid. It is the most widely used in different type of volumetric numerical methods, such as in the different classes of FD (e.g: [1,17,4, 14,15,16,17,18,19]), In FE methods (e.g. [20,21,22,23,24,25,26]) in SE

methods (e.g. 27,28,29]). In the TSN method, interactions between the halves of the ‘‘split nodes’’ occur exclusively through the tractions (frictional resistance and normal traction) acting between them, and they in turn are controlled by the jump conditions and a friction law. This method permits a partition of the equations of motion into separate parts governing each side of the fault surface [14,16].

The SG method, a class of ‘‘inelastic-zone’’ models [15], introduced by [1,17], represents the fault discontinuity through inelastic increments to stress components at a set of stress grid points taken to lie on the fault plane. With this type of scheme, the fault surface is indistinguishable from an inelastic zone with a thickness given by the spatial step Δx (or an integral multiple of Δx). The SG methods are very easy to implement in FD codes, as no modification to the difference equations is required, only modifications to the way stress is calculated from strain rate. However, from the study of [15], in which the different classes of fault representation methods in FD schemes have been evaluated, the SG method is less accurate than the TSN formulation. In a 3D test, as shown by [15], the SG inelastic-zone method achieved solutions that are qualitatively meaningful and quantitatively reliable to within a few percent, but full convergence is uncertain, and SG proved to be less efficient computationally, relative to the TSN approach.

For academic purpose, in appendix, we provide a matlab script attached to a formulation of the TSN method implemented in a FD 1D elastodynamic equation. This matlab script is intended to introduce the reader to a conceptual implementation of the TSN in a numerical code.

Theoretical formulation of the problem

The problem is formulated assuming an isotropic, linearly elastic infinite space containing a fault surface $\sum$ across which the displacement vector may have a discontinuity (Figure 1).

Assuming that surface $\sum$ is parallel to the *x*-*y* plane, that is, perpendicular to the *z* axis, the linearized elastodynamics equations of the continuous media surrounding the fault surface $\sum$ is represented, in its velocity-stress form, as:

$$\rho \frac{\partial \dot{u}_x}{\partial t} = \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xz}}{\partial z} \tag{1a}$$

$$\rho \frac{\partial \dot{u}_y}{\partial t} = \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \tau_{yz}}{\partial z} \tag{1b}$$

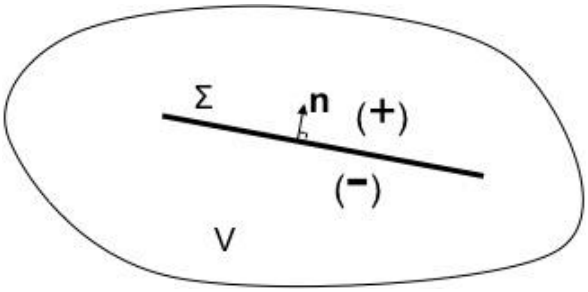


Figure 1. Schematic representation of an space of volume V containing a fault surface ∑ with normal unit vector **n** directed from negative side toward positive side of the fault.

$$\rho \frac{\partial \dot{u}_z}{\partial t} = \frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} \tag{1c}$$

and the constitutive law (Hooke's law) as:

$$\frac{\partial \sigma_{xx}}{\partial t} = (\lambda + 2\mu)\frac{\partial \dot{u}_x}{\partial x} + \lambda\left(\frac{\partial \dot{u}_y}{\partial y} + \frac{\partial \dot{u}_z}{\partial z}\right), \tag{2a}$$

$$\frac{\partial \sigma_{yy}}{\partial t} = (\lambda + 2\mu)\frac{\partial \dot{u}_y}{\partial y} + \lambda\left(\frac{\partial \dot{u}_x}{\partial x} + \frac{\partial \dot{u}_z}{\partial z}\right), \tag{2b}$$

$$\frac{\partial \sigma_{zz}}{\partial t} = (\lambda + 2\mu)\frac{\partial \dot{u}_z}{\partial z} + \lambda\left(\frac{\partial \dot{u}_x}{\partial x} + \frac{\partial \dot{u}_y}{\partial y}\right), \tag{2c}$$

$$\frac{\partial \tau_{xy}}{\partial t} = \mu\left(\frac{\partial \dot{u}_y}{\partial x} + \frac{\partial \dot{u}_x}{\partial y}\right), \tag{2d}$$

$$\frac{\partial \tau_{xz}}{\partial t} = \mu\left(\frac{\partial \dot{u}_z}{\partial x} + \frac{\partial \dot{u}_x}{\partial z}\right), \tag{2e}$$

$$\frac{\partial \tau_{yz}}{\partial t} = \mu\left(\frac{\partial \dot{u}_z}{\partial y} + \frac{\partial \dot{u}_y}{\partial z}\right), \tag{2f}$$

Parameters λ and μ are the Lame constants, ρ is density, $\dot{u} = \partial u / \partial t$ is the particle velocity formulated as the time derivative of the displacement u, σ is

the normal stress and τ is the shear stress. The fault surface $\sum$ has a (continuous) unit normal vector **n**. In our simple problem statement, in which no geometrical fault complexities are considered, this unit normal vector is always parallel to the axis z and directed toward the positive axis of z. A discontinuity in the displacement is permitted across the interface $\sum$. On $\sum$ we define negative and positive sides of the fault surface such that **n** (z axis) is directed from the former toward the latter. Taken $\sum$ to be the plane $z=0$, the limiting values of the displacement vector, u_v^+ and u_v^-, is

$$u_v^{\pm}(v,z=0,t) = \lim_{\varepsilon \to 0} u_v(v,z=\pm\varepsilon,t), \quad \varepsilon > 0 \tag{3}$$

The superscripts (+) and (-) denote, respectively, the plus-side and minus-side of the fault plane (Figure 1); v indicates the vector components x, y tangential to the fault or z normal to the fault. Then the slip vector, defined as the discontinuity of the vector of tangential displacement of the positive side relative to the negative side, is given by ($v=x,y$)

$$s_v(t) = u_v^+(t) - u_v^-(t) \tag{4}$$

and its time derivative (slip rate) is denoted by $\dot{\mathbf{s}}$. The magnitude of the slip and slip rate are denoted, respectively, by $|s|$ and $|\dot{s}|$. The open fault displacement ($v=z$) is formulated later.

The total shear traction vector (**T**) acting on the fault ($z=0$) that is continuous across $\sum$ with components $T_x = \tau_{xz}^0 + \Delta\tau_{xz}$ and $T_y = \tau_{yz}^0 + \Delta\tau_{yz}$ has its magnitude

$$|T| = \sqrt{T_x^2 + T_y^2} \tag{5}$$

where $\Delta\tau$ and τ^0 are, respectively, the shear stress change during rupture and initial shear stress.

As formulated in [14, 15, 16], the jump (rupture) conditions at the interface is given by

$$\tau_c - |T| \geq 0 \tag{6a}$$

$$\tau_c \dot{\mathbf{s}} - \mathbf{T}|\dot{s}| = 0 \tag{6b}$$

Equation (6a) stipulates that the total shear traction T is bounded by a nonnegative frictional strength τ_c, and equation (6b) stipulates that any

nonzero velocity discontinuity be opposed by an antiparallel traction (i.e., the negative side exerts traction -**T** on the positive side) with magnitude equal to the frictional strength τ_c. The frictional strength evolves according to some specified friction law

$$\tau_c = f(\sigma_n, s, \dot{s}, \varphi_1, \varphi_2 ...) \tag{7}$$

that may depends on normal stress (σ_n), slip (s), slip rate ($\dot{s}$), and other mechanical or thermal variables (φ_1, $\varphi_2 \ldots$).

Jump conditions (6a)–(6b), combined with the friction law (7) and appropriate initial stress conditions on $\sum$, provide a model of fault behavior. Under these conditions alone can model initial rupture, arrest of sliding and reactivation of slip.

When normal stress fluctuations are presents, the fault interface may undergo separation (fault opening) over portions of the contact surface $\sum$ if there is a transient reduction of the compressive normal stress to zero [30,31]. For the sake of completeness, as formulated by [14], we describe an extension of the set of jump conditions appropriate to also incorporate fault opening due to normal stress fluctuations. We denote the normal component of the displacement discontinuity on $\sum$ by U_n (fault opening displacement). From Equation 3, for v=z, the fault opening is given by

$$U_n(t) = u_z^+(t) - u_z^-(t) \tag{8}$$

The opening conditions, assuming negative normal stress in compression are

$$\sigma_n \le 0, \tag{9a}$$

$$U_n \ge 0, \tag{9b}$$

$$\sigma_n U_n = 0 \tag{9c}$$

σ_n is the total normal stress acting on the fault that is given by $\sigma_n = \sigma_{zz}^0 + \Delta\sigma_{zz}$ where $\Delta\sigma$ and σ^0 are, respectively, the normal stress change during rupture and initial normal stress.

Equation (9a) bounds the total normal stress by the condition that tensile normal stress is not permitted; equation (9b) guarantees no interpenetration; and equation (9c) stipulate that loss of contact is permitted only if accompanied by zero normal stress. Again, these jump conditions are adequate to cover multiple episodes of tensile rupture and crack closure.

Traction-at-split-node (TSN) fault representation formulation

The TSN boundary formulation treats the fault rupture as a true contact problem between two surfaces in which the kinematic shear discontinuity (slip) as well as the open discontinuity (fault opening) are explicitly modeled. This method (for shear discontinuity) was reviewed by [17] and described the formulation in detail by [14,15,16] for implementation in finite difference schemes. Dalguer and Day [16] adapted it for a fourth-order velocity-stress staggered finite difference code. Here we give a general description of the method following [14,15].

We position the fault on the x-y plane. As shown in Figure 2, a given fault-plane node is split into plus-side and minus-side parts, with respective lumped nodal masses M^+ and M^-. The separate contributions from each side

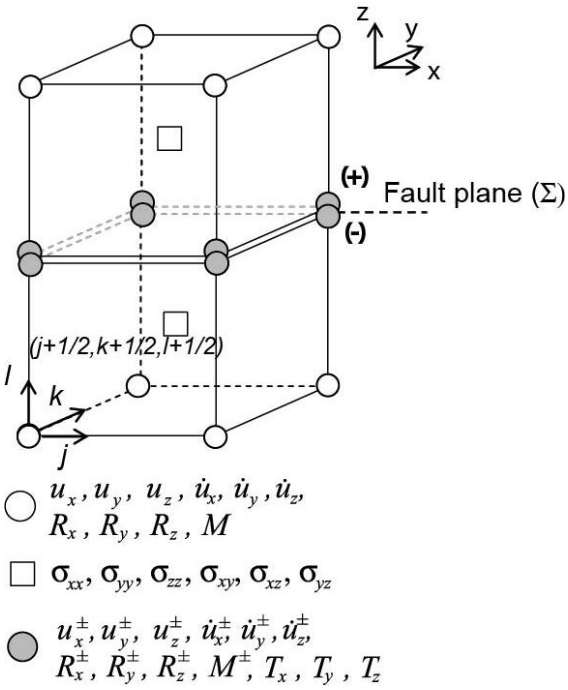


Figure 2. Traction at split node (TSN) fault representation method in a partially staggered cubic elements. Mass ($M^{\pm}$) is split, and separate elastic restoring forces ($R_v^{\pm}$) act on the two halves. The two halves of a split node interact only through shear and normal tractions (T_v) at the interface.

of the fault due to deformation of neighboring elements produce elastic restoring forces (nodal forces), R^+ and R^-. At a particular time (t), D'Alembert's principle leads to a nodal equilibrium equation of motion for each split node. At each step of integration the equation is solved by the central FD scheme to estimate the vector components of velocity ($\dot{u}_v^{\pm}$) and displacement ($u_v^{\pm}$) at a given node,

$$\dot{u}_v^{\pm}(t+\Delta t/2) = \dot{u}_v^{\pm}(t-\Delta t/2) + \frac{\Delta t}{M^{\pm}}\left\{R_v^{\pm}(t) \mp a\left[T_v(t) - T_v^0\right]\right\} \tag{10a}$$

$$u_v^{\pm}(t+\Delta t) = u_v^{\pm}(t) + \Delta t \dot{u}_v^{\pm}(t+\Delta t/2) \tag{10b}$$

where v indicates the vector components x, y tangential to the fault or z normal to the fault, Δt is the time step, a is the area of the fault surface associated with each split node, T_v is the nodal value of the traction-vector components, and T_v^0 is the corresponding initial equilibrium value. The slip and slip velocity vectors (for $v = x$ or y) are then

$$s_v(t) = u_v^+(t) - u_v^-(t) \tag{11a}$$

$$\dot{s}_v(t+\Delta t/2) = \dot{u}_v^+(t+\Delta t/2) - \dot{u}_v^-(t+\Delta t/2) \tag{11b}$$

and fault opening displacement and velocity (making $v = z$)

$$U_n(t) = u_z^+(t) - u_z^-(t) \tag{12a}$$

$$\dot{U}_n(t+\Delta t/2) = \dot{u}_z^+(t+\Delta t/2) - \dot{u}_z^-(t+\Delta t/2) \tag{12b}$$

To find the slip, slip velocity and fault opening displacement we need to solve equation 10 by evaluating T_v as follow.

Evaluation of T_v for shear traction (kinematic fault tangential discontinuity)

An appropriate methodology is defining a trial traction vector $\tilde{T}_v$ that would be required to enforce continuity of tangential velocity ($\dot{u}_v^+ - \dot{u}_v^- = 0$ for v equal to x and y) in equation (10a). The expression for $\tilde{T}_v$ is then estimated after few operations in equations 10-11 [14,15,16]

$$\tilde{T}_v \equiv \frac{\Delta t^{-1} M^+ M^- (\dot{u}_v^+ - \dot{u}_v^-) + M^- R_v^+ - M^+ R_v^-}{a(M^+ + M^-)} + T_v^0, \qquad v = x, y \tag{13}$$

where the velocities are evaluated at t-$\Delta t/2$, and the nodal tractions, restoring forces, and displacements are evaluated at t. The fault-rupture conditions stated in equations (6a,b) are satisfied if the fault-plane traction T_v of equation (10a) is

$$T_v = \begin{cases} \tilde{T}_v & \text{for} \quad \left[\left(\tilde{T}_x\right)^2 + \left(\tilde{T}_y\right)^2\right]^{1/2} \le \tau_c \\ \tau_c \dfrac{\tilde{T}_v}{\left[\left(\tilde{T}_x\right)^2 + \left(\tilde{T}_y\right)^2\right]^{1/2}} & \text{for} \quad \left[\left(\tilde{T}_x\right)^2 + \left(\tilde{T}_y\right)^2\right]^{1/2} > \tau_c \end{cases} \tag{14}$$

for $v = x,y$.

Evaluation of T_v for normal traction (kinematic fault normal discontinuity)

The same way as before, a trial fault normal traction $\tilde{T}_z$ (making $v=z$) that would be required to enforce continuity of normal displacement ($u_z^+ - u_z^- = 0$) in (10b) is estimated. After some operations in equations 10 and 12 the expression of $\tilde{T}_z$ is given by

$$\tilde{T}_z \equiv \frac{\Delta t^{-1} M^+ M^- (\dot{U}_n^{t-\Delta t/2} + \Delta t^{-1} U_n^t) + M^- R_z^+ - M^+ R_z^-}{a(M^+ + M^-)} + T_z^0 \tag{15}$$

where $\dot{U}_n^{t-\Delta t/2}$ is the fault opening velocity estimated at t-$\Delta t/2$ and U_n^t is the fault opening displacement estimated at t calculated using eq. (12). Assuming negative normal stress in compression and satisfying fault open conditions stated in equations (9), the fault normal traction T_z of equation 10a is

$$T_z = \begin{cases} \tilde{T}_z & \text{for} \quad \tilde{T}_z \le 0 \\ 0 & \text{for} \quad \tilde{T}_z > 0 \end{cases} \tag{16}$$

The conditions in (16) guarantee no interpenetration and nontensile normal stress (i.e. the fault resistant to tensile is zero), consequently loss of contact between the two surface of the fault (opening) occurs only if accompanied by zero normal stress. This open fault boundary condition is rather simplified approximation, since the fault opening may follow a pre- process in which a certain amount of tensile stress may be admitted to break the contact between the two surfaces of the fault, but this pre-process is ignored here.

Stress glut (SG) ‘‘inelastic-zone’’ fault representation formulation

The SG method has been documented by [1,17] and adapted it to a fourth-order velocity-stress staggered finite difference scheme by [15]. Considering the same grid element features in which the TSN formulation has been implemented above (Figure 2), the principal difference between the TSN method and the SG formulation is that the latter does not split the nodes neither place velocity nodes on the fault, but instead positions the fault to coincide with the standard grid points already containing the fault plane traction components (Figure 3). The fault discontinuity is not explicitly incorporated, rather it is represented through inelastic increments to those traction components. As

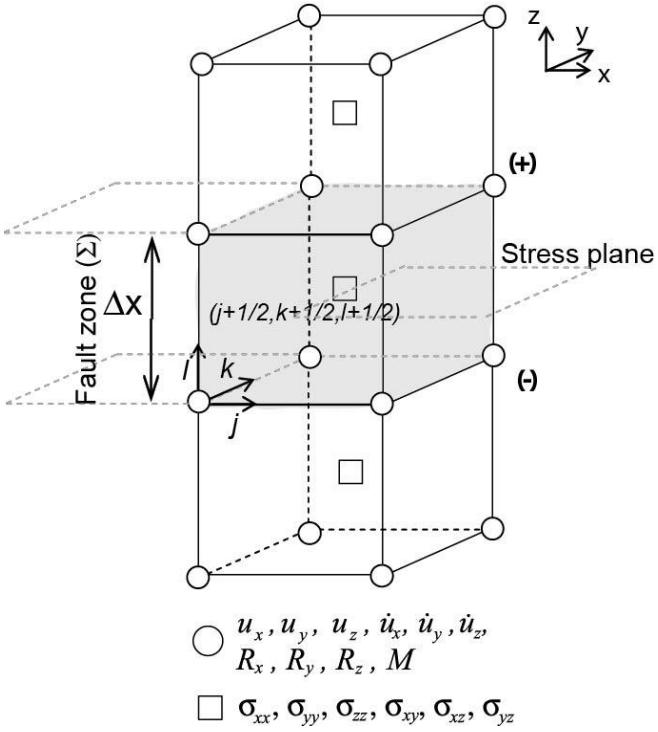


Figure 3. Inelastic-zone Stress glut (SG) fault representation method in a partially staggered cubic elements. The shear and normal tractions (T_v) acting on the fault are approximated by modifying the shearstress components located along the plane coinciding with the fault (labeled “stress plane”). This is equivalent to an inelastic zone of one grid-step (Δx) thickness.

shown in Figure 3, this formulation makes the fault indistinguishable from an inelastic zone of thickness Δx, where Δx is the dimension of the unit cell (assumed equal in all three coordinate directions for simplicity) of the grid. Due to this fault configuration, the fault normal discontinuity (fault opening) is also not explicit. So for this case, here we do not formulate fault opening boundary condition, and we limit our formulation to shear faulting boundary condition.

Here we reproduce the formulation stated in [15]. We again take the x-y plane as the fault surface. In the split-node method, we introduced extra grid variables T_x and T_y on the fault to represent the traction-vector components at the split nodes. In the SG method, no extra tractions have to be introduced to accommodate the fault; the faultplane traction components are located at the standard grid points for the tensor components σ_{xz} and σ_{yz}, respectively. However, we continue to use T_x and T_y to denote these two shear-traction components when they are located on the fault, for notational consistency with the split-node discussion.

Using the velocity-stress formulation of the equation of motion (Eq. 1 and 2), lets update nodal stresses assuming central differencing in time by using strain rates calculated from nodal velocities at $t-\Delta t/2$. Then, the shear stress components at a particular point acting on the fault plane take the form

$$\sigma_{vz}(t) = \sigma_{vz}(t-\Delta t) + \Delta t 2\mu \dot{\varepsilon}_{vz}(t-\Delta t/2) \tag{17}$$

where v indicates the vector components x, y tangential to the fault, $\dot{\varepsilon}_{vz}(t-\Delta t/2)$ is the strain rate and μ is the shear modulus. To implement the SG method, we modify this stress update scheme when calculating fault-plane traction components $T_v(t)$ by the addition of an inelastic component ($\dot{\varepsilon}^p_{vz}$) to the total strain rate:

$$T_v(t) = T_v(t-\Delta t) + \Delta t 2\mu\left[\dot{\varepsilon}_{vz}(t-\Delta t/2) - \dot{\varepsilon}^p_{vz}(t-\Delta t/2)\right] \tag{18}$$

Then, as proceeded for the TSN method, lets calculate a trial traction, $\tilde{T}_v(t)$, that would be required to enforce zero inelastic strain rate, i.e,

$$\tilde{T}_v(t) = T_v(t-\Delta t) + \Delta t 2\mu \dot{\varepsilon}_{vz}(t-\Delta t/2) \tag{19}$$

Then the fault-plane traction $T_v(t)$ that satisfy fault-rupture conditions stated in equations (6) is calculated using eq. (14). The inelastic strain rate $\dot{\varepsilon}^p_{vz}$ is estimated after some operations between equation (14), (18) and (19)

$$\dot{\varepsilon}_{yz}^{p}(t-\Delta t/2)=\frac{\tilde{T}_v(t)-T_v(t)}{2\mu\Delta t} \tag{20}$$

Fault slip is estimated through inelastic increments distributed in an inelastic zone of thickness Δx. Then the total slip rate on the fault is calculated by integrating the inelastic strain rate over the spatial step Δx in the direction normal to the fault, which gives

$$\dot{s}_v(t-\Delta t/2)=2\Delta x\dot{\varepsilon}_{yz}^{p}(t-\Delta t/2) \tag{21}$$

from which the slip is then updated by central differencing,

$$s_v(t)=s_v(t-\Delta t)+\Delta t\dot{s}(t-\Delta t/2) \tag{22}$$

Frictional shear strength: Slip weakening friction model

As described in Eq. (7), the frictional shear strength τ_c in its general form evolves according to some specified friction law, and may depend upon normal stress, slip, slip velocity, and other mechanical or thermal variables. For simplicity, here we use the simple slip-weakening friction model in the form given by [1,2]. This friction law, first proposed by [32,33] by analogy to cohesive zone models of tensile fracture, is extensively used for shear dynamic rupture simulations (e.g.[1,2,4,6,34,35,36,37,8].

The frictional strength τ_c is assumed to be proportional to normal stress σ_n (taken negative in compression)

$$\tau_c=-\mu_f\sigma_n \tag{23}$$

The coefficient of friction μ_f depends on the slip path length through the linear slip-weakening relationship [2]

$$\mu_f=\begin{cases}\mu_s-(\mu_s-\mu_d)|s|/d_0 & \text{for } |s|<d_0\\ \mu_d & \text{for } |s|\geq d_0\end{cases} \tag{24}$$

where μ_s and μ_d are coefficients of static and dynamic friction, respectively, d_0 is the critical slip-weakening distance, and $|s|$ is the magnitude of the slip vector.

Despite its limitations of the slip weakening model as a model for natural earthquakes (as noted in, e.g., [14], this friction law provides a suitable starting point for testing numerical methods. Other friction models are out of the scope of this chapter, in which interface frictional properties may be

better represented by more complicated relationships that account for rate and state effects (e.g. [38,39]) and thermal phenomena such as flash heating and pore pressure evolution (e.g., [40,41,42,43]).

Cohesive zone

The cohesive zone (Figure 4a) is the portion of the fault plane behind the crack tip where the shear stress decreases from its static value to its dynamic value and slip $|s|$ satisfies $0 < |s| < d_0$ (e.g. [32]). The cohesive model was first introduced by [44] in which constant cohesive zone was considered. Subsequently [32,33] proposed a cohesive zone model linearly dependent on distance to the crack tip; and Andrews [2] proposed a model linearly dependent on slip. Basically the models of [2,32,33] are equivalent and well know as slip-weakening model as formulated above. In this friction model, the cohesive zone, as shown in Figure 4a, examines the crack tip phenomena at a level of observation, in which the fracture energy G_c, (Figure 4b) is a mesoscopic parameter which contains all the dissipative processes in the volume around the crack tip, such as off-fault yielding, damage, micro-cracking etc. In the event that the normal stress and frictional parameters are constant over the entire fault, as will be the case in the test problem considered later, this idealized model results in constant fracture energy G_c with

$$G_c = (\tau_s - \tau_d) d_0 / 2 \tag{25}$$

where τ_s and τ_d are, respectively, the peak shear stress (static yielding stress) and dynamic yielding stress, given by

$$\tau_s = \mu_s |\sigma_n| \tag{26}$$

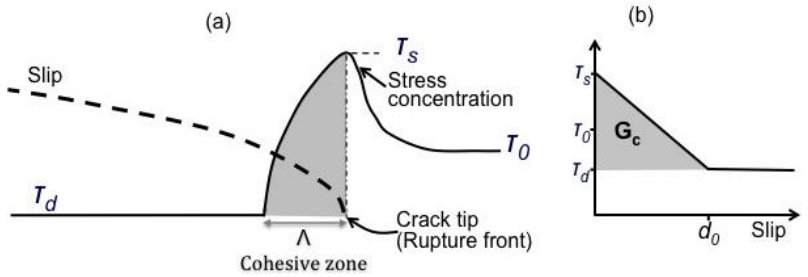


Figure 4. (a) Schematic representation of stress and slip along a shear crack and cohesive zone for a slip-weakening crack; (b) Stress-slip relationship of a slip-weakening model [2] and fracture energy G_c representation.

$$\tau_d = \mu_d |\sigma_n| \tag{27}$$

Note that in this context, the fracture energy G_c is not the surface energy defined by Griffith [45] in linear elastic fracture mechanics.

In the cohesive zone, shear stress and slip rate vary significantly, and proper numerical resolution of those changes is crucial for capturing the maximum slip rates and the rupture propagation time and speeds. Therefore an estimate of the cohesive zone width to calibrate numerical resolution would be useful. A review of some concepts of linear fracture mechanics and simple estimates for the cohesive zone size in two-dimensional cases of mode II and mode III was presented by [14]. These authors provide two ways to estimate the cohesive zone size and calibrate numerical resolution: the zero-speed cohesive zone width Λ_0 given by

$$\Lambda_0 = \frac{9\pi}{32} \frac{\mu_m d_0}{(\tau_s - \tau_d)} \tag{28}$$

for m = *II, III*, respectively mode II and mode III rupture; where $\mu_{II} = \mu$; $\mu_{III} = \mu/(1-\nu)$, with ν as the Poisson's ratio. [14] also approximate solution for Λ at large propagation distances (for mode III crack problems) given by

$$\Lambda = \frac{9}{16} \left(\frac{\mu d_0}{\Delta\tau} \right)^2 L^{-1} \qquad for \quad L >> L_0 \tag{29}$$

where $\Delta\tau = (\tau_0 - \tau_d)$ is the stress drop, τ_0 the initial stress, L propagation distance, and L_0 is half of the critical crack length for a 2D crack given by

$$L_0 = \frac{\mu d_0 (\tau_s - \tau_d)}{\pi \Delta\tau^2} \tag{30}$$

As pointed out by [14], the two estimates of the cohesive width are complementary. The Λ_0 estimate shows that regardless of the background stress or rupture propagation distances, the numerical resolution is already constrained by the choice of the frictional parameters and elastic bulk properties; and it provides a convenient upper bound for the cohesive zone size (it is an upper bound in the sense that any nonzero rupture speed would shrink this zone even further due to Lorentz contraction [14]).

As stated in [14] the Λ estimate attempts to incorporate the background stress level (through the stress drop $\Delta\tau$) and the reduction of the cohesive zone (Lorentz contraction) due to the increasing rupture speed for large propagation distances L.

To relate numerical accuracy to the degree to which the cohesive zone is resolved, the authors in [14] have expressed the grid-size dependence of the solution in terms of the dimensionless ratio N_c. Where N_c is the ratio of the width of the cohesive zone, Λ, to the grid interval Δx.

$$N_c = \Lambda / \Delta x \tag{31}$$

This ratio provides a non-dimensional characterization of the resolution of a given numerical solution. As discussed in [14], even thought N_c is a local measure of resolution, because Λ varies as the rupture propagates, both the Λ_0 estimate from (28) and the Λ estimate (29) should give good initial guidance as to what kind of spatial resolution will be needed in dynamic rupture propagation problems. However, as pointed out by [14], one should not expect a perfect quantitative agreement, as the estimates are derived with a number of simplifying assumptions.

Numerical test

SCEC benchmark problem version 3

Here we present some results collected from the series of papers [14,15,16], in which we have solved a three-dimensional (3D) problem of spontaneous rupture propagation for a planar fault embedded in a uniform infinite elastic isotropic space, using different numerical methods of Finite Difference and Boundary Integral (BI). The formulation and parameters of the test case correspond to Version 3 of the Southern California Earthquake Center (SCEC) benchmark problem [46]. The problem geometry is shown in Figure 5.

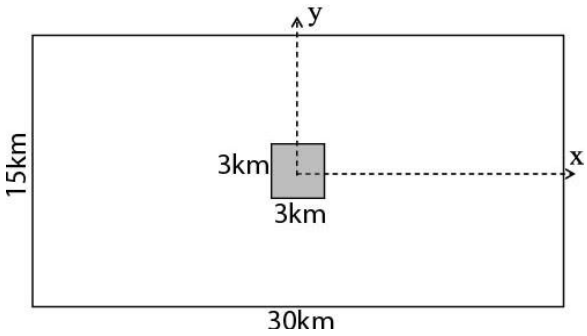


Figure 5. Fault geometry to test dynamic rupture simulation. The square in the center is the nucleation area where rupture initiates.

Table 1. Stress parameters for the numerical test of spontaneous dynamic rupture simulation.

Parameters	Within Fault Area of 30 km x 15km		Outside Fault Area
	Nucleation	Outside nucleation	
Initial shear stress (τ_0), MPa	81.6	70.0	70.0
Initial normal stress (σ_n), MPa	120.0	120.0	120.0
Static friction coefficient (μ_s)	0.677	0.677	infinite
Dynamic friction coefficient (μ_d)	0.525	0.525	0.525
Static yielding stress ($\tau_s = \mu_s\sigma_n$), MPa	81.24	81.24	infinite
Dynamic yielding stress ($\tau_d = \mu_d\sigma_n$), MPa	63.0	63.0	63.0
Dynamic stress drop ($\Delta\tau = \tau_0 - \tau_d$), MPa	18.6	7.0	7.0
Strength excess ($\tau_s - \tau_0$), MPa	-0.36	11.24	infinite
Critical slip distance, d_0 , m	0.40	0.40	0.40

We take the fault plane to be the *x*-*y* plane. The shear pre-stress is aligned with the *x* axis, and the origin of the coordinate system is located in the middle of the fault, as shown in Figure 5. The fault and pre-stress geometries are such that the *x* and *y* axes are axes of symmetry (or antisymmetry) for the fault slip and traction components. As a result, the *xz* plane undergoes purely in-plane motion, and the *yz* plane purely anti-plane motion.

Rupture is allowed within a fault area that extends 30 km in the x direction and 15 km in the y direction. A homogeneous medium is assumed, with a P wave velocity of 6000 m/s, S wave velocity of 3464 m/s, and density of 2670 kg/m^3. The distributions of the initial stresses and frictional parameters on the fault are specified in Table 1.

Rupture nucleation

The rupture initiation of this kind of dynamic rupture problems is artificial and nucleation procedure can affect the rupture propagation (e.g. [47]). Here we adopt the criterion of overloading the initial stress at the nucleation patch, so rupture can initiates because the initial shear stress in the nucleation is set to be slightly (0.44%) higher than the initial static yield stress in that patch. Then the rupture propagates spontaneously through the fault area, following the linear slip-weakening fracture criterion (25). The nucleation size for the problem can be roughly estimated using equation (30) that give a value of L_0= 1.516km, which is half of the nucleation size. We assume that the nucleation shape is a square, so it will give a 3 km x 3 km square area centered on the fault, as shown in Figure 5.

Estimate of spatial resolution of the numerical model

As mentioned before, the cohesive zone developed during rupture propagation need to be accurately solved to obtain reliable solution of the problem. Then before simulation it is convenient to have some estimates of the degree of the numerical accuracy by calculating the spatial resolution to which the cohesive zone is resolved. For this purpose the approximate analytical cohesive zone Λ_0 from (28) and the Λ of (29) are calculated to estimate dimensionless ratio N_c of equation (31) as good initial guidance to define the spatial resolution needed for the test problem. Using the data of the test problem, we obtain zero-speed cohesive zone Λ_0 = 620m for mode III, and Λ_0 = 827m for mode II. They can be considered as the upper bound of our problem. The cohesive zone, Λ, at the maximum propagation distance L=7.5km along the mode III is Λ=251m. Notice that the estimate of Λ for mode II cannot be derived analytically, it needs some numerical procedure not included in this work [14].

Assuming a grid size Δx=100m, the N_c value, from Eq. (31), is 6 to 8 for the upper bound, and 2.5 for the propagation distance. Those estimates indicate that a good spatial resolution for our problem requires $\Delta x \leq$100m. The accuracy reached by this resolution will depend on the numerical method used to model the fault as well as the numerical technique used, as evaluated in [14,15,16].

Numerical techniques

The test problem is solved by two numerical techniques:

1) The so-called 3D dynamic fault model (DFM) code in which the TSN fault representation method is implemented [4,5,14, 48]. In the DFM the spatial difference operators are constructed by specializing trilinear elastic finite elements to the Cartesian mesh, approximating integrals by one-point quadrature, and diagonalizing the mass matrix (see more details of it in [14]). The method approximates temporal derivatives by explicit, central differencing in time. On a uniform mesh, the method is second-order accurate in space and time. In that case, the differencing scheme that results from this procedure is equivalent (away from the fault surface) to the second-order partly staggered grid method, which has been reviewed by [49] (see also in [50], p. 884, formula 25.3.22].
2) The 3D, four-order velocity-stress staggered (VSSG) wave propagation code of [51]. In this code we have implemented the SG and TSN fault representation method described earlier. The TSN formulation for the

VSSG FD scheme has been proposed by [16], as called by these authors, this implementation is referred to as the SGSN (staggered-grid split-node) method.

Numerical results

Numerical solutions for the DFM, SGSN and SG fault representations methods are briefly qualitatively discussed here. A complete quantitative and qualitative assessment of these methods and the solutions for this problem has been extensively discussed in our series of papers [14,15,16].

The highest grid resolution used for DFM, SGSN and SG methods are respectively 50m, 100m and 50m and referred respectively as DFM50, SGSN100 and SG50. The rupture arrival time (referred to as ‘‘rupture time’’ in the following) is a sensitive indicator of numerical precision, because this sensitivity reflects the nonlinearity of the problem. Relatively small inaccuracies in the calculated stress field can be expected to very significantly and affect the timing of rupture breakout from the nucleation zone as well as the subsequent rupture velocity. Therefore we have used rupture time differences as a primary means to show differences between our solutions. Figure 6 shows contours of rupture time for the three methods. The computed evolution of the rupture time is virtually identical for the DFM and SGSN solutions (Figure 6a), so that the contours for these two cases overlay and are nearly indistinguishable. The SG and DFM models (Figure 6b) have rupture contours that are very close together right after the initiation of the rupture, with differences increasing with the rupture propagation. As discussed in [15], rupture-time differences between SG and DFM cannot be accounted for by a simple time delay due to differences in nucleation, but

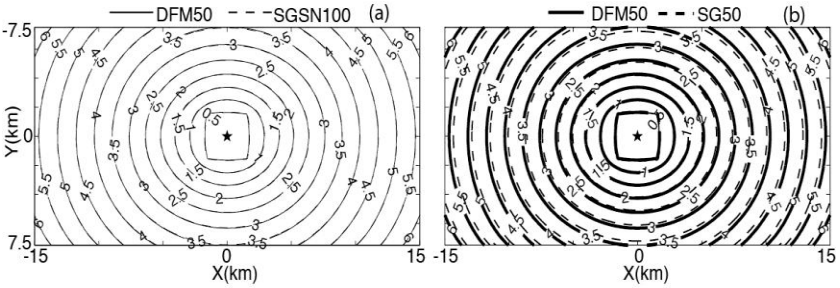


Figure 6. Contour plot of the rupture front for the dynamic rupture test problem: (a) comparison between DFM50 (grid size Δx = 50m) and SGSN100 (Δx = 100m) solutions; (b) comparison between DFM50 and SG50 (Δx = 50m) solutions.

represent systematic differences in rupture velocity over the entire rupture. Cohesive zone development for these methods along both *x* (inplane) and *y* (antiplane) axes are shown in Figure 7. DFM and SGSN are practically identical, the SG solutions produce a rupture with a cohesive-zone width that varies with propagation distance in a manner similar to the DFM and SGSN, but it is systematically narrower, but the cohesive-zone-width curves for the three methods have roughly the same shape. Qualitatively the three solutions provide comparable results. A relevant feature of the cohesive development is that as the crack velocity increases, the cohesive zone shrink in the direction of rupture propagation. This feature involves small-scale processes that need to be accurately solved, consequently it leads to numerical challenges in which calculations of such numerical simulations pose high demands in terms of required memory and processor power (e.g., [14]).

A quantitative estimation of the rupture time misfit as a function of grid interval for the three methods is shown in Figure 8. The rms misfits estimated in our papers [14,15,16] for the SG, DFM and SGSN methods use as reference solution the one calculated by the boundary integral (BI) method with grid size 100m presented in [14]. The BI method might provides

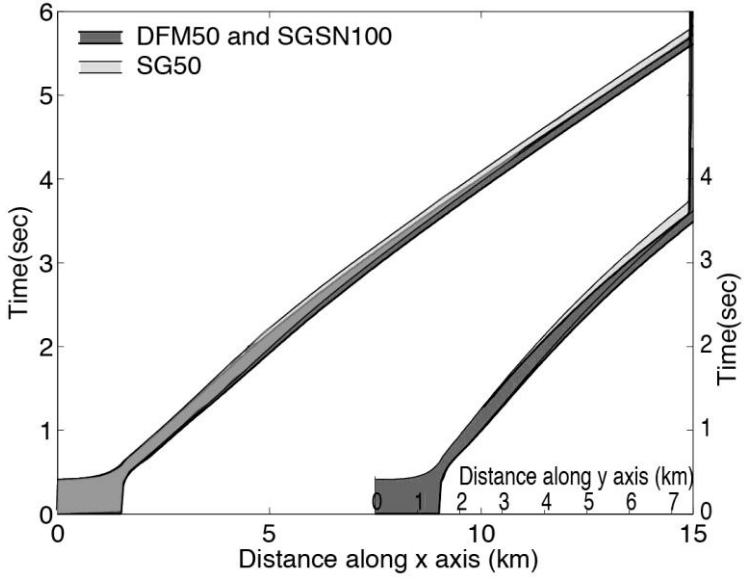


Figure 7. Cohesive zone evolution during rupture, along both inplane (x axis) and antiplane (y axis) directions for DFM50, SGSN100 and SG50.

semianalytical solutions for this problem, therefore it gives a suitable reference solution. The misfits of the DFM, SGSN and SG solutions as functions of Δx, or equivalently, as functions of resolution number N_c (Eq. 31) are shown in Figure 8. For reference, we also plot the results of the BI from [14].

As noted by [14], the DFM solutions follow a remarkably well-defined power law in the grid size, with exponent, or convergence rate, of approximately 3. DFM and BI methods share a nearly identical convergence rate and that both achieve misfits comparable. As presented by [16] the rupture-time differences for SGSN show a bilinear scaling with the grid size. The first scaling line corresponds to solutions with $\Delta x \leq 0.3$ and the second line for $\Delta x > 0.3$. The transition between these two scaling lines occurs between $\Delta x = 0.3$ and $\Delta x = 0.4$, corresponding to a grid interval slightly less

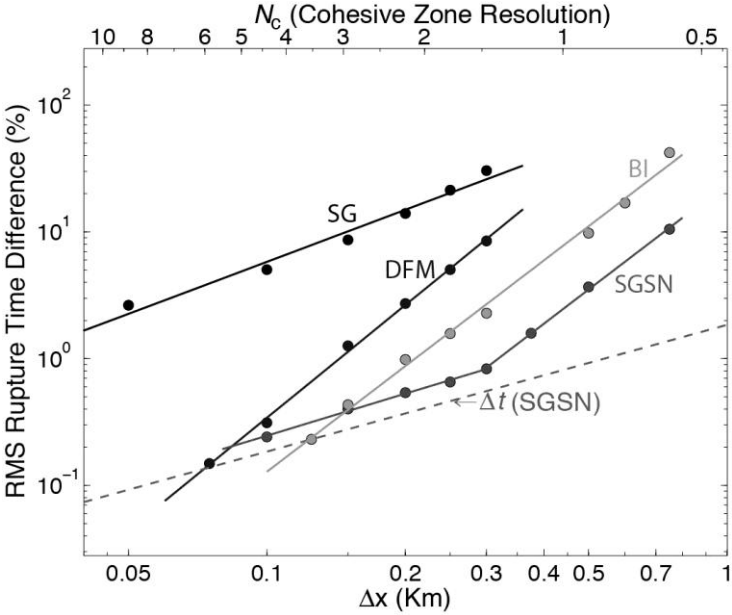


Figure 8. Misfit in time of rupture, relative to reference solution, shown as a function of grid interval Δx. Misfits are RMS averages over the fault plane for DFM, SGSN and SG solutions. All the solutions are relative to BI100m (Δx=100m) calculated by Day et al 2005. The dashed line shows the (approximate) dependence of time step Δt on Δx. The upper axis characterizes the calculations by their characteristic N_c values, where N_c is median cohesive zone width in the in-plane direction divided by Δx (Eq. 31).

than the median cohesive-zone width (L = 0.44 km) (see [16]). In the second line, for $\Delta x > 0.3$, the RMS time differences exceed 1.5% and the dependence upon Δx appears to follow a power law of exponent ~3 similar to the DFM and BI (see further discussion on it in [14,15,16]). But the SGSN has an exceptional performance. Very low misfit of order of ~1% is already achieved for $\Delta x = 0.3$, corresponding to N_c ~1.5. In contrast, the SG misfits follow a convergence rate with low power law of exponent ~ 1.4, suggesting that this method is computational less efficient than the others.

A very insightful nature of this kind of dynamic rupture models is the rupture evolution that involves: initiation, evolution and stopping of the slip, and the evolution of the stress after the slipping ceases. So we reproduce the evaluation discussed in [14] of the slip rate and shear stress time history profiles along the x axis (in-plane direction) (Figure 9a) and the y axis (antiplane direction) (Figure 9b). We show results for the DFM50 only presented in [14]. For other solutions, SGSN100 and SG50, the feature discussed here are identical. As shown in these figures the pulses associated with the P and S waves returning from the borders of the fault are observed in the time histories of slip rate and stress. In Figures 9a and 9b we annotate these fault-edge-generated pulses. The P waves from the left and right borders of the fault traveling along the in-plane direction are denoted by ‘‘P’’ in Figure 9a. The pulses associated with the edge-generated S wave are indicated by ‘‘Si’’ and ‘‘Sa,’’ with Si corresponding to the pulses coming back from the left and right borders of the fault, traveling predominantly along the in-plane direction, and Sa corresponding to the pulses coming back from the top and bottom borders, traveling predominantly along the antiplane direction. In addition to these stopping phases, a late reactivation of slip, after its initial arrest, can also be seen in these figures. This feature is associated with the Si pulse, and its behavior is explained as follows. The P wave coming back from the boundary reduces the shear stress on the fault, causing slip to stop, leaving the shear stress somewhat below the dynamic friction value (dynamic overshoot). The subsequent Si fault edge pulse has to overcome that stress deficit in order to reinitiate slip. As it approaches the center of the fault, this pulse becomes weak. This wave experiences constructive interference at the center of the fault in which there is an encounter between the Si waves coming from the left and right side of the fault. As can be seen in the figures of shear stress, the Si pulse crosses the center and continues traveling to the other side of the fault, but always below the dynamic friction level, and therefore unable to produce further slipping.

Note that our solution procedure assumes, for simplicity, that once the dynamic frictional strength τ_d is reached at a point on the fault, the strength will

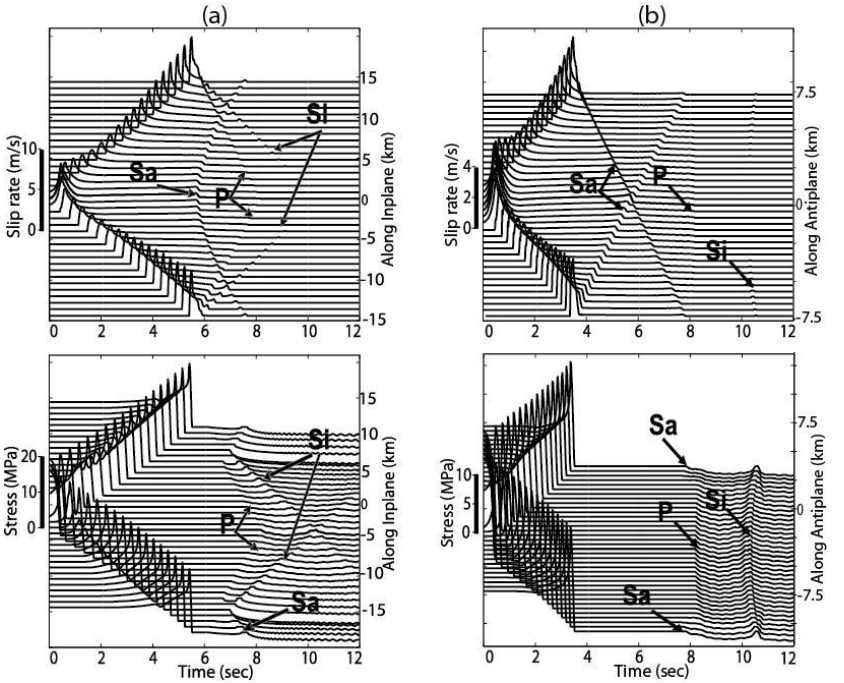


Figure 9. Time history of (top) slip rate and (bottom) shear stress for points along the axis of in-plane motion (x axis) (left) and antiplane motion (y axis) (right) for the DFM50 solution. The labels P and Si correspond to the P and S waves, respectively, generated at the left and right edges of the fault (i.e., propagating predominantly along the axis of in-plane motion). The label Sa identifies the S waves generated at the top and bottom of the fault (propagating predominantly along the antiplane axis).

not increase to larger values on the timescale of the computation, even if the point reaches zero slip velocity. That is, it is assumed that there is no healing for times of order of seconds. However, rock interfaces in the lab do exhibit healing at rest or small sliding velocities, and a more complete constitutive description would include that effect, but it is out of the scope of this work.

Large aspect-ratio fault (L>>W)

One interesting application of dynamic rupture models is to study earthquake rupture in large aspect-ratio strike-slip faults with L>>W, in which L and W are respectively the length and width of the fault. It is

expected that large earthquakes, such as the 2002 Mw 7.9 Denali and the 2008 Mw 8.0 Wenchuan earthquake in China, both with fault length about 300km, rupture the entire seismogenic thickness and are originated in large aspect-ratio faults. Therefore, the understanding of the rupture mechanism of this kind of fault is very important to address questions such as on why rupture extends so long, and what are the conditions to rupture stops before it becomes a large event.

Strike-slip faulting in a large aspect-ratio fault is dominated by the inplane rupture mode (mode II in fracture mechanics). Previous studies, such as from [4, 31] shows that in this kind of fault, the rupture is highly affected by the width (W). The main mechanism dominated in this kind of fault has been already explained in [4], that is, the fault initially ruptures as a crack-like (a simply-connected patch) around the hypocenter, but subsequently, at a time greater than that required for the rupture to cross the fault width, the rupture bifurcates into two separate pulses traveling in opposite directions due to the stopping phases coming from the top and bottom of the fault (see Figure 9 of the evolution of this stopping phases). When this process occurs in the bi-material case [31], it evolves interacting with the normal stress perturbation (characteristics of bimaterial fault rupture) and under very limited conditions it can lead to unilateral rupture (see details of this mechanism in [31]).

As complementary to the studies described above, here we model this kind of rupture problem to investigate the W effect on spontaneous rupture propagation in homogeneous strike slip faults. We fix the frictional parameters and nucleation rupture to be those of the SCEC benchmark problem, version 3, described in the previous section. Then we explore the sensitivity of rupture to variability of the fault width (W) in a fault with rupture propagation distance along strike of up to 400km. The grid size for these calculations is 50m.

Our results show that W takes an important role on rupture arresting and the generation of steady-state pulse-like rupture due to the arrival of the stopping phases (described in Figure 9) at the rupture front. Figures 10a,b,c shows respectively the rupture time, final slip and peak-slip rate along the inplane axis direction for different fault widths. Rupture is arrested for model with W <=5.9. For models larger than this width, the rupture propagation becomes self-sustained, increasing the rupture speed with increasing W.

Notice that rupture initiation for all the models is identical. All models reach the rupture speed limit (Rayleigh waves speeds) early, but then, when the rupture reaches the top and bottom of the fault, the ruptures speed, final

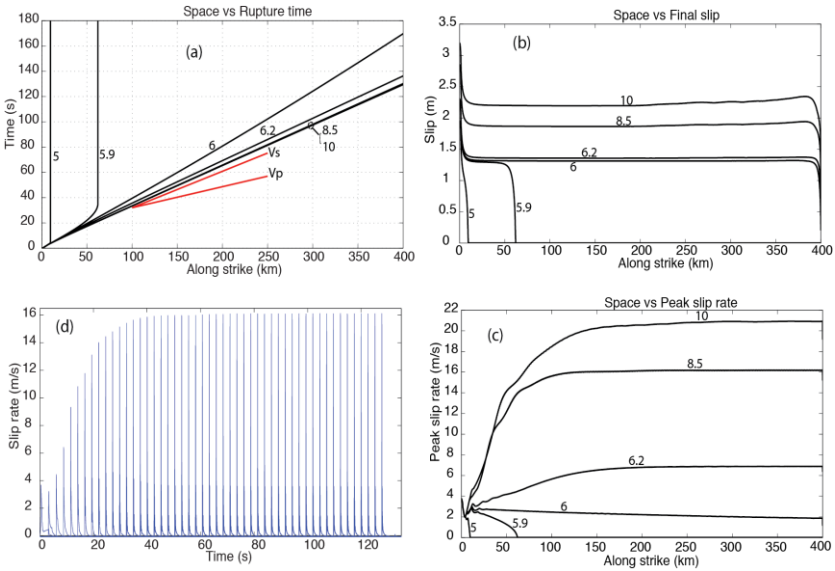


Figure 10. Rupture time (a), final slip (b) and peak slip rate (c) along the inplane axis of a strike slip fault with rupture propagation length of 400km, for different fault width (W). The number next to the line specifies the fault width. Figure (d) shows slip-rate vs time at each 8km interval along the inplane axis for the model with fault width W=8.5km.

slip and peak slip rate are affected. Interesting, at rupture distance L>>W when the rupture is self-sustained, rupture propagates with a steady-state velocity pulse, i.e., the slip-rate pulse travels without altering its shape and amplitude, as shown in Figure 10d. This steady-state mechanism suggests that the cohesive zone length in the rupture front remains constant.

It is clear that the main mechanism dominating this kind of fault is due to the effect of stopping phases, as explained above. When this process occurs in a very narrow fault, the S-wave stopping phase reaches the rupture front early, and they are loaded with enough energy to arrest the rupture. But when the stopping phase reaches the rupture front late, the rupture front is already self-sustained, producing a complicated interaction between the stopping phase and the pulse dominated in the rupture front; consequently, the pulse becomes steady state.

From an energetic point of view, initially the fault is loaded with elastic energy that is dissipated during rupture propagation. The energy dissipated

during rupture increases with L, whereas the available elastic energy is proportional to W. For L>>W, the dissipated energy becomes larger than the available elastic energy, leading to an eventual arresting of the rupture.

Remarks

Here we have described the numerical algorithms of two well known methods to represent fault discontinuity for spontaneous rupture dynamic calculation: the so-called traction at split-node (TSN) scheme and the inelastic-zone stress glut (SG) method. The main goal of this work is to introduce to the reader the conceptual implementation of these methods and its application in a simple test problem. For academic purpose, in appendix we provide the TSN implementation in a 1D wave equation that includes a matlab script, so the reader can follow the formulation and build his/her own code.

Advanced papers referred in Introduction are recommended to read for applications of these methods for different type of problems. There are recent development of fault representation and wave propagation technique not cited before, such us those used in Finite Volumes (FV) methods (e.g. [52,53]) and high order discontinues Galerkin (DG) methods (e.g. [54,55]). The nature of the fault representation in these methods is different than the TSN and SG method described here. The VF and DG incorporate formulations of fluxes to exchange information between the two surfaces of contact by solving the Riemann problem (e.g. [56]). These methods appear to be elegantly powerful and suitable to solve problems in extreme complex media and fault geometries. Another new generation algorithms emerging recently are the so-called adaptive mesh refinement formulations (e.g. [57]). Since rupture dynamic problems require to solve small scale in space and time during rupture propagation, these adaptive mesh algorithms appear to be the future application for this kind of problems.

Acknowledgments

The content of many part of this chapter has been published in our series of papers [14, 15 and 16] as specified in the text. Therefore I express my gratitude to the authors of these papers, Steven Day, Nadia Lapusta and Yi Liu. I also would like to thanks to Jean Paul Ampuero for enjoyable discussions on the last part of this chapter in which we discuss rupture dynamic on large aspect-ratio faults. Some of the simulations were done at the Swiss National Supercomputing Center (CSCS), under the production project “Development of Dynamic Rupture Models to Study the Physics of Earthquakes and Near-Source Ground Motion.

Appendix

Numerical implementation of the traction at split-node (TSN) fault representation in a 1D Elastodynamic equation for rupture dynamic problems

Let assume the fault plane is perpendicular to the *z* axis and located at *z*=0. To simplify the problem, we will implement the mixed boundary condition in a 1D wave equation, so all the fields depend only on *z*. This reduces to the condition that exactly the same thing is happening at every points along an infinitely fault plane.

Let use the velocity-stress form of the elastodynamic equations, in which the velocity *v(z,t)* and shear stress $\tau(z,t)$ are the dependent variables:

$$\frac{\partial v}{\partial t} = \frac{1}{\rho}\frac{\partial \tau}{\partial z} \qquad \text{[A1]}$$

$$\frac{\partial \tau}{\partial t} = \mu \frac{\partial v}{\partial z} \qquad \text{[A2]}$$

Where μ is the shear module and ρ the density.

Let use the standard staggered grid finite difference for the spatial discretization of the equation (Figure A1). The fault normal is in the z direction and located at z=0. For simplicity, even un-realistic, we assume the existence of free-surface on the plus and minus domain of the discretization (see Fig. A1).

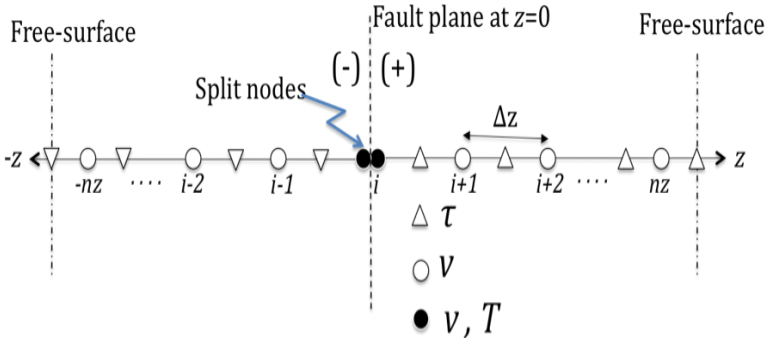


Figure A1. Staggered-grid discretization of the 1D elastodynamic equation with grid cells (split nodes) adjacent to the fault plane.

Approximation of spatial derivatives of equations A1 at the split nodes

Write separate equations for each side of the fault, taking into account the shear traction T acting at the interface, and its initial static equilibrium value T_0. Introduce the following one-sided difference approximations for τ, applicable to the plus and minus sides of the fault, respectively.

$$\left(\frac{\partial \tau}{\partial z}\right)_i^{\pm} = \pm \frac{\tau_{i\pm 1/2} - (T - T_0)^{-}}{0.5\Delta z} \qquad \text{[A3]}$$

The time derivatives of equation (A1) at time t approximates by second-order central differences

$$\left(\frac{\partial v}{\partial t}\right)_t = \frac{v(t + \Delta t/2) - v(t - \Delta t/2)}{\Delta t} \qquad \text{[A4]}$$

where Δt is the time step.

Approximation of spatial derivatives of equations A1 and A2 at interior grid points

Approximate the derivatives with a second-order spatial difference

$$\left(\frac{\partial \phi}{\partial z}\right)_i = \frac{\phi_{i+1/2} - \phi_{i-1/2}}{\Delta z} \qquad \text{[A5]}$$

where ϕ represents an arbitrary stress τ or velocity component v

Approximation of free-surface boundary condition

Positioning the free-surface at the stress node (see figure A1), we satisfy the free-surface condition setting stress at this node to be zero

$$\begin{aligned} \tau_{nz+1/2} &= 0 \\ \tau_{-nz-1/2} &= 0 \end{aligned} \qquad \text{[A6]}$$

Matlab script

Combining these equations (A1-A6) and the equations (5), (6), (11)-(15) on the fault described in the main text, we have wrote a matlab scripts at the end of this appendix. The matlab script is self explanatory. The example test

uses data assuming the fault is in the interface of two materials, plus side and minus side of the fault. The main feature modeled in this test are: 1) The test show evolution in time of slip velocity, slip and stress on the fault; 2) wave radiated from the fault toward the free-surface. Due to different material properties the test shows wave propagating at different speeds; 3) The effect of the free-surface on the radiated wave.

Data used in the matlab script test

Model geometry

L = 5000m; domain size (m) on each side of the fault

Δz = 25m; grid size

Material properties

c^+ = 4000.0 m/s; wave speed plus side of the fault

ρ^+ = 2670.0kg/m^3; density plus side of the fault

$\mu^+ = \rho^+ (c^+)^2$ shear module plus side of the fault (Pa)

c^- = 2000.0 m/s; wave speed minus side of the fault

ρ^- = 2670.0kg/m^3; density minus side of the fault

$\mu^- = \rho^- (c^-)^2$ shear module minus side of the fault (Pa)

Friction and initial stress

σ_n= 120e6 Pa; initial normal stress on the fault

μ_s =0.677; static friction coefficient

μ_d =0.525; dynamic friction coefficient

τ_0 =82.0e6 Pa; initial shear stress

d_0=0.4 m; critical slip distance (m)

Simulation time
$tmax = 1.5*L/c^+$;

Time discretization (CFL=0.5)

$dt = 0.5*\Delta z/c^{+}$; time step

nt = integer($tmax/dt$)+1; Number of time steps

Spatial discretization

nz = ingeger($L/\Delta z$)+1; Number of grid points

Suggestions for other tests

You can play with the grid size to evaluate convergence and numerical oscillations. Use the s Use the same data above, but for Δz =10m, 50m, 100m 200m 400m.

References

1. Andrews, D. J. 1976a, Rupture propagation with finite stress in antiplane strain, J. Geophys. Res. 81, 3575-3582.
2. Andrews, D. J. 1976b, Rupture velocity of plane-strain shear cracks, J. Geophys. Res. 81, 5679-5687.
3. Das, S., and K. Aki, Fault planes with barriers: A versatile earthquake model, J. Geophys. Res., 82, 5648-5670.
4. Day, S. M. 1982a, Three-dimensional finite difference simulation of fault dynamics: rectangular faults with fixed rupture velocity, Bull. Seismol. Soc. Am., 72, 705-727.
5. Day, S. M. 1982b, Three-dimensional simulation of spontaneous rupture: the effect of nonuniform prestress, Bull. Seismol. Soc. Am., 72, 1881-1902.
6. Olsen, K. B., R. Madariaga, and R. Archuleta 1997, Three Dimensional Dynamic Simulation of the 1992 Landers Earthquake, Science. 278, 834-838.
7. Dalguer, L.A; Irikura K; Riera J. And Chiu H.C. 2001, The Importance of the Dynamic Source Effects on Strong Ground Motion During the 1999 Chi-Chi (Taiwan) Earthquake: Brief Interpretation of the Damage Distribution on Buildings. Bull. Seismol. Soc. Am., 95, 1112-1127.
8. Dalguer, L.A; K. Irikura and J. Riera 2003a, Generation of New Cracks Accompanied by the Dynamic Shear Rupture Propagation of the 2000 Tottori (Japan) Earthquake, Bull. Seismol. Soc. Am., 93, 2236-2252.
9. Dalguer, L.A., H. Miyake, S.M. Day and K. Irikura 2008, Surface Rupturing and Buried Dynamic Rupture Models Calibrated with Statistical Observations of Past Earthquakes. Bull. Seismol. Soc. Am. 98, 1147-1161, doi: 10.1785/0120070134.
10. Peyrat, S.; Olsen, K. and Madariaga, R. 2001, Dynamic modeling of the 1992 Landers earthquake, J. Geophys. Res. 106, 26,467-26,482

11. Oglesby, D. and Day, S.M. 2001, Fault Geometry and the Dynamics of the 1999 Chi-Chi (Taiwan) Earthquake. Bull. Seismol. Soc. Am., 91, 1099-1111.
12. Aochi, H. and Fukuyama, E. 2002, Three-dimensional nonplanar simulation of the 1992 Landers earthquake. J. Geophys. Res., 107, NO. B2, 2035, 10.1029/2000JB000061.
13. Olsen, K., S.M. Day, L.A. Dalguer, J. Mayhew, Y. Cui, J. Zhu, V.M. Cruz-Atienza, D. Roten, P. Maechling, T.H. Jordan, D. Okaya and A. Chourasia 2009, ShakeOut-D: Ground motion estimates using an ensemble of large earthquakes on the southern San Andreas fault with spontaneous rupture propagation, Geophys. Res. Lett., 36, L04303, doi:10.1029/2008GL036832.
14. Day, S. M., L.A. Dalguer, N. Lapusta, and Y. Liu 2005, Comparison of finite difference and boundary integral solutions to three-dimensional spontaneous rupture, J. Geophys. Res., 110, B12307, doi:10.1029/2005JB003813.
15. Dalguer, L. A.,and S.M. Day 2006, Comparison of Fault Representation Methods in Finite Difference Simulations of Dynamic Rupture. Bull. Seismol. Soc. Am., 96, 1764-1778.
16. Dalguer, L. A., and S.M. Day 2007, Staggered-Grid Split-Node Method for Spontaneous Rupture Simulation. J. Geophys. Res., 112, B02302, doi:10.1029/ 2006JB004467.
17. Andrews, D. 1999, Test of two methods for faulting in finite-difference calculations, Bull. Seism. Soc. Am., 89, 931-937.
18. Brietzke, G. B., A. Cochard, and H. Igel 2007, Dynamic rupture along bimaterial interfaces in 3D, Geophys. Res. Lett., 34, L11305, doi:10.1029/2007GL029908.
19. Moczo, P., J. Kristek, M. Galis, P. Pazak, and M. Balazovjech 2007, The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion, Acta physica slovaca, 57(2), 177-406.
20. Oglesby, D., R. Archuleta, and S. Nielsen 1998, The three-dimensional dynamics of dipping faults, Bull. Seism. Soc. Am., 90, 616-628.
21. Oglesby, D., R. Archuleta, and S. Nielsen 2000, Earthquakes on dipping faults: the effects of broken symmetry, Science, 280, 1055-1059.
22. Aagaard, B., T. Heaton, and J. Hall 2001, Dynamic earthquake rupture in the presence of lithostatic normal stresses: Implications for friction models and heat production, Bull. Seism. Soc. Am., 91, 1765-1796.
23. Ma, S., S. Custodio, R. J. Archuleta, and P. Liu 2008, Dynamic modeling of the Mw 6.0 Parkfield, California, earthquake. J. Geophys. Res., 113, B02301, doi:10.1029/2007JB005216.
24. Duan, B. 2010, Role of initial stress rotations in rupture dynamics and ground motion:Acase study with implications for theWenchuan earthquake. J. Geophys. Res., 115, B05301, doi:10.1029/2009JB006750.
25. Galis, M., P. Moczo, and J. Kristek 2008, A 3-D hybrid finite-differencefinite-element viscoelastic modelling of seismic wave motion, Geophys. J. Int., 175, 153-184.
26. Barall, M. 2010, Home of FAULTMOD Finite-Element Software, WWW.FAULTMOD.COM.

27. Vilotte, J.-P., G. Festa, and J.-P. Ampuero 2006, Dynamic fault rupture propagation using nonsmooth spectral element method, Eos Trans. AGU, 87(52), FallMeet. Suppl., Abstract S52B-05.
28. Kaneko, Y., N. Lapusta, and J.-P. Ampuero 2008, Spectral element modeling of spontaneous earthquake rupture on rate and state faults: Effect of velocity-strengthening friction at shallow depths, J. Geophys. Res., 113, B09,317.
29. Galvez, P., J.-P. Ampuero, L. Dalguer, and T. Nissen-Meyer 2011, Dynamic rupture modeling of the 2011M9 Tohoku earthquake with unstructured 3D spectral element method, Eos Trans. AGU, 1(1), Fall Meet. Suppl., Abstract S24.
30. Day, S. M. 1991, Numerical simulation of fault propagation with interface separation (abstract), Eos Trans. AGU, 72, 486.
31. Dalguer, L. A. and S. M. Day 2009, Asymmetric Rupture of Large Aspect-ratio Faults at Bimaterial Interface in 3D. Geophysical Research Letters, 36, L23307, doi:10.1029/2009GL040303.
32. Ida, Y. 1972, Cohesive force across the tip of a longitudinal-shear crack and Griffith's specific surface energy, J. Geophys. Res., 77, 3796-3805.
33. Palmer, A. C., and J. R. Rice 1973, The growth of slip surfaces in the progressive failure of overconsolidated clay slopes, Proc. R. Soc. Lond., A332, 537.
34. Fukuyama E. and R. Madariaga 1998, Rupture dynamic of a planar fault in a 3D elastic medium: rate- and slip-weakening friction, Bull. Seismol. Soc. Am., 88, 1-17.
35. Madariaga R., Olsen K., and Archuleta R. 1998, Modeling Dynamic Rupture in a 3D Earthquake Fault Model, Bull. Seism. Soc. Am. Vol. 88, 1182-1197.
36. Harris, R. A., and S. M. Day 1999, Dynamic 3D simulations of earthquakes on en echelon faults, Geophys. Res. Letters, 26, 2089-2092.
37. Dalguer, L.A; K. Irikura and J. Riera, 2003b, Simulation of Tensile Crack Generation by 3D Dynamic Shear Rupture Propagation During an Earthquake. J. Geophys. Res., 108(B3), 2144, doi:10.1029/2001JB001738.
38. Dieterich, J.H. 1979, Modeling of rock friction, 1, Experimental results and constitutive equations. J. Geophys. Res., vol. 84. pp. 2161-2168.
39. Ruina, A. 1983, Slip Instability and State Variable Friction Laws, J. Geophys. Res., 88, 10359-10370.
40. Lachenbruch, A. H. 1980, Frictional heating, fluid pressure and the resistance to fault motion, J. Geophys. Res., 85, 6097-6112.
41. Mase, C. W., and L. Smith 1985, Pore-fluid pressures and frictional heating on a fault surface, Pure Appl. Geophys., 122, 583-607.
42. Mase, C. W., and L. Smith 1987, Effects of frictional heating on the thermal, hydrologic, and mechanical response of a fault, J. Geophys. Res., 92, 6249-6272.
43. Rice, J. R. 2006, Heating and weakening of faults during earthquake slip, J. Geophys. Res., 111(B5), B05311, doi: 10.1029/2005JB004006.
44. Barenblatt, G.I. 1959, The formation of equilibrium cracks during brittle fracture: General ideas and hypotheses, axially symmetric cracks, Applied Mathematics and Mechanics (PMM) 23, pp. 622-36.
45. Griffith, A. A. 1920, The phenomena of rupture and flow in solids, Phil. Trans. Roy. Soc., Ser. A, 221, 163-198.

46. Harris, R. A., M. Barall, R. Archuleta, E. M. Dunham, B. Aagaard, J. P. Ampuero, H. Bhat, V. Cruz-Atienza, L. Dalguer, P. Dawson, S. Day, B. Duan, G. Ely, Y. Kaneko, Y. Kase, N. Lapusta, Y. Liu, S. Ma, D. Oglesby, K. Olsen, A. Pitarka, S. Song, and E. Templeton 2009, The SCEC/USGS dynamic earthquake-rupture code verification exercise, Seismological Research Letters, 80(1), 119-126, doi:10.1785/gssrl.80.1.119.
47. Bizzarri, A. (2010), How to promote earthquake ruptures: Di_erent nucleation strategies in a dynamic model with slip-weakening friction, Bull. Seismol. Soc. Am., 100, 923-940, doi:10.1785/0120090179.
48. Day, S. M., and G. P. Ely 2002, Effect of a shallow weak zone on fault rupture: Numerical simulation of scale-model experiments, Bull Seismol. Soc. Am., 92, 3022-3041.
49. Moczo, P., J. O. A. Robertsson, and L. Eisner 2007, The Finite-Difference Time-Domain Method for Modelling of Seismic Wave Propagation. In Advances in Wave Propagation in Heterogeneous Earth, 421-516, Wu, R.-S., Maupin, V., eds., Advances in Geophysics 48, Dmowska, R., ed., Elsevier – Academic Press, doi: 10.1016/S0065-2687(06)48008-0.
50. Abramowitz, M., and I. A. Stegun (1964), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, U.S. Dept. of Commer., Natl. Inst. of Stand. And Technol., Gaithersburg, Md., 1964.
51. Pitarka, A. 1999, 3D elastic finite-difference modeling of seismic motion using staggered grid with nonuniform spacing, Bull. Seismol. Soc. Am., 89, 54-68.
52. Benjemaa, M., N. Glinsky-Olivier, V. Cruz-Atienza, J. Virieux, and S. Piperno 2007, Dynamic non-planar crack rupture by a finite volume method, Geophys. J. Int., 171, 271 285.
53. Benjemaa, M., N. Glinsky-Olivier, V. Cruz-Atienza, and J. Virieux 2009, 3-D dynamic rupture simulations by a finite volume method, Geophys. J. Int., 178, 541560, doi:10.1111/j.1365-246X.2009.04088.x.
54. de la Puente, J., J.-P. Ampuero, and M. Kaser 2009, Dynamic rupture modeling on unstructured meshes using a discontinuous Galerkin method, J. Geophys. Res., 114, B10, 302, doi:10.1029/2008JB006271.
55. Pelties, C; J. de la Puente; J.-P. Ampuero; G. Brietzke and M. Kaeser (2012) Three-dimensional dynamic rupture simulation with a high-order Discontinuous Galerkin method on unstructured tetrahedral meshes J. Geophys. Res., 117, B02309, doi:10.1029/2011JB008857.
56. LeVeque, R. 2002, Finite Volume Methods for Hyperbolic Problems, Cambridge University Press, Cambridge.
57. Kozdon, J. E. and Dunham, E.M. 2011, Adaptive Mesh Refinement for Earthquake Rupture Simulations. 2011 SIAM Conference on Mathematical & Computational Issues in the Geosciences. MS19 Computational Challenges in Earthquake Simulation. March 21-24, 2011, Long Beach, CA.

Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

The Mechanics of Faulting: From Laboratory to Real Earthquakes, 2012: 125-152
ISBN: 978-81-308-0502-3 Editors: Andrea Bizzarri and Harsha S. Bhat

5. The birth of forward models: From Coulomb criterion to cohesive force laws

Raúl Madariaga
Laboratoire de Géologie UMR CNRS 8538, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France

Abstract. The propagation of seismic ruptures along a fault subject to an initial stress distribution and a set of frictional parameters has been studied extensively over the years. When prestress is independent of time, rupture is relatively simple: it accelerates steadily to the terminal velocity and, under certain circumstances it can become supershear for mode II. For mode III, on the other hand, the terminal velocity is the shear wave speed. In this paper we review the most important results obtained in fracture mechanics and seismology for antiplane cracks in the first years of earthquake dynamics. First, we study a model of a rupture front that moves at constant speed with the load following it, as in rupture pulses. Then we study a model of a rupture front that appears spontaneously and propagates afterwards at variable speed. We show that these two types of rupture behave differently as speed increases. Steady propagating rupture pulses, like dislocations, can not approach the terminal speed. On the other hand spontaneous ruptures moving at variable speed under time independent load increase their speed steadily until they reach the shear wave speed. We then discuss seismic radiation, the generation of high frequency waves by seismic ruptures. We show

Correspondence/Reprint request: Dr. Raúl Madariaga, Laboratoire de Géologie UMR CNRS 8538, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France, E-mail: madariag@geologie.ens.fr

that radiation can only occur when rupture speed changes. Cracks moving at constant speed simply "pull" their static elastic field behind the rupture front without emitting seismic waves. We obtain a complete solution for radiation from an arbitrarily moving crack in 2D. We finally introduce friction following the original work by Ida who studied different slip weakening friction models. We show that the cohesive zone follows a simple scaling relation with a numerical coefficient that depends on the details of the slip weakening law.

1. Introduction

Earthquake are due to the fast propagation of shear ruptures along pre-existing seismic faults. Several studies in the late 60s, 70s and early 80s pioneered our understanding of seismic rupture and introduced simple models of faulting in two dimensions, typically using homogeneous distributions of stress and friction parameters. Classical examples of such models are the self-similar circular rupture model introduced by Kostrov (1966) and the circular crack solved numerically by Madariaga (1976). Extensive research then followed in order to properly pose the problem of propagation and radiation by a seismic rupture starting from the classical work by Kostrov (1964, 1966) and Eshelby (1969). Very soon it became clear that friction also played a fundamental role in the initiation, development of rupture and the healing of faults. The classical Coulomb model of a sudden drop in friction from a static to a kinematic coefficient led to an impasse, with infinite stress singularities and many other physical problems. The reason is that this model lacks an essential length scale needed to define a finite energy release rate near the rupture front. Slip weakening friction laws were introduced in dynamic rupture modelling by Ida (1972) and Andrews (1976a,b) for plane (2D) ruptures and by Day (1982b) for 3D fault models. A closely related friction law that is very amenable to computation was introduced by Palmer and Rice (1973) who found a complete analytical solution for the rupture front. These authors showed that slip weakening regularizes the numerical model of the rupture front, distributing stress and slip concentrations over a distance controlled by the length scale in the friction law. Other models of friction at low slip rates were studied in the laboratory by Dieterich (1978, 1979) and Ruina (1983), who proposed the model of rate- and state-dependent friction. Ohnaka and Kuwahara (1990), Ohnaka (1996) concluded that their experiments could be explained with a simple slip-weakening friction law. In fact, for many practical purposes, the rate-and-state and slip weakening friction laws can be reconciled remarking that both models contain a finite length scale that controls the behaviour of the rupture front. Extensive reviews on rupture dynamics until 1990 were published by Kostrov and Das (1989), Scholz (1989) and Freund (1990).

Recent studies of rupture processes for selected earthquakes have shed new light on our understanding of earthquake ruptures. These models suggest a complexity of the rupture process that the early models of rupture in a uniformly loaded medium were unable to explain. Although in the late seventies, Das and Aki (1977b), Mikumo and Miyatake (1978, 1979), Madariaga (1979) and Andrews (1980, 1981) pointed out the deficiencies of the classical crack models, it was not until the late 80s that good quality near-field strong motion data became available. By that time numerical methods had become mature and 3D dynamic models became common.

In this chapter we review what we believe are the most important results obtained in the field of earthquake rupture in the early stages of development of a fracture model of earthquakes. In order to keep the exposition simple we will restrict our study to antiplane cracks, but in-plane models can be studied with the same methods as long as rupture is sub-shear (see, Kostrov, 1975, Madariaga, 1977). In section 1 we review the early models of earthquake rupture and discuss the generation of seismic waves by an elastic shear fault model. The most important result is that cracks moving at constant speed do not radiate, only changes in the rupture front dynamics produce seismic waves. Then we introduce some of the most fundamental friction laws used in modelling earthquakes. Unfortunately, very few friction models can be solved analytically so that earthquake models have to resort to numerical methods even for the simplest models.

2. The shear crack model

From the mid sixties to the mid-eighties earthquake models were developed at a fast path, from dislocation to crack models without friction and, finally, fault models including friction. In this section we will briey review the simpler crack model before we delve into the problem of the interaction between rupture propagation and friction. We will deliberately keep the exposition at the level of basic results; details will be provided in the following section where we will discuss in greater detail a complete model of fracture and radiation.

Let us consider now a properly posed source model embedded in a homogeneous elastic model of the earth. Extension to more complex elastic media, including realistic wave propagation media poses no major technical difficulties except, of course, that in general media only numerical solutions are possible.

Consider the 2D elastic wave equation:

$$\rho \frac{\partial^2}{\partial t^2} \mathbf{u} = \nabla \cdot \sigma, \tag{1}$$

where **u**(**x,** *t*) is the displacement vector field, a function of both position **x** and time *t*, and ρ (**x**) is the density of the elastic medium. Associated with the displacement field **u** the stress tensor σ (x; *t*) is defined by

$$\sigma = \lambda \nabla \cdot \mathbf{u}\, \mathbf{I} + \mu \left[(\nabla \mathbf{u}) + (\nabla \mathbf{u})^T \right], \tag{2}$$

where λ(**x**) and μ (**x**) are Lamé's elastic constants, **I** is the identity matrix and superscript T indicates matrix transpose.

2.1. Boundary conditions on the fault

Assume that the earthquake occurs on a fault surface perpendicular to the axis *z*. Due to frictional instability a rupture zone propagates along the fault. At time *t* the rupture front is assumed to be located at position *l*(*t*). At a point inside the fault, *x* <= *l*(*t*), displacement and particle velocities are discontinuous so that

$$D_i(x,t) = u_i^+(x^+,t) - u_i^-(x^-,t) \tag{3}$$

is the slip across the fault. $x^{\pm}$ denote points immediately above or below the fault, and $u_i^{\pm}$ are the corresponding displacement components.

When slip *D* occurs there is a concomitant *change in the traction* σ_{iz} across the fault. This can be computed solving the wave equation (1):

$$\Delta\sigma_{iz}(x,t) = \Delta\Sigma_i[D] \qquad \text{for} \qquad x < \ell(t), \tag{4}$$

where $\Delta\Sigma$ [*D*] is a short hand notation for a functional of *D* and its temporal and spatial derivatives.

The problem posed above can be split into two simpler problems: an inplane or model II problem where the *x* component of displacement is discontinuous on the fault so that slip is parallel to the fault axis; and an antiplane problem (mode III) where the *y* component of displacement is discontinuous. The antiplane problem is simpler and does not require the use of complex variables. In the following we will use the antiplane fault as a simple model of an earthquake. For subshear inplane cracks, the antiplane solutions can be used to develop mode II solutions by a technique proposed by Kostrov (1975) and used by Madariaga (1977). Their method relies on the use of fixed coordinates. An alternative method for in-plane problems was proposed by Fossum and Freund (1975) who solved the inplane crack problem in coordinates moving with the rupture front. In the following we

will follow Kostrov's approach because it can be used to compute the full radiated field from a moving crack. A completely different method was proposed by Eshelby (1969) based on the use of certain properties of potential fields.

3. Study of a 2D antiplane rupture moving along a flat fault

The simplest possible model of a shear fault is that of the antiplane crack moving along a plane fault in an elastic medium (see Fig 1). This problem is sufficiently simple that it can be completely solved for general loads. The elastic field of an antiplane crack was first solved for a rupture moving at constant speed by Craggs (1960) and Yoffe (1951) for particular loads. Kostrov (1966) and Eshelby (1969) later showed that it had an exact solution for any load. The solution for plane shear cracks moving at constant speed was obtained by Fossum and Freund (1975). The super-shear in plane crack moving at constant speed was solved by Burridge et al (1979). The crack running at constant speed has however a number of limitations as an earthquake model. The most obvious one is that the load (stress drop inside the crack) has to follow the rupture front at the same speed as the front, the other is that cracks moving at constant speed produce no seismic radiation.

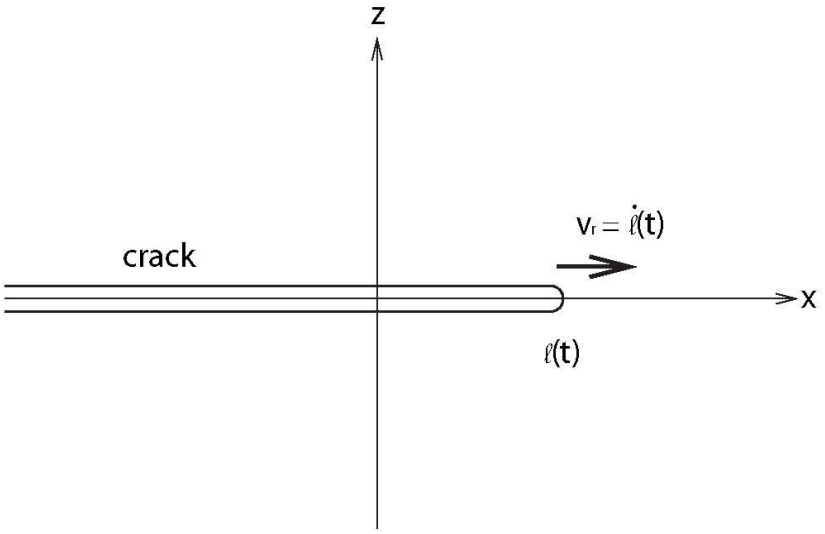


Figure 1. Geometry of the problem. A crack is running at constant or non-uniform speed along the *x*-axis. The interior of the crack $x < l$ slips due to stress drop.

3.1. The mode III antiplane crack model running at constant speed

Let us consider a simple crack moving at constant speed along a flat 2D fault as shown in Figure 1. The only component of displacement and stress that are relevant in this problem are u_y, σ_{yz} and σ_{yx}.

In coordinates fixed in the elastic medium with uniform properties (μ and ρ), the seismic waves generated by this crack satisfy the equation

$$\frac{1}{\beta^2}\frac{\partial^2}{\partial t^2}u_y(x,z,t) = \frac{\partial^2}{\partial x^2}u_y(x,z,t) + \frac{\partial^2}{\partial z^2}u_y(x,z,t) \tag{5}$$

where the shear wave speed $\beta = \sqrt{\mu/\rho}$.

If the crack moves at constant speed v_r in the direction of positive x **and** the load follows the crack at the same speed we can simplify the solution of the problem using moving coordinates. Following Ida (1972) and Burridge and Halliday (1971) this problem can be solved using the following transformation to steady state variables:

$$\begin{aligned} x - v_r t &\rightarrow X \\ \sqrt{1 - v_r^2/\beta^2}\, z &\rightarrow Z \end{aligned} \tag{6}$$

In this coordinate system the equation of motion (5) can be rewritten as

$$\frac{\partial^2}{\partial X^2}u_y(X,Z) + \frac{\partial^2}{\partial Z^2}u_y(X,Z) = 0 \tag{7}$$

In order to obtain 7 we used the property $v_r\partial_x = \partial_t$. Thus, the solution of the steady state crack reduces to the solution of Laplace equation for a static antiplane crack, for which there are many solutions. We solve this equation with crack boundary conditions on $Z = 0$:

$$\begin{aligned} u_y(X,0) &= 0 \quad \text{for} \quad X > 0 \\ \sigma_{yz}(X,0) &= \Delta\sigma(X) \quad \text{for} \quad -\infty < X < 0 \end{aligned} \tag{8}$$

These boundary conditions define a mixed boundary value problem that can be solved by a number of methods using complex potentials or integral equations. We adopt here the latter approach. For any slip distribution in the crack $X < 0$ we can write the following boundary integral that expresses the stress field on the fault in terms of the slip distribution:

$$\sigma_{yz}(X,0) = -\frac{\mu}{2\pi}\sqrt{1-v_r^2/\beta^2}\; P\int_{-\infty}^{0} \frac{\partial D_y(\xi)}{\partial \xi}\frac{d\xi}{X-\xi} \tag{9}$$

where $D_y(x) = u_y(X,0^+) - u_y(X,0^-)$ is the slip discontinuity inside the crack zone $X \le 0$. *P* indicates the principal value of the integral (see Ida, 1972 eq. 1). This expression can be derived from the representation theorem of Burridge and Knopoff (1964). The discrete version of (9) is usually known as the displacement discontinuity method and is widely used to solve crack problems in many areas, including mining, fracture, etc.

Equation (9) can be inverted in order to express the slip distribution as a function of the stress drop inside the crack. This solution is attributed to many of the creators of fracture dynamics. It is based on a theorem shown by Muskhelishvili (1953). The result (see Ida, 1972, eq 10) is

$$\dot{D}_y(X) = \frac{2v_r}{\pi\mu}\frac{1}{\sqrt{1-v_r^2/\beta^2}}\frac{1}{\sqrt{-X}}P\int_{-\infty}^{0}\frac{\sqrt{-\xi}\,\sigma_{yz}(\xi,0)}{X-\xi}d\xi \qquad \text{for} \quad X<0 \tag{10}$$

that expresses the sliprate $(\dot{D}_y)$ as a function of the stress drop inside the crack. This equation is rarely discussed in the seismological literature but it does not converge unless the integral on the RHS converges. We remark that for D_y to be finite, the integral over the stress has to have a finite value.

A companion expression useful to study friction relates the stress field outside the crack (for $X > 0$) to stress inside the crack:

$$\sigma_{yz}(X,0) = -\frac{1}{\pi}\frac{1}{\sqrt{X}}\int_{-\infty}^{0}\frac{\sqrt{-\xi}\,\sigma_{yz}(\xi,0)}{X-\xi}d\xi \qquad \text{for} \quad X>0 \tag{11}$$

(see Ida 172, eq. 13). We notice the remarkable similarity between the two expressions (10 and 11). We have now the complete solution for the antiplane shear crack propagating along a flat fault at constant speed. Similar expressions can be found for in-plane cracks (e.g. Fossum and Freund, 1975; Kostrov, 1975).

3.2. Stress intensity factor and energy release rate

Near the crack tip, close to $X = 0$, it is possible to simplify equations (10 and 11) under the condition that σ_{yz} is sufficiently well behaved. We find on $Z = 0$:

$$\sigma_{yz}(X,0) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{X}}K_{III} \qquad \text{as} \quad X \to 0^+ \tag{12}$$

$$\dot{D}_y(X,0) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{-X}} V_{III} \qquad \text{as} \quad X \to 0^- \tag{13}$$

in (12) the stress intensity factor K_{III} is defined as follows

$$K_{III} = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{0} \frac{\sigma_{yz}(\xi)}{\sqrt{\xi}} d\xi \tag{14}$$

and the velocity intensity factor in (13)

$$V_{III} = \frac{2v_r}{\mu} \frac{1}{\sqrt{1 - v_r^2/\beta^2}} K_{III} \tag{15}$$

The stress and velocity fields present inverse square root singularities in stress and velocity on opposite sides of the crack tip. These singularities constitute a natural property of a running crack, they appear because of the mixed boundary conditions on the two sides of the crack tip.

Slip near the tip of the crack can be computed integrating (13):

$$D_y(X) = \frac{2}{\sqrt{2\pi}} V_{III} \sqrt{-X} \qquad \text{as} \quad X \to 0^- \tag{16}$$

For a running crack we can now compute the energy release rate, that is, the amount of energy absorbed by unit advance of the crack tip, that we will call G_c. Using an argument by Kostrov and Nikitin (1966) and Palmer and Rice (1973) we find that

$$G_c = \frac{1}{2\mu} \frac{1}{\sqrt{1 - v_r^2/\beta^2}} K_{III}^2 \tag{17}$$

and a similar expression for mode II. These expressions lead to a paradox. When the rupture speed approaches the shear wave speed, velocity intensity V_{III}, slip amplitude and energy release rate G_c tend to infinity unless the stress intensity factor K_{III} tends to zero as the crack speed reaches the terminal speed.

4. A transient antiplane shear crack propagating at variable speed

The steady state solution discussed in the previous subsection explains some of the most general features of antiplane shear faults. The previous results can be generalized to a fault propagating at variable subsonic

(subshear) speeds. We will prove that the steady state solution remains valid in the case of an arbitrary stress drop, provided that the stress field varies smoothly near the rupture front. The solution for an antiplane shear crack propagating at variable speed was found by Kostrov (1966) using a very general procedure that he later extended to the solution for the propagation of an inplane shear crack (Kostrov, 1975). Actually the results of the previous section can be completely derived from the solution for a crack propagating at variable speed.

4.1. Exact solution for a semi-infinite shear fault

Let a semi-infinite shear fault as shown in Figure 1 extend from $-\infty$ to the current rupture front position $\ell(t)$. We study the stress and velocity field around this fault. We assume a homogeneous elastic body loaded internally by a stress field in the shear crack. The boundary conditions are

$$\begin{aligned} \sigma_{yz} &= -\Delta\sigma(x,t) \quad \text{for} \quad x < \ell(t) \\ u_y &= 0 \qquad\quad \text{for} \quad x \geq \ell(t) \end{aligned} \tag{18}$$

where $\Delta\sigma(x, t)$ is the traction change (usually called stress-drop) between the two-sides of the fault. As will be discussed later this stress drop has to be determined from a non-linear friction law. Here we will assume the simplest friction and consider a constant kinematic friction. In this case $\Delta\sigma(x, t)$ measures the difference between the initial static stress field and the kinematic friction. In order to set up the integral equation and its solution for some simple models of loading we assume initially that stress drop is a known function of space and time.

Because of the symmetry of the problem about the x axis, we can write the solution as a linear boundary integral equation relating stress outside the crack, on $\ell(t) < x < \infty$, to the known stress change in $x < \ell(t)$. For the applications we have in mind we assume that at time $t = 0$ the semi-infinite crack appears instantaneously along $x < 0$, so that $\ell(0) = 0$. The solution to this problem was found by Kostrov (1966) who proposed the following expression for stress outside the crack as a function of stress inside it:

$$\sigma_{yz}(x,t) = -\frac{1}{\pi}\frac{1}{\sqrt{x - x_\ell(x,t)}}\int_{x-\beta t}^{x_\ell(x,t)} \frac{\sqrt{x_\ell(x,t) - \xi}}{x - \xi}\Delta\sigma(\xi, t - (x-\xi)/\beta)\, d\xi \tag{19}$$

for $\ell(t) < x < \beta t$ and $\sigma_{yz}(x, t) = 0$ for $x > \beta t$.

The integral in (19) is calculated along the backward characteristic drawn from the point (x, t) where the stress field is calculated. The geometry is shown in 2. Characteristics of the wave equation (5) in the plane (ξ, τ) are lines slope $\pm 1/\beta$. As can be also seen from this figure, for a subsonic rupture front the stress field calculated from (19) is zero for all the points $x > \beta t$, located ahead of the shear wave front.

The intersection of the backward characteristic with the rupture front (x_ℓ, t_ℓ) can be found solving the implicit equations

$$x_\ell = \ell(t_\ell); \quad x - x_\ell = \beta(t - t_\ell) \tag{20}$$

Given a rupture history $\ell(t)$, this system may be solved analytically or numerically for x_ℓ, t_ℓ.

As shown by Ida (1973) and Madariaga (1983) it is also possible to solve exactly for the slip on the crack $x \le \ell(t)$. The relation is

$$D_y(x,t) = \frac{2}{\pi\mu} \int_{x-\beta t}^{x_\ell(x,t)} \int_{t_m}^{t_M} \frac{\Delta\sigma(\xi,\tau)}{\sqrt{(t-\tau))^2 - (x-\xi)^2/\beta^2}} d\tau d\xi \tag{21}$$

where the upper limit of the integral over τ is $t_M = t - |x - \xi|/\beta$. These are the characteristics through the current point (x, t) as shown in Figure 3. The

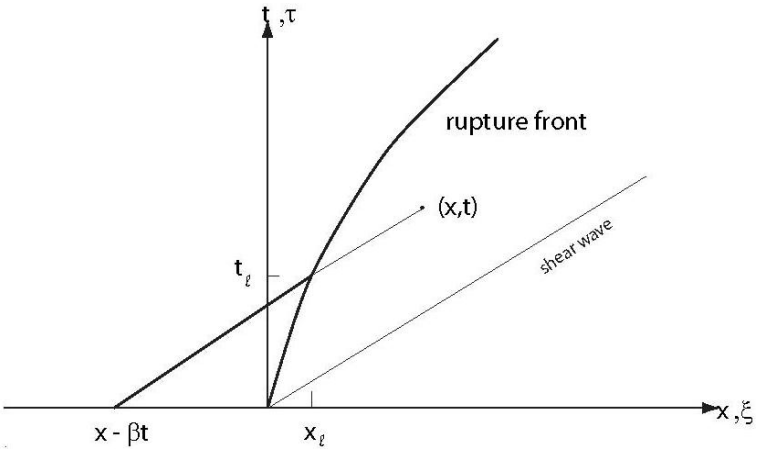


Figure 2. Computation of the stress field autside a non-uniformly moving crack. The stress at point x at time t is computed along the backward characteristic from (x, t).

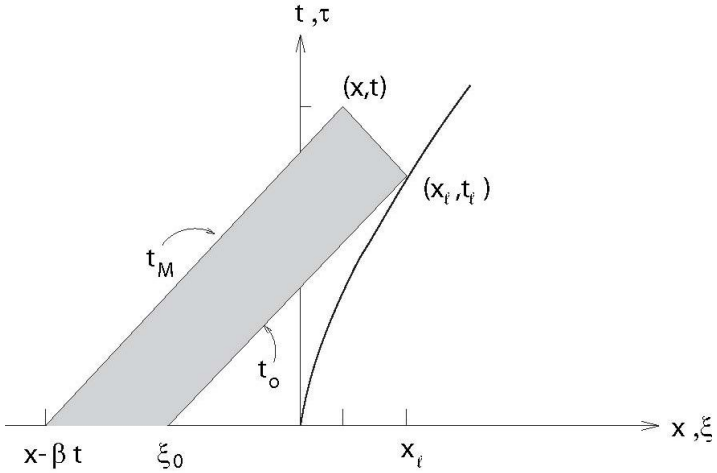


Figure 3. Computation of slip inside the crack at point (x, t) for an antiplane shear crack moving at non-uniform speed. The grey area is the region of integration for 21.

intersection (x_ℓ, t_ℓ) of the backward characteristic through (x, t) with the rupture front is calculated by

$$x_\ell = \ell(t_\ell); \quad x - x_\ell = -\beta(t - t_\ell) \tag{22}$$

The lower limit t_m = max(0, t_0) of the integral over τ in (21) is defined by the inter-section of the characteristic through x_ℓ, t_ℓ with the x-axis as shown also on Figure 3. The equation of the characteristic can be calculated using (20) in order to determine (x_ℓ, t_ℓ) so that

$$t_0(x, t, \xi) = t - \frac{2x_\ell(x, t) - x - \xi}{\beta} \tag{23}$$

Finally the point of intersection of this characteristic with the x axis is given by $t_0(x, t, \xi_0) = 0$, from which we get

$$\xi_0 = 2x_\ell(x, t) - x - \beta t \tag{24}$$

In the following applications we will need the slip velocity inside the crack instead of the slip. Taking the derivative of (21) we find (see appendix):

$$\dot{D}_y(x,t) = \frac{2}{\pi\mu}\int_{x-\beta t}^{x_\ell(x,t)}\int_{\max(0,t_0)}^{t_M}\frac{\Delta\dot{\sigma}(\xi,\tau)}{\sqrt{(t-\tau))^2-(x-\xi)^2/\beta^2}}d\tau d\xi$$
$$- \frac{1}{\pi\mu}\int_{x-\beta t}^{\xi_0}\frac{\Delta\sigma(\xi,0)}{\sqrt{t^2-(x-\xi)^2/\beta^2}}d\xi$$
$$- \frac{2}{\pi\mu}\frac{\dot{x}_\ell}{\sqrt{x_\ell-x}}\int_{\xi_0}^{x_\ell}\frac{\Delta\sigma(\xi,t_0)}{\sqrt{x_\ell-\xi}}d\xi \qquad (25)$$

This expression gives the slip velocity inside the fault for any distribution of traction change $\Delta\sigma(\xi,\tau)$. It contains three terms: (1) the first is radiation from changes in stress rate. Generally, the contribution of stress rate to radiation is neglected in most seismological studies, although it is an important term in the energy balance for seismic faults. (2) The second term in (25) is due to instantaneous rupture of the fault along the negative semi-infinite axis at time $t = 0$. For faults starting from a small initial region, this term can be neglected or replaced by another one. (3) The last term represents the radiation by the propagating rupture front, the most important term for the following discussion.

4.2. Stress and slip velocity for a time independent stress drop

These expressions for a general stress drop can be further simplified when stress drop is constant with time. This is one of the most common assumptions in earthquake simulations. For time-independent stress drop the first term in (25) is zero so that the slip velocity $\dot{D}_y$ reduces to the from the simpler expression:

$$\dot{D}_y(x,t) = -\frac{1}{\pi\mu}\frac{\dot{x}_\ell}{\sqrt{x_\ell-x}}\int_{\xi_0}^{x_\ell(x,t)}\frac{\Delta\sigma(\xi)}{\sqrt{x_\ell-\xi}}d\xi \qquad (26)$$

This is a simplification of (25) where we only consider the radiation emitted by the rupture front as it propagates.

4.3. Stress and velocity intensity factors and energy flow rate

A fundamental property of shear cracks is that near the rupture front stresses and slip velocities have universal features that can be derived for general loading. Let us first compute stress near the rupture front. We start from the expression (19) for the stress ahead of the crack tip. We assume here that stress is time independent.

$$\sigma_{yz}(x,0,t) = -\frac{1}{\pi}\frac{1}{\sqrt{x - x_\ell(x,t)}}\int_{x-\beta t}^{\ell(x,t)} \Delta\sigma(\xi)\frac{1}{\sqrt{x_\ell(x,t) - \xi}}d\xi \qquad (27)$$

for $x \to \ell(t)$, where $x_\ell(x, t)$ is the retarded position of the crack tip shown in Fig 2. In order to compute the stress intensity factor we have to replace x_ℓ by the current position of the rupture front t_ℓ. For that purpose we use the following approximation valid when $x \to \ell(t)$:

$$x - \ell(t) = (1 - v_r/\beta)(x - x_\ell(x,t)) \qquad (28)$$

where we used the property $\dot{\ell}(t) = v_r$.

We can now rewrite (27) in the simple universal form:

$$\sigma_{yz}(x,0,t) = \frac{1}{\sqrt{2\pi}}K_d[\ell(t)](x - \ell(t))^{-1/2} \qquad (29)$$

where the dynamic stress intensity factor is

$$\begin{aligned} K_d(x) &= \sqrt{1 - v_r/\beta}\, K_0(x) \\ K_0(x) &= \sqrt{\frac{2}{\pi}}\int_{x-\beta t}^{x} \frac{\sigma_{yz}(\xi)}{\sqrt{x - \xi}}d\xi \end{aligned} \qquad (30)$$

K_0 is the stress concentration of an equivalent static crack of the same length as the moving one. This separation into a velocity dependent factor and an intrinsic stress intensity factor is one of the most important characteristics of dynamic fracture mechanics. K_0 depends only on the stress drop inside the crack and it has no information about rupture history for subshear cracks. It is instructive to compare our expression (30) with that derived for a steady state crack (12). For a steady state crack the stress intensity factor does not include the universal velocity dependent term. For this reason we use the notation K_{III} for the stress intensity factor of cracks propagating at constant speed, and K_d for cracks with constant stress load.

Let us now turn to the slip velocity field in the vicinity of the crack tip. In the case of a time independent stress drop, the slip velocity is given by the last term in (26). Taking the limit of this term we find

$$\dot{D}(x,t) = \frac{1}{2\pi}\frac{K_0}{\mu}\frac{v_r}{1 + v_r/\beta}(\ell(\tau) - x)^{-1/2} \qquad (31)$$

where $\ell(\tau)$ is the retarded position of the rupture front. Thus the slip rate has an inverse square root singularity inside the crack that is the counterpart of the singularity in shear stress. Changing to the current position $\ell(t)$ using (28) we get asymptotically for $x \to \ell(t)$.

$$\dot{D}(x,t) = \frac{1}{2\pi}\frac{K_0}{\mu}\frac{v_r}{\sqrt{1+v_r/\beta}}(\ell(t)-x)^{-1/2} \tag{32}$$

we can now define the velocity intensity factor

$$V_d(x) = \frac{K_0(x)}{\sqrt{2\pi}\mu}\frac{v_r}{\sqrt{1+v_r/\beta}} \tag{33}$$

which, just as K_d, separates into a term that depends on the instantaneous rupture speed and another that depends only on the stress drop inside the crack.

Finally we compute the energy release rate as for steady state cracks:

$$\begin{aligned} G_c(x) &= \frac{1}{2\mu}\frac{1}{\sqrt{1-v_r/\beta}}K_d^2 \\ &= \frac{1}{2\mu}\sqrt{\frac{1-v_r/\beta}{1+v_r/\beta}}K_0^2 \end{aligned} \tag{34}$$

This expression shows the well-known fact that for a transient crack with static load, the energy release rate decreases to zero as the crack approaches the terminal speed β for antiplane cracks. It is important to realize that this is not a universal property of cracks, only of those that grow under static load. For a steady crack, on the other hand, G_c does not approach zero as the speed increases, it actually increases without limit as $v_r \to \beta$.

4.4. Why is steady state intensity factor different from Kostrov's

Steady state propagating cracks assume that the rupture front moves at constant speed **and** that the stress field moves at constant speed behind it. That is stress drop in a steady state crack is

$$\Delta\sigma(x,t) = \Delta\sigma(x - v_r t) \tag{35}$$

using the expressions for constant rupture speed $\ell(t) = v_r t$ we find

$$x_\ell(x,t) = \frac{v_r}{\beta}\frac{\beta t - x}{1 - v_r/\beta} \tag{36}$$

And $t_\ell\,(x,\,t) = x_\ell(x,\,t)/v_r$. Thus when $x \to v_r t$, $x_\ell \to x$.

Inserting these expression in (27) we get

$$\sigma_{yz}(x,0,t) = -\frac{1}{\pi}\frac{\sqrt{1-v_r/\beta}}{\sqrt{x-v_r t}}\int_{x-\beta t}^{x}\Delta\sigma\left((1-v_r/\beta)(x-\xi)\right)\frac{1}{\sqrt{x-\xi}}d\xi \tag{37}$$

for $x \to \ell(t)$ where $\tau = (x_\ell - \beta t)/v_r$.

We can now integrate (37) using the change of variables

$$\xi \to \zeta = (1-v_r/\beta)(x-\xi) \tag{38}$$

so that

$$\begin{aligned}\sigma_{yz}(x,0,t) &= -\frac{1}{\pi}\frac{1}{\sqrt{x-v_r t}}\int_{-\beta t/(1-v_r/\beta)}^{0}\Delta\sigma(\zeta)\frac{1}{\sqrt{\zeta}}d\zeta \\ &= -\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{x-v_r t}}K_{III}^2\end{aligned} \tag{39}$$

Thus, steady state cracks behave differently from transient cracks in which the load (stress drop) is only a function of position. An example of steady state crack is a pulse-like crack that propagates at constant speed. Stress intensity for such a crack is invariant with rupture speed, so that in contrast with classical crack solutions it can not approach the terminal speed. In general then, crack-like solutions approach the terminal speed, pulses do not.

5. Radiation from an antiplane shear crack propagating at variable speed

The steady state solution discussed earlier in this paper has a fundamental defect: it does not produce seismic waves, so that in fact it can not be used directly to model earthquakes. A crack moving at constant speed does not radiate, just like an electric charge moving at constant speed does not produce electromagnetic waves. In order to produce seismic waves the rupture front of the crack must change. Let us recall (34); this expression can be considered as an equation of motion for the rupture front. It determines

rupture speed (v_r) as a function of the load represented by K_0 and the energy release rate G_c required by the rupture front to grow. If stress intensity and energy release rate are constant, the crack moves at constant speed and it does not radiate seismic waves.

Seismic waves are emitted when rupture velocity changes rapidly as a consequence of changes in either G_c or K_0. As discussed by Madariaga (1983) these two changes correspond to the usual concept of barrier and asperities used in seismology to explain rupture heterogeneity. Let us consider a simple example of seismic radiation studied initially by Eshelby (1969) for antiplane cracks and by Madariaga (1977) for inplane and antiplane cracks. We assume that a crack is running at constant rupture velocity v_r and that at a certain point x_{stop}, at time t_{stop} it suddenly stops because it meets an unbreakable barrier; a place on the fault where G_c is effectively infinite. When the crack stops, the stress intensity factor K_d suddenly increases from $K_d = \sqrt{1 - v_r/\beta} K_0$ to K_0, where K_0 defined in (30) is the so-called zero-speed rupture velocity. In other words when the crack is running at high speed its stress intensity factor is reduced by the Lorenz contraction term $(\sqrt{1 - v_r/\beta})$, but when it stops the Lorenz term no longer reduces the stress intensity. The radiation from the sudden arrest of the crack, or stopping phase, may be easily computed by subtracting to a steadily propagating crack another one that suddenly starts from (x_{stop}, t_{stop}) with a stress intensity that has the opposite sign. The stopping phase can be computed in a number of ways, as shown by Eshelby (1969) using potential methods and Madariaga (1977) using complex analysis.

The wave emitted by the sudden arrest of the crack is

$$\dot{u}_y(R,t) = -\frac{1}{\sqrt{2\pi}} \frac{K_0}{\mu} \frac{v_r}{1 - v_r/\beta \cos\theta} \frac{\sin\theta/2}{\sqrt{R}} H(t - R/\beta) \tag{40}$$

where, see figure 4, R and θ are cylindrical coordinates centred on the point where the crack stopped (x_{stop}, 0), time is measured from the time of arrest. The left hand side is the velocity produced by the passage of the stopping phase. Across the stopping phase wave front, velocity suddenly jumps from the value given by (40) to zero. The different terms in (40) are easy to understand. The amplitude of the stopping phase is proportional to the zero-speed stress intensity factor K_0. The third term is the directivity due to the motion of the crack front. The following term is the radiation pattern divided by the inverse square root typical of two dimensional waves. Finally the time dependence is the Heaviside function moving with the shear wave speed. The radiation pattern may seem surprising because it is different from that of a simple double couple. The reason it has this peculiar form is that the crack

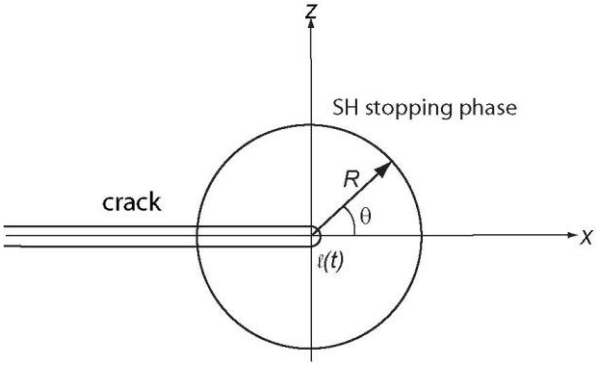


Figure 4. Emission of an stopping phase by a rupture front that suddenly stops at the arrest point ℓ(t).

front separates a forward domain (θ = 0) where velocity is zero from the interior of the crack where slip velocity has an inverse square root singularity. The sine function has opposite sign on the two sides of the fault ($\theta = \pm\pi/2$).

In the frequency domain a sudden jump in velocity means that the Fourier spectrum of ground velocity has an ω^{-1} behaviour at high frequencies. We can easily integrate (40) to obtain ground displacement. In this case the stopping phase is associated with a slope discontinuity in displacement. The spectral amplitude of a change in slope is characterized by a high frequency spectral decay of the ω^{-2} type. This is the classical high frequency decay of seismic waves emitted by seismic sources (Aki, 1967, Brune, 1970). This is of course not the only way to produce an ω^{-2} decay at high frequencies but it is a very good candidate for it. Madariaga (1977) associated the ω^{-2} asymptote as the origin of Brune's model of high frequency spectral decay. He showed that when a circular crack that was growing at high subsonic speed suddenly stops, it will produce ω^{-2}-like stopping phases emitted from the border of the circular crack. Andrews (1980) proposed an alternative model in which high frequency waves are controlled by a rupture front moving at constant speed on a fault that contains heterogeneities that have a particular power spectrum. Andrews and later work did not consider the stopping phases that would be radiated if rupture were to stop.

5.1. Seismic radiation by a crack moving at variable speeds

Madariaga (1983) showed that the high frequency radiation produced by crack moving at variable speed can be computed exactly as long as stress

does not change with time. When stress changes with time there are additional terms that can be computed, but they are not as easy to interpret. As the rupture front advances at constant or smoothly varying speed it produces a local stress field near rupture front. Away from the plane of the fault, however, the stress field is very smooth when the crack moves with slowly varying speed. The rupture front only produces high frequencies when the rupture velocity changes. We can find a complete solution for the seismic waves radiated from this crack using the similar approach to that used by Kostrov (1966) to solve for the slip rate on the plane of a crack that moves at variable speed (25).

The seismic waves emitted by the rupture front can be written in the very simple form:

$$\dot{D}_y(x,z,t) = \frac{1}{\sqrt{2\pi}} \frac{K_0(\ell)}{\mu} \frac{\dot{\ell}}{1-\dot{\ell}/\beta\cos\theta} \frac{\sin\theta/2}{\sqrt{R}} \tag{41}$$

In this expression the velocity field observed at a point (x, z, t) comes from the intersection of the backward influence cone drawn from the current observation point with the rupture front (see Figure 5). The point of intersection can be computed by solving the simultaneous equation

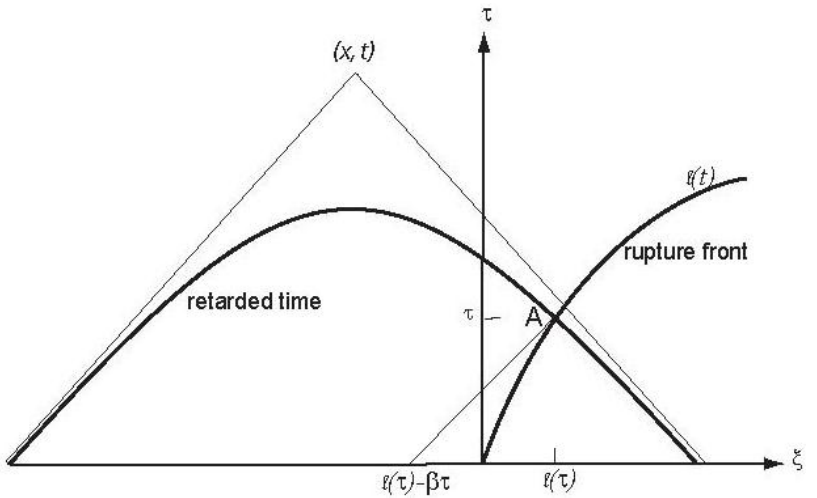


Figure 5. Computation of the velocity field emitted by a moving crack. Radiation reaching the point (x, z) at time t comes from the rupture front situated at point A. The stress intensity factor is computed along the backward characteristic issued from A.

$$\beta(t-\tau) = \sqrt{(x-\ell)^2 + z^2} \quad \text{and} \quad x = \ell(\tau) \tag{42}$$

for τ and $\ell(\tau)$. This is a very non-linear equation but it is not difficult to solve. Once ℓ is known we can compute $K_0(\ell)$, the distance R and radiation angle θ.

Expression (41) is a simplification of a more general expression derived by Madariaga (1983). Here we only consider the radiation (41) emitted by the rupture front as it propagates. There are other terms that take into account radiation by the initial stress on the fault, but in general those terms produce weaker seismic waves. In the following we consider that the dominant term is produced by the motion of the rupture front.

We notice first that when rupture is moving at constant speed $\dot{\ell} = v_r$ is constant. So that the velocity field of a crack moving at constant speed is just the static field of the crack that is dragged by the rupture front multiplied by a factor proportional to rupture speed and inversely proportional to directivity. As we already explained earlier, no high frequency waves are emitted by a crack moving at constant speed. Seismic waves are generated only when the crack front speed $\dot{\ell}$ changes rapidly.

6. Slip weakening friction

In the previous sections we reviewed the basic properties of a simple antiplane crack moving at variable speed in an almost arbitrary time and space variable stress field. The radiation emitted by the crack is simply related to the stress concentration at the crack tip. It is unlikely that such a simple result can be extended to cracks with friction because friction is intrinsically non-linear. There is however one situation in which simple results can be obtained. If the region where "slip-weakening" occurs is small compared to any other dimensions of the fault, the end zone can be studied independently of the rest of the crack. This is the case we will study in the following.

A crucial assumption in seismic source dynamics is that traction across the fault is related to slip at the same point through a *friction law* that can be expressed in the general form

$$T(D, \dot{D}, \theta_i) = T_{\text{total}} \tag{43}$$

so that friction T is a function of at slip, slip rate $\dot{D}$ and several state variables denoted by θ_i. For more details on state variables see Dieterich (1978, 1979),

Ruina (1983), but see also Ohnaka (1996) for an alternative point of view. The traction that appears in the friction law is the total traction T_{total} on the fault which can be expressed as the sum of pre-existing stress $T^0(x)$ and the stress change ΔT due to slip on the fault. The pre-stress is caused by tectonic load of the fault and will usually be a combination of purely tectonic loads due to internal plate deformation, plate motion etc., plus the stress field left over from previous seismic events on the fault and its vicinity.

In the early days of dynamic earthquake modelling the most frequently used friction law were the simple slip weakening laws proposed by Ida (1972). It is an adaptation to shear faulting of the Barenblatt-Dugdale friction laws used in hydro-fracturing and tensional (mode I) cracks. In this friction law, slip is zero until the total stress reaches a peak value (yield stress) that we denote by T_u. Once this stress has been reached, slip D starts to increase from zero and $T(D)$ decreases linearly to T_f as slip increases:

$$\begin{aligned} T(D) &= (T_u - T_f)\left(1 - \frac{D}{D_c}\right) + T_f \qquad \text{for} \qquad D < D_c \\ T(D) &= T_f \qquad\qquad\qquad\qquad \text{for} \qquad D > D_c, \end{aligned} \tag{44}$$

where D_c is a characteristic slip distance and T_f is the residual friction at high slip rate, sometimes called the "kinematic" friction. There is considerable discussion in the literature about how large this residual friction is. Many authors following the observation that there is a very broad heat flow anomaly across the San Andreas fault in California have proposed that faults are "weak", meaning that T_f is close to zero. Other authors propose that kinematic friction is high and faults are strong. We can not go into any details about this discussion here, interested readers may consult the papers by Scholz (2000) and Townend and Zoback (2000). For most applications of earthquake dynamics, only stress change is important so that without loss of generality we can assume that $T_f = 0$ in much of the following.

The slip weakening friction law (44) was used in numerical simulations of rupture starting with Andrews (1976a,b), Day (1982b) and many others. In many earlier studies of earthquake dynamics, a simpler version of (44) was used in which D_c was effectively zero. This numerical version of slip-weakening was called the Irwin criterion by Das and Aki (1977a) and was widely used by many authors although it is obviously grid-dependent (see, e.g. Virieux and Madariaga, 1982).

Once slip is larger than the slip weakening distance D_c, friction becomes a function of slip rate $\dot{D}$ and one or more state variables that represent the memory of the interface to previous slip. A very simple *rate dependent*

friction law was proposed by Knopoff and Burridge (1968) and has been used in simulations by Cochard and Madariaga (1986):

$$T(\dot{D}) = T_s \frac{V_0}{V_0 + \dot{D}} + T_f, \tag{45}$$

where V_0 is a characteristic slip velocity and $T_s \leq T_u$ is the limit of friction when slip rate decreases to zero. The applicability of rate weakening to seismic ruptures is much more controversial than slip weakening, although there is plenty of indirect evidence for its presence in seismic faulting. Heaton (1990) proposed that it was the cause of short rise times; rate-dependence at steady slip velocities is also an intrinsic part of the rate and state friction laws proposed by Dieterich (1978) and Ruina (1983). Those friction laws are very important at low slip rate, but at high rates they are very similar to slip weakening.

6.1. Steady propagation of a seismic rupture under friction

Ida (1972) and Palmer and Rice (1973) assumed that the state of stress near the crack tip is controlled by a far field state of stress and that friction acts locally to reduce stress and velocity intensity near the crack tip to zero. The idea is to separate the study of large scale slip-rate on the fault from the study of the effect of friction inside so-called process zone. The dimensions of the process zone are assumed to be small compared to other length scales in the crack problem. This is a boundary layer approach that separates the external field from the local process zone. A more general approach is possible but it requires the use of numerical methods because the problem is very non-linear. This is the approach used by Andrews (1976a,b) and later authors to study the growth of a plane crack.

We adopt as external field away from the crack tip the simple steady state crack solution presented in equations (12) and (13). The slip rate behind the process zone is

$$\dot{D}_y(X,0) = \frac{1}{\sqrt{2\pi}} \frac{2v_r}{\mu} \frac{K_{III}}{\sqrt{1 - v_r^2/\beta^2}} \frac{1}{\sqrt{-X}} \qquad \text{for} \quad X < 0 \tag{46}$$

We use the notation K_{III} for the stress intensity factor in order to indicate that this is the stress intensity factor for a steady running crack.

Inside the process zone a cohesive stress field appears due to the friction law. Let this stress be $\sigma_{zy}(X)$. This stress field moves behind the rupture front

at constant speed v_r. The slip rate distribution produced by the cohesive stress can be computed with equation (10) so that the total slip rate is:

$$\dot{D}_y(X) = \frac{1}{\sqrt{2\pi}} \frac{2v_r}{\mu} \frac{1}{\sqrt{1 - v_r^2/\beta^2}} \frac{1}{\sqrt{-X}} \left[K_{III} - \sqrt{\frac{2}{\pi}} \int_{-\infty}^{0} \sqrt{-\xi} \frac{\sigma_{yz}(\xi, 0)}{X - \xi} d\xi \right] \quad (47)$$

from which we can derive a first condition for the cohesive zone noting that the total velocity intensity will disappear if and only if $\dot{D}_y(0) = 0$, that is if

$$K_{III} = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{0} \frac{\sigma_{yz}(\xi, 0)}{\sqrt{\xi}} d\xi \quad (48)$$

thus the stress field inside the cohesive zone must exactly balance the stress concentration that would prevail near the crack tip if there was no friction.

Since the intensity factor due to friction eliminates the stress intensity due to the external load it is evident that the energy release rate into the crack tip must be equal to the rate of work of the cohesive forces.

$$G_c = \frac{1}{\mu} \frac{1}{\sqrt{1 - v_r^2/\beta^2}} K_{III}^2 \quad (49)$$

and $G_c = \int (T_u - T_f) dD$ for simple slip weakening models. This is the same as equation (17), derived for a crack propagating at constant speed, but here G_c has a well defined form for each slip weakening model. (49) can also be shown using the energy flow method proposed by Kostrov and Nikitin (1970) or by the Rice's J-integral (see Rice, 1980 for more details).

Equation (47) together with the friction law (43) define a well-posed non-linear eigen-value problem that can be solved by the iterative method proposed by Ida (1972). His method is however very involved because he did not introduce the requirement (48) from the beginning. He used several expressions that are equivalent to the energy release rate G_c in his derivation. The problem leads to a non-linear eigenvalue problem that can only be solved numerically. Ida (1972) presented solutions for a number of models for the friction law.

There is one example of friction law that can be solved exactly without resorting to numerical methods. This is a model in which friction is assumed to be constant inside the process zone as shown by the broken line in Figure 6. For this model friction is

$$\sigma_{yz} = (T_u - T_f) \qquad \text{for} \qquad D < D_c \quad (50)$$

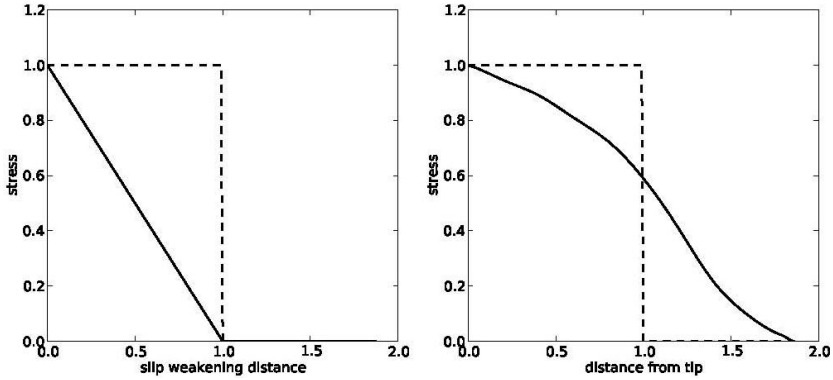


Figure 6. Slip weakening friction laws. On the left friction as a function of slip, on the right friction as a function of distance from the tip. Friction is normalized by $T_u - T_r$, slip by the slip weakening distance D_c and distance by L_c.

and 0 otherwise. Here D_c is the slip weakening distance and $(T_u - T_f)$ is the excess stress in the process zone.

The energy release rate for this friction model is $G_c = (T_u - T_f)\, D_c$. Since friction in constant in the process zone, we only need to determine the length L_c of the process zone using (48). Integrating we find

$$K_{III} = 2\sqrt{\frac{2}{\pi}}(T_u - T_f)\,\sqrt{L_c} \tag{51}$$

We can now use (49) to express L_c in terms of the slip weakening distance

$$L_c = \frac{\pi}{4}\frac{\mu}{T_u - T_f}\sqrt{1 - v_r^2/\beta^2}\; D_c \tag{52}$$

so that the size of the process zone scales linearly with D_c. Thus, D_c provides a length scale to the rupture process that was absent in the simple crack models.

For the linear slip weakening model there is no simple analytical approximation, but it is not difficult to compute numerically. The numerical solution in shown by the continuous line in Figure 6. In this figure slip is measured by D_c and stress by $(T_u - T_f)$. The figure on the right plots the friction inside the process zone where distance is measured by L_c for the constant stress model. From the figure on the right we get

$$K_{III} = 1.17(T_u - T_f)\,\sqrt{L_c} \tag{53}$$

Inserting into (49) we get

$$L_c = 0.735 \frac{\mu}{T_u - T_f} \sqrt{1 - c_r^2/\beta^2}\, D_c \tag{54}$$

again, the ratio between L_c and D_c depends only on the state of stress near the crack tip.

Palmer and Rice (1973) proposed a different approximation to the friction of the end-zone. Instead of solving the non-linear eigenvalue problem for a particular friction law $T(D)$, they assumed a particular form of the variation of σ_{xz} with distance from the crack tip:

$$\sigma_{xz} = (T_u - T_f)(1 - X/L_c) \tag{55}$$

where again L_c is the size of the cohesive zone. Inserting this expression into the expression for the stress intensity factor (48) we get

$$K_{III} = \frac{4}{3}\sqrt{\frac{2}{\pi}}(T_u - T_f)\sqrt{L_c} \tag{56}$$

Finally inserting this expression into (49), we get

$$L_c = \frac{9\pi}{16} \frac{\mu}{T_u - T_f} \sqrt{1 - c_r^2/\beta^2}\, \bar{D}_c \tag{57}$$

where $\bar{D}_c$ is an equivalent slip weakening distance defined by Palmer and Rice (1973). $\bar{D}_c$ is about one half the value of D_c as defined in (44) so that Palmer and Rice's model is very similar to that of linear slip weakening.

Since most recent models are computed numerically, the important result is that (52, 54 and 57) have the form of a scaling relation

$$L_c = C_c \frac{\mu}{T_u - T_f} \sqrt{1 - c_r^2/\beta^2}\, D_c \tag{58}$$

where C_c is a numerical coefficient of order 1. L_c to D_c ratios vary very little for different models of the friction law studied by Ida (1973).

We conclude that the length of the process zone L_c is roughly $\mu/(T_u - T_f)$ times the size of the slip weakening zone, reduced by the Lorenz contraction term. Since the ratio between μ and the peak stress drop $(T_u - T_f)$ is generally of the order of 10^3–10^4, the ratio L_c / D_c will be of the same order of magnitude. Another important consequence of (58), noted by Andrews

(1976), is that as the rupture velocity approaches the shear wave speed, L_c tends to zero if D_c has a fixed value independent of rupture speed. It must be noticed, though, that when the rupture velocity approaches the terminal speed, the description of the end-zone as an inverse squared-root singularity is no longer correct. Thus at the terminal speed numerical computations are required.

7. Conclusions

We have reviewed some of the most fundamental results on earthquake fracture dynamics derived from 1965 to 1985. Some of these results were derived adapting to shear faults results previously obtained for mode I cracks. Friction, however, did not play a significant role in mode I cracks, so that a number of original methods were developed very rapidly in the early 70s to deal with this problem. An interesting property of mode II shear cracks that we did not review here is supershear rupture velocities that will be reviewed in another chapter of this book. We put emphasis on the propagation of shear cracks at variable speed because cracks propagating at constant speed will not produce seismic waves, except in the very unlikely situation that the stress concentration near the crack tip changes without a concomitant change in rupture speed.

Thanks to improvements in speed and memory capacity of parallel computers it is no longer a problem to model the propagation of seismic ruptures along a fault, or a fault system, embedded in an elastic 3D medium. The enhanced computational power can be used to improve classical models in order to study 3D effects, It is nowadays possible to do dynamic forward models of earthquakes of any magnitude.

8. Appendix

Proof of expression (25):
Let us take the derivative of (21)

$$\dot{D}(x,t) = \frac{2}{\pi\mu}\int_{x-\beta t}^{x_\ell(x,t)} \frac{d}{dt}\left[\int_{t_m}^{t_M} \frac{\Delta\sigma(\xi,\tau)}{\sqrt{(t-\tau))^2-(x-\xi)^2/\beta^2}}d\tau\right] d\xi \qquad (59)$$

where the dot indicates a time derivative. Let us remark that the time derivatives with respect to the limits of the integrals over ξ do not contribute to the derivative because the integrals over τ for $\xi = x_\ell$ and $\xi = x - \beta t$ are zero.

The derivative of the integral over τ in (59) has to be computed very carefully in order to take into account the variation of its limits. We get

$$\frac{d}{dt}\left[\int_{t_m}^{t_M}\frac{\Delta\sigma(\xi,\tau)d\tau}{\sqrt{(t-\tau)^2-(x-\xi)^2/\beta^2}}\right] = \int_{t_m}^{t_M}\frac{\Delta\dot{\sigma}(\xi,\tau)d\tau}{\sqrt{(t-\tau)^2-(x-\xi)^2/\beta^2}} - \frac{d(t-t_m)}{dt}\frac{\Delta\sigma(\xi,t_m)}{\sqrt{(t-t_m)^2-(x-\xi)^2/\beta^2}} \quad (60)$$

The derivative at the upper point t_M does not contribute, because $d(t - t_M)/dt = 0$.

As seen from Figure 3, $t_m = 0$ for $\xi < \xi_0$ where ξ_0 was defined in (24). Thus $d(t - t_m)/d_t = 1$ for $\xi < \xi_0$. For $\xi > \xi_0$, on the other hand, $t_m = t_0$ defined in (23). In this case we get $dt_m/dt = 2\dot{x}_\ell/\beta$ and

$$\sqrt{(t-t_0))^2-(x-\xi)^2/\beta^2} = \sqrt{(2x_\ell-x-\xi)^2-(x-\xi)^2}/\beta = \frac{2}{\beta}\sqrt{x_\ell-x}\sqrt{x_\ell-\xi} \quad (61)$$

with this simplification we can finally derive (25).

9. Acknowledgments

R. Madariaga's work was supported by Agence Nationale pour la Recherche, program RISK 08 under contract DEBATE.

10. References

1. Aki, K. (1967). Scaling law of seismic spectrum, *J. Geophys. Res*. **72**, 1217-1231.
2. Andrews, J. (1976a). Rupture propagation with finite stress in antiplane strain, *J. Geophys. Res*. **81**, 3575-3582.
3. Andrews, J. (1976b). Rupture velocity of plane strain shear cracks, *J. Geophys. Res*., **81**, 5679-5687.
4. Andrews, J. (1980). A stochastic fault model. I. Static case, *J. Geophys. Res*., **85**, 3867-3877.
5. Andrews, J. (1981). A stochastic fault model. II. Time-dependent case, *J. Geophys. Res*. **87**, 10821-10834.
6. Brune, J. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res*. **75**, 4997-5009.

7. Burridge, R. and L. Knopoff (1964). Body force equivalents for seismic dislocations *Bull. Seismol. Soc. Am*. **54**, 1875-1888.
8. Burridge, R., and Halliday, R. (1971). Dynamic shear cracks with friction as models for shallow earthquakes, *Geophys. J. Roy. Astron. Soc*. **25**, 261-283.
9. Burridge, R., Cohn, G., and L.B. Freund (1979). The stability of a plane strain shear crack with finite cohesive force running at intersonic speeds, *J. Geophys. Res*., **84**, 2210-2222.
10. Craggs, J. W. (1966). On the propagation of a crack in an elastic-brittle solid, *J. Mechanics Phys. Solids*, **8**, 66-75.
11. Cochard, A., and Madariaga, R. (1996). Dynamic faulting under rate-dependent friction, *J. Geophys. Res* **142**, 419-445.
12. Das, S., and Aki, K. (1977a). A numerical study of two-dimensional spontaneous rupture propagation, *Geophys. J. Roy. astr. Soc*. **50**, 643-668.
13. Das, S., and Aki, K. (1977b). Fault plane with barriers: a versatile earthquake model, *J. Geophys. Res*. **82**, 5658-5670.
14. Day, S.M. (1982a). Three-dimensional finite difference simulation of fault dynamics: rectangular faults with fixed rupture velocity, *Bull. Seis. Soc. Am*. **72**, 795-727.
15. Day, S.M. (1982b). Three-dimensional simulation of spontaneous rupture: the effect of non-uniform prestress, *Bull. Seis. Soc. Am*. **72**, 1881-1902.
16. Dieterich, J. (1978). Time-dependent friction and the mechanics of stick-slip, *Pageoph* **116**, 790-806.
17. Dieterich, J. (1979). Modeling of rock friction. 1. Experimental results and constitutive equations, *J. Geophys. Res*. **84**, 2161-2168.
18. Eshelby, J. D. (1969). The elastic field of a crack extending non-uniformly under general anti-plane loading *J. Mechanics. Phys. Solids*, **17**, 177-199.
19. Fossum, A.F. and L. B. Freund (1975). Nonuniformly moving shear crack model of a shallow focus earthquake mechanism, *J. geophys. Res*., **80**, 3343-33347.
20. Freund, L. B., Dynamic Fracture Mechanics (1990). cambridge University Press, Cambridge, U.K.
21. Fukuyama, E., and Madariaga, R. (1998). Rupture dynamics of a planar fault in a 3D elastic medium: Rate- and slip-weakening friction, *Bull. Seismol. Soc. Am*. **88**, 1-17.
22. Heaton, T. (1990). Evidence for and implications of self-healing pulses of slip in earthquake rupture, *Phys. Earth. Planet. Int*. **64**, 1-20.
23. Ida, Y. (1972). Cohesive force across the tip of a longitudinal-shear crack and Griffith's specific surface energy, *J. Geophys. Res*. **77**, 3796-3805.
24. Kostrov, B. (1964). Self-similar problems of propagation of shear cracks, *J. Appl. Math. Mech*. **28**, 1077-1087.
25. Kostrov, B. V. (1966). Unsteady propagation of longitudinal-shear cracks, *Appl. Math. Mech*., **30**, 1241-1248.
26. Kostrov, B. V. (1975). On the crack propagation with variable velocity, *Int. J. fracture*, **11**, 47-56.
27. Kostrov, B. and Das, S. (1989). *Principles of Earthquake Source Mechanics*, Cambridge University Press.

28. Kostrov, B V and L.V. Nikitin (1970). Some general problems of mechanics of brittle fracture *Archiwum Mechaniki Stosowanej*, **22**, 749-776.
29. Madariaga, R. (1976). Dynamics of an expanding circular fault, *Bull. Seismol. Soc. Am*. **66**, 639-667.
30. Madariaga, R. (1977). High frequency radiation from crack (stress-drop) models of earth-quake faulting. *Geophys. J. Roy. astr. Soc*. **51**, 625-651.
31. Madariaga, R., (1979). On the relation between seismic moment and stress drop in the presence of stress and strength heterogeneity, *J. Geophys. Res*. **84**, 2243-2250.
32. Madariaga, R. (1983). High frequency radiation from dynamic earthquake fault models. *Ann. Geophys*., **1**, 17-23.
33. Mikumo, T. and Miyatake, T. (1978). Dynamical rupture process a three-dimensional fault with non-uniform friction and near-field seismic waves, *Geophys. J. Roy. astr. Soc* **54**, 417-438.
34. Mikumo, T. and Miyatake, T. (1979). Earthquake sequences on a frictional fault model with non-uniform strength and relaxation times, *Geophys. J. Roy. astr. Soc* **59**, 497-522.
35. Muskhelishvili, N. I., 1973, *Singular Integral Equations*, P. Noordhoff, Groningen, Netherlands.
36. Ohnaka, M. (1996). Non-uniformity of the constitutive law parameters for shear rupture and quasi-static nucleation to dynamic rupture: a physical model of earthquake generation processes, *Proc. Natl. Acad. Sci. USA* **93**, 3795-3802.
37. Ohnaka, M. and Kuwahara, Y. (1990). Characteristic features of local breakdown near crack-tip in the transition zone from nucleation to dynamic rupture during stick-slip shear failure, *Tectonophysics* **175**, 197-220.
38. Palmer, A. C. and J. R. Rice (1973). The growth of slip surfaces in the progressive failure of over-consolidated clay. *Proc. Roy. Soc. London*, **A 332**, 527-548.
39. Rice, J. R. (1980), The mechanics of earthquake rupture, in Physics of the Earth Interior: Proceedings of the International School of Physics Enrico Fermi, edited by A. M. Dziewonski and E. Boschi, Ital. Phys. Soc., Bologna.
40. Ruina, A. (1983). Slip instability and state variable friction laws, *J. Geophys. Res*. **88**, 10359-10370.
41. Scholz, C. (1989). *The Mechanics of Earthquake and Faulting*, Cambridge University Press.
42. Scholz, C. (2000). Evidence for a strong San Andreas fault, *Geology* **28**, 163-166.
43. Townend, J. and Zoback, M.D. (2000). How faulting keeps the crust strong, *Geology*, **28**, 399-402.
44. Virieux, J. and Madariaga, R. (1982). Dynamic faulting studied by a finite difference method, *Bull. Seismol. Soc. Am*. **72**, 345-369.
45. Yoffe, E. H. (1951). The moving Griffith crack, *Philos. Magazine*, **42**, 739-750.

Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

The Mechanics of Faulting: From Laboratory to Real Earthquakes, 2012: 153-207
ISBN: 978-81-308-0502-3 Editors: Andrea Bizzarri and Harsha S. Bhat

6. Models of earthquakes and aseismic slip based on laboratory-derived rate-and-state friction laws

Nadia Lapusta[1,2] and Sylvain Barbot[2]
[1]Division of Engineering and Applied Science, [2]Division of Geological and Planetary Sciences, California Institute of Technology, USA

Abstract. Faults accommodate slow tectonic loading through both earthquakes and slow slip. Here, we discuss fault models capable of reproducing the entire range of fault slip behaviors, fueled by the increasing stream of high-quality laboratory experiments, observational data, and computational resources. The success of the laboratory-based models opens a possibility of predictive physics-based modeling, in which a range of potential fault behaviors is uncovered. However, much remains to be done toward that goal.

1. Introduction

Fault processes involve both dynamic events – seismic slip perceived as earthquakes – and complex patterns of quasi-static (aseismic) slip. Understanding physics and mechanics of this behavior in its entirety is a fascinating scientific problem. However, even for the more pragmatic goal of understanding only the behavior of destructive large dynamic events, it is still important to consider the entire earthquake cycle, since aseismic slip and

Correspondence/Reprint request: Dr. Nadia Lapusta, Division of Engineering and Applied Science, Division of Geological and Planetary Sciences, California Institute of Technology, USA. E-mail: lapusta@its.caltech.edu

smaller events may determine where large earthquakes would nucleate as well as modify stress and other initial conditions before dynamic rupture. Moreover, large events are relatively rare; hence improving our understanding of earthquake physics has to be based in part on observations of smaller events and aseismic slip phenomena.

Many fundamental questions in earthquake science require understanding of both seismic and aseismic slip, and their interaction. What controls the spatio-temporal distribution of slip on faults? How do earthquakes nucleate and arrest? What can we tell about large destructive rare events from smaller ones? What is the stress state on the faults? What causes aftershock sequences? Which aspects of fault physics are relevant to ground motions and estimates of earthquake hazard?

One approach to addressing this set of interconnected questions is to (i) formulate fault models with constitutive relations that are based on laboratory experiments and theories of how fault materials deform under the wide range of conditions during the earthquake cycle, (ii) simulate the response of those models in terms of fault slip, and (iii) find the relevant models by comparing the simulated features with observations over a range of temporal scales, in terms of individual seismic events, event sequences, and aseismic slip.

Here, we discuss the progress on implementing this approach, fueled by the increasing stream of high-quality laboratory experiments, observational data, and computational resources, with the focus on rate-and-state fault models.

2. Rate-and-state friction laws as laboratory-based description of fault behavior

The rate-and-state friction framework incorporates the current "state-of-the-art" understanding of macroscopic frictional properties. This section summarizes the laboratory-based inferences about frictional resistance of nominally planar shear zones which can be either interfaces between two contacting surfaces or narrow layers of fault gouge. Discussing frictional resistance implies that the shear zone is compressed by the normal stress σ (positive in compression) and there is no opening. If the shear zone is permeated by fluids, the typical assumption is that the effect of fluids on frictional resistance can be represented by the effective normal stress [1]:

$$\bar{\sigma} = \sigma - p \tag{1}$$

where p is the fluid pore pressure. The relative shear motion of the two surfaces is called slip, and the time derivative of that is called slip velocity or

slip rate. In the following, the terms “slip velocity” and “slip rate” are used interchangeably.

2.1. Standard formulations based on low-velocity experiments

The standard, so-called Dieterich-Ruina, rate-and-state friction laws were developed to incorporate experimental observations [2–5] of frictional resistance of shear zones at slip rates of the order of 0.01-1000 μm/s. Such range is typically called “low-velocity”, since the velocities are small compared to the seismic ones of the order of 1 m/s. The laws model variations of frictional shear strength due to its dependence on slip rate and evolving properties of the contact population or shearing layer; these evolving properties are described by a state variable, or variables [4–10].

The rate-and-state friction laws contain features of simpler friction laws. The first systematic study of friction was done by Leonardo da Vinci in the fifteenth century, and then his findings were rediscovered 200 years later by Amontons [11]. Amontons found that the frictional force is independent of the size of the surfaces in contact and that friction is proportional to the normal load, relating, in modern terms, the frictional shear stress τ and the effective normal stress $\bar{\sigma}$ as

$$\tau = f\bar{\sigma}, \quad f = \text{constant}, \tag{2}$$

where *f* is the friction coefficient thought to be a constant independent of the sliding velocity or the accumulated slip. Coulomb, nearly 100 years later, was one of the researchers who tried to explain the properties of friction [11], as well as the observation that static friction is (often) higher than the kinetic friction, and the law (2), which is still widely used, became known as the Coulomb friction law. First explanations of the independence of the frictional force on the contact area were given by Bowden and Tabor [12, 13] in their work on friction of metals. They noted that, because of the surface roughness, the surfaces touch only in certain places, called “asperities”, so that the real area of contact is much smaller than the apparent one. Bowden and Tabor then derived the friction law (2) for ductile metals. This explanation was extended to materials with largely elastic asperity contact first by Archard [14], who represented surface roughness by hierarchical structure of elastic spheres, and then, in a more realistic model of a rough surface with a random distribution of asperity heights, by Greenwood and Williamson [15].

While the Coulomb law (2) gives a sufficient description of friction for many elementary applications, it is not suitable for studies of unstable frictional phenomena (stick-slip), including sequences of earthquakes, even if

one agrees to neglect finer features of frictional behavior discovered by experiments [2–5] and discussed below. The problem is that the simple Coulomb law does not allow for unstable slip between identical materials (unless in a configuration for which normal stress is altered by slip). That is why other formulations have been used to study unstable slip. They have the general form (2), but the coefficient of friction is no longer considered to be constant.

In purely velocity-dependent (also called rate-dependent) laws, the coefficient of friction depends on the slip velocity V, so that one writes:

$$\tau = f\bar{\sigma}, \quad f = f(V) \; . \tag{3}$$

Often, the function $f(V)$ is decreasing with increasing slip velocity V for rocks, as experiments suggest for steady sliding at different slip velocities V (e.g., [16, 17]) and consistently with the well-known concept of static-kinetic friction. Such dependence is called velocity weakening - the faster the slip is, the easier it is to slip. If function $f(V)$ is increasing, then the law describes velocity strengthening. Another set of friction laws in use is slip-dependent laws, in which the coefficient of friction f depends on slip. The simplest law of such type allows f to degrade linearly with slip until a certain amount of slip (called the slip-weakening distance and denoted by D_c or d_c) is reached; during the subsequent slip, the friction coefficient stays constant (e.g., [18]). Such laws were motivated by cohesive zone models of tensile fracture [19, 20]. They can also be viewed as a generalization of the static-kinetic concept of friction, but contain no account of rate dependence. Such laws cannot model gradual regain of strength without additional ingredients.

The purely velocity-dependent or slip-dependent laws were introduced as plausible descriptions that allowed simulations of spontaneous slip events. When the growing interest in the subject and advances in experimental techniques made it possible to study the frictional response in more detail, it turned out that the response possesses remarkable features that have intrinsically mixed velocity-dependent and slip-dependent aspects. These features were first documented in the works by Rabinowicz [21,22], and then they were rediscovered and quantified in the so-called velocity-jump experiments by Dieterich [2, 3], Ruina [4, 5], and others [23–29].

In the velocity-jump experiments (Figure 1), frictional resistance settles down to a constant value as sliding occurs at a constant velocity. Then, as a sudden jump in slip velocity is imposed, friction also instantaneously jumps, with the same sign of the change. This is called the positive instantaneous velocity dependence, or positive direct dependence, or positive direct effect. After the direct effect, the frictional response evolves, through a certain amount

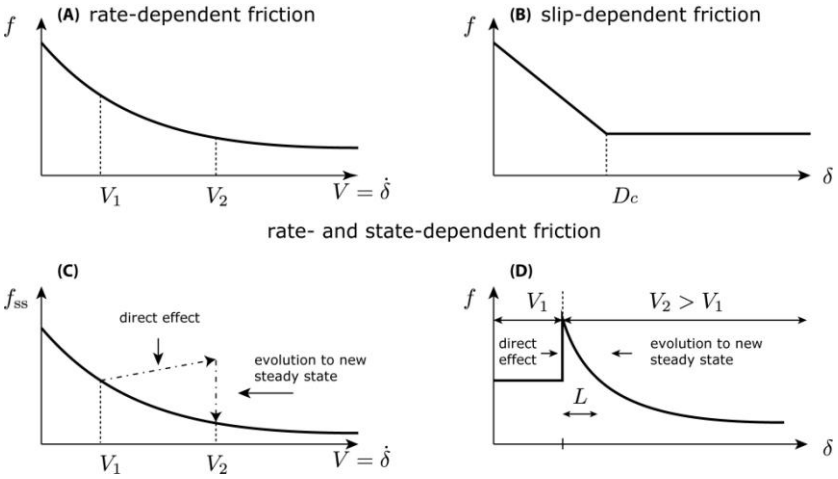


Figure 1. Schematic depiction of three types of friction laws. (a) In purely rate-dependent laws, the coefficient of friction f depends on the slip velocity V. Rate-weakening friction is illustrated, in which the resistance is weaker for higher slip rates. Rate-strengthening and non-monotonic responses have also been observed in experiments. (b) In slip-dependent laws, f depends on slip. The oftenused linear slip-weakening law is illustrated. Non-monotonic laws have also been proposed. (c-d) Rate-and-state laws simplify to a rate-dependent response (c) when resistance is reported for steady sliding at a given rate. However, jumps in slip rate bring about additional features, such as the positive direct effect and then evolution to the new steady-state friction level. The evolution occurs with slip and it is described through a state variable. Note that these features add up to the effective slip dependence at the crack front that resembles slip-dependent laws (b), as discussed in section 4.1.

of slip, to a new value appropriate for the new value of the slip velocity. The amount of slip necessary for the evolution is called the characteristic slip distance and is often denoted by L (also sometimes denoted by d_c or D_c). Hence the frictional resistance has an instantaneous and "eventual" velocity-dependent behavior, but exhibits transient slip-dependent behavior as well. (Over slip scales much larger than L, which is typically of the order of 1-100 μm in experiments, longer-term evolution of strength can also occur, due to wear effects and, in rapid slip, shear heating; see section 2.2 for the discussion of some of the additional effects).

How can such features be explained and quantified? The prevailing consensus is that they are caused by the combined dependence of friction strength on slip velocity and the "state" (principally, the maturity) of the asperity contact population. To describe the latter, a state variable

(or variables) is introduced. For constant in time effective normal stress $\bar{\sigma}$, shear resistance τ obeying rate-and-state friction laws is typically written as:

$$\tau = \bar{\sigma}[f_0 + a\ln(V/V_0) + b\ln(V_0\theta/L)] \ , \tag{4}$$

where V is slip velocity, θ is the state variable, L is the characteristic slip distance, $a > 0$ and b are rate-and-state parameters of the order of 0.01, and f_0 is the reference friction coefficient at the reference slip velocity V_0. Several forms of the state evolution equation have been proposed based on laboratory experiments (e.g., [2–5, 30]), including the aging form:

$$\frac{d\theta}{dt} = 1 - V\theta/L \ , \tag{5}$$

the slip form:

$$\frac{d\theta}{dt} = -V\theta/L\ln(V\theta/L) \ , \tag{6}$$

and the composite form

$$\frac{d\theta}{dt} = \exp(-V/V_c) - V\theta/L\ln(V\theta/L) \ . \tag{7}$$

The rate-and-state laws were originally formulated purely empirically, to match the experiments. Currently, there is no full physical explanation of all the features and dependencies uncovered in the past 30+ years, but progress has been made toward such an explanation. In particular, the positive direct velocity effect ($a > 0$), which is generally found to involve a positive proportionality to ln(V), was plausibly attributed to a thermally activated creep process at stressed asperity contacts (e.g., [9, 31–36]). (Note that parameter b has also been found positive in all cases in which it was quantified, but currently there is no theoretical explanation for the sign of b.) Based on this insight, the rate-and-state law (4) is typically regularized around $V = 0$ for use in fault modeling (e.g., [34])

$$\tau = a\bar{\sigma}\,\mathrm{arcsinh}\left[\frac{V}{2V_0}\exp\left(\frac{f_0 + b\,\ln(V_0\theta/L)}{a}\right)\right] \ . \tag{8}$$

The characteristic slip distance L, which governs the evolution of frictional response under a constant velocity (and hence governs the

evolution of the state variable), can be interpreted as the slip required to renew the population of contacts. In this case, the state variable is interpreted as the lifetime of the asperity population. Dieterich and Kilgore [37, 38] made direct optical observations of the contact population and its change with slip (and also with time) in various transparent materials, including quartz; they linked the characteristic slip distance L (which they called d_c) to changes in the asperity population.

The evolution equation for the state variable is also used to incorporate the observed healing of frictional interfaces during slide-hold-slide experiments [2–5, 10, 27, 39, 40], which is consistent with its interpretation as the lifetime of the asperity population. Intuitively, the shear strength of two surfaces in contact under compression should increase, and this is exactly what is observed. In the aging form of the state variable evolution (5), the state variable θ increases linearly with time when the interface is locked ($V = 0$), and hence the friction coefficient increases with the logarithm of time, incorporating the experimentally observed healing. The name aging reflects this feature of the increasing contact lifetime during the locked stage. In the slip form (6), the state variable can evolve only at non-zero slip rate V (hence the name "slip" law), and no healing occurs during stationary contact as a result; this is a potential deficiency of this form [27]. However, the slip form is a better match to the velocity-jump experiments [27, 41]. That is why other state evolution laws have been proposed, with additional parameters, e.g. the composite form (7) [30].

At a constant velocity V, the state variable and hence friction evolve toward constant values (called steady-state values) θ_{ss} and τ_{ss}. The steady-state values of friction at different sliding velocities are different in general, and often faster sliding velocity corresponds to lower steady-state friction (steady-state velocity-weakening behavior). Steady-state velocity-strengthening behavior is promoted by lower confining stresses, higher temperatures, and presence of some minerals [10, 24, 25, 28, 42–44]. From the formulation (4-6), one gets:

$$\theta_{ss} = L/V, \quad \tau_{ss} = \bar{\sigma}[f_0 + (a-b)\ln(V/V_0)] \ , \tag{9}$$

showing that the sign of $(a - b)$ encapsulates the velocity-weakening ($a - b < 0$) and velocity strengthening ($a - b > 0$) response. Shear zones with $a - b = 0$ are called velocity-neutral.

Despite the fact that the rate-and-state effects are usually small (of the order of 1-10%) compared to the baseline frictional strength, these effects are fundamentally important for the physically and mathematically meaningful

stability properties of frictional sliding. In studies of 2D linearized elastodynamic stability of frictional sliding for a general class of rate-and-state laws with the properties discussed above and steady-state velocity, it has been established [36] that (i) the perturbations with the largest wavenumbers (the shortest wavelengths) are always stable, which guarantees the well-posedness of the problem in response to a generic perturbation, and (ii) the dynamic response converges to the quasi-static one for sufficiently small sliding velocities, i.e., the problem has a quasistatic range, as one would intuitively expect. However, compromises from the full rate-and-state constitutive framework, in the direction of the classical friction laws like purely velocity-dependent friction of velocity-weakening type, do not allow a quasi-static range of sliding velocities and, in fact, lead to paradoxical predictions (supersonic propagation of all perturbations) or ill-posedness [36].

2.2. Formulations with additional features

The rate-and-state laws (4-7) encapsulate the most commonly observed experimental features and have been quite successfully used to interpret a range of earthquake source phenomena, as discussed in section 4. However, a number of important extensions have been proposed and used based on experimental and theoretical studies.

Evolution of shear stress in response to normal stress changes: In (4), frictional resistance is directly proportional to normal stress, as has been established in numerous experiments dating back to Leonardo da Vinci. However, experiments have shown that, for fast enough variations in normal stress, the frictional resistance does not immediately assume the proportional values but rather evolves with slip [45, 46], over slip scales comparable to those of the state variable evolution. This evolution effect has been incorporated into rate-and-state friction formulations by modifying the state evolution equation [45]. It is important to include this modification whenever the normal stress variations occur on slip scales smaller than or comparable to the characteristic slip distance L. For example, this gradual evolution is critically important in proper formulations of problems that involve slip between elastically dissimilar materials (e.g., [36, 47, 48]) where the normal stress change is coupled with slip and can occur quite rapidly at the rupture tip. However, the effect appears to be unimportant for a range of models where the effective normal stress varies due to pore pressure evolution (e.g., [49,50]), because, for the parameters assumed in those studies, normal stress changes occur over slip scales large compared to L.

Dilatancy: Inelastic shear dilatancy results when the pore space in the shearing gouge varies during the shear deformation due to particle motion or breakage; both positive and negative dilatancy (or compaction) occurs with shear. It has been observed in lab experiments and shown to correlate with the frictional resistance (e.g., [24, 29]); its effect may be at least partially responsible for the state-evolution effects discussed in section 2.1. The dilatancy also affects frictional resistance by modifying pore pressure p in the effective stress expression (1) and this effect has been shown to be important for earthquake nucleation and balance of seismic and aseismic slip [51–54]. Evolution of dilatancy with deformation and its effect on friction is an active area of research in geomechanics (e.g., [55, 56]). Based on the velocity-jump experiments in which variations in dilatancy were also measured [24], the work of Segall and Rice [51] linked inelastic shear dilatancy to the state variable of the formulation (4-5):

$$\frac{d\phi_{\rm pl}}{dt} = -\epsilon\frac{\dot{\theta}}{\theta}\,, \tag{10}$$

where $d\phi_{\rm pl}/dt$ is the inelastic dilatancy (or the rate of change of the pore space volume) and ϵ is a constant which is of the order of 10^{-4} based on matching experimental measurements [24].

Enhanced coseismic weakening: The standard rate-and-state formulations have been proposed based on friction experiments at relatively slow slip velocities (10^{-9} to 10^{-3} m/s) in comparison to seismic values of the order of 1 m/s. There is mounting experimental and theoretical evidence that, at larger slip velocities and slips, fault behavior is affected by additional processes, as reviewed by Tullis [57]. Several of the processes are due to shear heating (inevitable during fast sliding that accumulates significant slip) such as flash heating of contacting asperities [58–62], pore pressure evolution [49, 50, 59, 63–71], and melting [72, 73]. These additional processes are likely to dominate rate-and-state effects during dynamic slip, potentially causing significant additional weakening of fault surfaces. The effect of enhanced coseismic weakening on the long-term fault behavior is briefly discussed in section 5.

Let us give an example of how coseismic effects of shear heating can be added to the rate-and-state formulation. One shear-heating weakening mechanism that has laboratory support is flash heating, in which tips of contacting asperities heat up and weaken. Such weakening may be activated even for very small slips and could be important even for microseismicity. To include flash heating, one can modify the rate-and-state formulation (4) to:

$$\tau_f = \bar{\sigma} f = (\sigma - p)\left[\frac{f_0 - f_w + a\ln(V/V_0) + b\ln(V_0\theta/L)}{1 + L/\theta V_w} + f_w\right], \quad (11)$$

where V_w is the characteristic slip velocity at which flash heating starts to operate, f_w is the residual friction coefficient, and pore pressure p could be evolving due to dilatancy and shear heating as discussed below. Based on laboratory experiments and flash heating theories, V_w is of the order of 0.1m/s.

Evolution of pore pressure: Pore pressure p in the shearing layer would clearly affect the frictional resistance, and this is encapsulated in the friction law (4) that includes effective normal stress.

Several processes can affect the pore pressure, including shear heating of the pore fluid and dilatancy that changes the pore volume. If pore pressure in the shearing layer is different from that in the surrounding medium, diffusion processes should start taking place, governed in part by permeability of the surrounding materials. The permeability can vary by orders of magnitude both in space (e.g., [74]) and presumably in time (e.g., due to coseismic damage or interseismic healing).

A formulation to compute the coupled temperature and pore pressure evolution is given by [49, 50, 59]

$$\frac{\partial T}{\partial t} = \alpha_{\rm th}\frac{\partial^2 T}{\partial y^2} + \frac{\omega(y)}{\rho c},$$
$$\frac{\partial p}{\partial t} = \alpha_{\rm hy}\frac{\partial^2 p}{\partial y^2} + \Lambda\frac{\partial T}{\partial t} - \frac{F(y)}{\beta_c}\frac{d\phi_{\rm pl}}{dt}, \quad (12)$$

where y is the space coordinate normal to the fault, T is the temperature, $\alpha_{\rm th}$ and $\alpha_{\rm hy}$ are the thermal and hydraulic diffusivities, $\omega(y)$ is the heat generation rate, the integral of which over y equals to τV, ρc is the specific heat, Λ is pore pressure change per unit temperature change under undrained condition, β_c is the specific storage, and $F(y)$ is a function representing the distribution of the inelastic porosity change. Both the heat source, $\omega(y)$, and inelastic pore space generation factor, $F(y)$, are distributed within the width of the shearing layer. One typical assumption is to take these terms to represent the effect of uniform sliding in the fault zone of thickness w, which would result in $\omega(y) = \tau_f V/ wH(y - w/2)H(-y - w/2)$ and $F(y) = H(y - w/2)\mathrm{H}(-y - w/2)$, where $H(y)$ is the Heaviside step function. Note that several studies on the effect of

dilatancy (e.g., [51, 52]) used a simplified model for the pore pressure diffusion:

$$\frac{\partial p}{\partial t} = \frac{1}{\bar{t}}(p^{\infty} - p) + \Lambda \frac{\partial T}{\partial t} - \frac{1}{\beta_c}\frac{d\phi_{\mathrm{pl}}}{dt} \ , \tag{13}$$

where $\bar{t}$ is the time scale for the diffusion and p^{∞} is the background pore pressure.

Other effects: The rate-and-state formulations discussed are capable of capturing a number of effects observed in the experiments, but they have limitations. The laws for the state variable evolution remain empirical and fit various experimental responses to a different degree (e.g., [10, 27, 29, 41, 53]). In fact, formulations with more than one state variable were considered at some point [4,75,76]. Variations in temperature appear to have effects on low-velocity frictional resistance similar to those in slip rate [31] but these effects are not included in the standard descriptions. The characteristic slip distance L may not be a constant material property but vary with a number of shear characteristics including the width of the shearing layer [26]. Dilatancy in the presence of fluids may affect rate-and-state parameters [77]. More discussion on some of these and other effects is given in the reviews on rate-and-state friction by Marone [10] and Dieterich [78].

3. Stability of slipping on rate-and-state interfaces

In applying rate-and-state laws to earthquake source processes, one is most interested in how faults governed by such laws would respond to slow loading provided by the motion of tectonic plates. Significant insight into this problem has been obtained from targeted stability studies that consider specialized scenarios. Most of such studies have employed one or more of the following strategies: (i) analyzing the response of spring-slider models and interpreting the results in terms of continuum models, (ii) considering linearized stability of steady sliding of an interface between two infinite elastic half-spaces, and (iii) constructing estimates using theoretical developments with assumptions based on numerical simulations.

All such studies with the standard Dieterich-Ruina formulation (4-6) have resulted in the same basic conclusions. Velocity-strengthening fault zones ($a - b > 0$) respond to slow loading with similarly slow slip and cannot produce spontaneously accelerating slip. Of course, slip there can be perturbed, e.g. by applying external stress changes, but the stability properties of velocity-strengthening regions are such that the perturbations would tend to die out. In the velocity-weakening regions ($a - b < 0$), small enough

regions slip slowly under slow loading, but once the slipping region is large enough, it spontaneously accelerates toward inertially driven rupture.

The estimates of the critical length scale for seismic, wave-producing slip in velocity-weakening regions, often called the nucleation size, have been obtained in several ways. Linearized stability studies of steady motion with a constant slip rate [7,36] show that the motion is stable to perturbations of small enough wavelengths and unstable to perturbations of large enough wavelengths, with the critical wavelength in quasi-static two-dimensional (2D) problems given by:

$$h^*_{\mathrm{RR}} = \frac{\pi}{4}\frac{\mu^* L}{(\sigma - p)(b - a)}, \tag{14}$$

where $\mu^* = \mu$ for antiplane (mode III) problems, $\mu^* = \mu/(1 - \nu)$ for inplane (mode II) problems, μ is the shear modulus, and ν is the Poisson's ratio. This result is valid for both aging and slip formulations (5-6) since they have the same linearized expression about steady-state sliding. Note that the critical wavelength for the velocity-neutral interfaces ($a - b = 0$) is infinite based on (14), which is consistent with velocity-strengthening regions being always stable. The critical wavelength (14) can be regarded as an estimate of the nucleation size. Another estimate for the aging laws has been obtained [79] by considering acceleration to instability of a one-degree-of-freedom spring-slider system which approximates a fixed patch:

$$h^*_D = C\frac{\mu^* L}{(\sigma - p)b}, \tag{15}$$

where C is a model-dependent constant that enters the relation between the patch size h and its effective stiffness k_{eff} through $k_{\mathrm{eff}} = C\mu^*/h$. Numerical simulations of slip in the velocity-weakening regions governed by the aging formulation with $a/b > 0.5$ have shown that the stable slip there tends to take the form of a quasi-statically extending crack [80]; the energy balance for that scenario leads to the following estimate:

$$h^*_{\mathrm{RA}} = \frac{2}{\pi}\frac{\mu^* L b}{(\sigma - p)(b - a)^2}\ . \tag{16}$$

The regime of $a/b > 0.5$ includes typical experimentally measured values for a and b that have been widely used in simulations (e.g., [34, 81–83]). In the regime $a/b < 0.4$, the same study [80] found that a fixed-patch estimate of the

type (15) works well. Formulations with the slip law tend to produce smaller nucleation sizes [53, 84].

The estimates can be written in the form

$$h^*_{\rm all} = C\frac{\mu^* L}{(\sigma - p)F(a,b)} \ , \tag{17}$$

where C is a model-dependent constant of order 1 and $F(a, b)$ is a combination of parameters a and b. 3D estimates would be larger than 2D estimates by a factor of two to three; $h^*_{\rm RA}$ needs to be increased by a factor of $\pi^2/4$ (A. Rubin, private communication). The resulting estimate

$$h^* = \frac{\pi}{2}\frac{\mu^* L b}{(\sigma - p)(b-a)^2} \ , \tag{18}$$

matches well nucleation sizes in 3D long-term simulations of earthquake sequences [82, 83] and this is the estimate mentioned in sections 4.1-4.3.

The typical nucleation sizes h^* expected from such estimates can be obtained using representative values of μ = 30000 MPa, L = 1 to 100 microns, $\sigma - p$ = 200 MPa (which is representative of over-burden minus hydrostatic pore pressure at 10 km depth), b = 0.015, and a = 0.01, resulting in h^* of the order of 0.1 to 10 meters. This consideration predicts that all earthquakes due to instability of frictional interfaces should have sizes comparable to or larger than 0.1 to 10 meters. This is consistent with the observational study of microseismicity in mines [85] which estimated the size of the smallest "friction-dominated" earthquakes to be about 20 m; smaller events had a different power spectrum and were classified as "fracture-dominated".

To estimate the moment magnitudes of the smallest allowable events according to the nucleation size estimates, let us assume a simple model of a circular patch with radius r and a constant stress drop $\Delta\tau$, for which the seismic moment is given by (e.g., [86]):

$$M_0 = \frac{16}{7}\Delta\sigma r^3 \tag{19}$$

and the moment magnitude is given by ([87]):

$$M_{\rm w} = \frac{2}{3}\log_{10} M_0 - 6.07 \ , \tag{20}$$

where M_0 is in Nm. Taking the nucleation estimates of 0.1 to 10 m as the patch size and considering typical stress drops of 1 to 10 MPa, we get a range

of the minimum moment estimates from $M_w = -3.8$ to $M_w = 0.8$, with the former value corresponding to the radius of 0.1 m and stress drop of 1 MPa, and the latter value corresponding to the radius of 10 m and stress drop of 10MPa. Note that the lower estimate of $M_w = -3.8$ can be further decreased somewhat by assuming a smaller stress drop or a smaller nucleation size (e.g., because of locally elevated normal stress). Such low moment magnitude estimates mean that, due to observational limitations, it is not easy to confirm or deny the existence of such cut-offs on real faults. However, the estimates are consistent with observations, in the sense that they allow seismic events down to M_w of -4.

In the nucleation size estimates (17), two parameters can vary in a broad range: the effective normal stress $(\sigma - p)$ and the characteristic slip distance L. Since pore pressure p can reach near-lithostatic values [88,89], the nucleation size estimates can become arbitrarily large for p arbitrarily close to σ. This implies that some segments may have velocity-weakening properties but slip stably, due to sufficiently low effective normal stresses that would make their nucleation sizes larger than the segments themselves. Moreover, the estimates predict that as p changes, e.g. due to fluid flow, so does the nucleation size that determines the stability of a fault segment. This has been explored to model slow slip events [88, 90]. Similarly, values of L may vary more broadly on real faults than in laboratory experiments and may be process-dependent [26].

The already rich fault behavior predicted by the stability analysis of the standard rate-and-state formulations (4-6) becomes even more complex when additional factors (section 2.2) are included (e.g., [51, 75, 91]).

Let us focus here on inelastic dilatancy (10) and the associated pore pressure effects [51]. Clearly, the permeability of the materials that surround the fault zone is quite important for processes that involve pore pressure: if the materials are sufficiently permeable, then any change in pore pressure in the shear zone would be immediately counteracted by the fluid flow and the pore pressure would not change. In this regime, called "drained", pore pressure does not change due to dilatancy. On the other hand, if the materials are effectively impermeable on the relevant time scales, then the fluid flow in and out of the shearing layer is negligible and the effect of dilatancy on pore pressure p and hence fault stability is maximized. Such a regime is called "undrained".

The ratio $u = t_{diff}V/L$, where t_{diff} is the characteristic diffusion time that depends on the permeability and L/V is the characteristic time for the state variable evolution, measures how drained or undrained the deformation is; $u \ll 1$ and $u \gg 1$ correspond to effectively drained and undrained conditions, respectively. For drained conditions, the nucleation size estimates (17) hold, with p equal to the background pore pressure p^∞. For undrained conditions, the behavior depends on the dilatancy parameter ϵ from (10):

$$h^*_{\rm dil} = C\frac{\mu^* L}{(\sigma - p^\infty)(b-a) - \frac{f_o \epsilon}{\beta L}} \text{ if } \epsilon \ll \epsilon_{\rm crit} \ , \tag{21}$$

and

$$h^*_{\rm dil} = \infty \text{ if } \epsilon \gg \epsilon_{\rm crit} \ , \tag{22}$$

where

$$\epsilon_{\rm crit} = (\sigma - p^\infty)(b-a)\beta L / f_o \ . \tag{23}$$

Hence, in the undrained limit, the nucleation size is increased by dilatancy for small enough values of ϵ and becomes infinite for large enough values of ϵ. Note that, as the fault slip accelerates and slip rate V increases, parameter u increases as well and the character of the deformation changes in the direction of being more undrained, assuming that $t_{\rm diff}$ and L stay constant. This suggests that the nucleation size that the slipping zone needs to achieve to accelerate to seismic slip rates increases as the fault accelerates, from the undrained estimates (17) to the values in (21). These characteristics have also been exploited for modeling slow slip events [53, 54].

To summarize, velocity-weakening segments can produce both rapid and slow slip under tectonic loading even in the standard rate-and-state formulations. If their properties and effective normal stress correspond to the nucleation sizes larger than the segment in question, they can be completely aseismic. Such effects are even more pronounced if dilatancy is taken into account. Note that the consideration above ignores shear heating which would tend to increase pore pressure and counteract the effects of dilatancy, a possibility even for relatively slow slips during nucleation processes [92]. Shear-heating weakening mechanisms may also act in the regions that are velocity-strengthening at low rates, if activated during rapid slips characteristic of earthquakes that penetrate into the velocity-strengthening regions due to stress concentrations on their fronts [93] (manuscript in preparation).

4. Success of rate-and-state fault models in reproducing observed phenomena

The fault models with rate and state laws have been successfully used to reproduce and analyze a wide range of earthquake phenomena, including earthquake nucleation, stick-slip and creeping fault regions, spatio-temporal

slip complexities, earthquake aftershocks, and slow slip events (e.g., [10,11,33,34,49–54,78–84,88,90,92–130]). In this section, we give several examples of such modeling.

To study slip on a fault governed by rate-and-state laws, including numerical simulations of the slip evolution with time, the fault needs to be embedded into a representation of the surrounding medium, to couple the fault resistance with fault tractions that evolve due to loading as well as due to slip of the fault itself. Often, planar faults embedded into elastic bulk are considered. Even in such models, simulations of long-term slip histories punctuated by earthquakes are quite challenging because of the variety of temporal and spatial scales involved [34, 83]. Slow loading requires hundreds to thousands of years in simulated time and fault zone dimensions are in tens to hundreds of kilometers. At the same time, rapid changes in stress and slip rate at the propagating dynamic rupture tip occur over distances of order meters and times of order a small fraction of a second. It is especially challenging to account for all inertial effects during seismic events while simulating earthquake sequences.

One methodology for such simulations uses spectral boundary integral methods. It was developed first for 2D models [34], then 3D models [83], and then extended to problems with coupled temperature and pore pressure evolution [50] based on prior studies [94, 102, 104, 131]. The approach allows us to resolve slow aseismic slip, fast seismic slip, and the gradual transition between them. During earthquake rupture, all inertial effects are accounted for. It is accurate and efficient for simulating slip on planar faults embedded in homogeneous elastic media but it becomes less efficient for complex fault geometries and cannot be straightforwardly applied to problems with heterogeneous or inelastic bulk.

Other approaches has been proposed (e.g., [97,120,127,132–134]), which adopt simplified treatments of either slow tectonic loading and hence aseismic slip, or inertial effects during dynamic rupture, or transition between interseismic periods and dynamic rupture; the simplifications either allow for computational efficiency, or make the approaches more flexible with respect to fault geometries and bulk representations, or both. In particular, the quasi-dynamic approach [97] significantly simplifies the treatment of inertial effects during simulated earthquakes by ignoring wave-mediated stress transfers; it has been widely used in earthquake studies (e.g., [88,97,99,102, 108,109,111,129]).

Note that useful insights about fault slip have been obtained using spring-slider models, e.g. some of the stability studies discussed in section 3. However, such models have significant limitations and should be used with

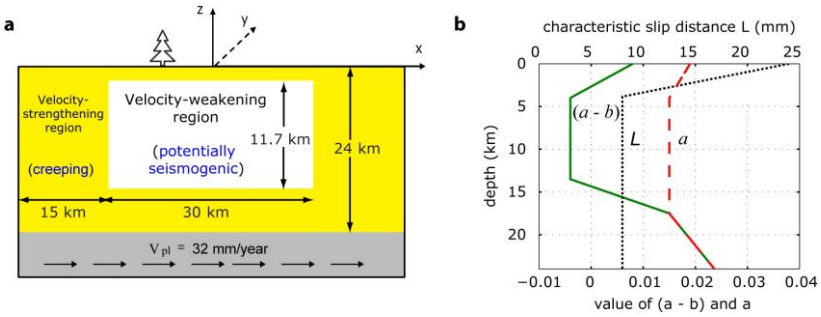


Figure 2. Properties of the fault segment used to illustrate the response of rate-and-state faults (Figure 2 from [83]). (a) Rate-and-state friction acts on the top 24 km of the fault. A potentially seismogenic region of velocity-weakening properties (white) is surrounded by velocity-strengthening regions (yellow). Below the depth of 24 km, steady motion of 32 mm/year is imposed. (b) Depth dependence of friction parameters $(a - b)$, a, and L in the velocity-weakening region. The effective normal stress $\bar{\sigma}$ is constant and equal to 50 MPa. Note that a more realistic distribution of $\bar{\sigma}$ would have smaller values closer to the free surface; near-free-surface normal stress variations do not affect the overall behavior in this model [83].

care. Models with one slider have only one degree of freedom and hence cannot represent many important effects such as the evolving size (and hence effective stiffness) of a slipping zone; such effects can lead to qualitative changes in the model behavior (e.g., [113]). Models with multiple sliders connected by springs typically restrict elastic interactions to nearest neighbors, and hence they cannot reproduce the long-range interactions inherent in the elastic medium. The absence of long-range interaction can qualitatively change the system behavior, e.g., by limiting stress concentration at the rupture tip and promoting rupture arrest (e.g., [97]).

4.1. Rate-and-state laws as a unified description for earthquake-producing and creeping fault segments

Let us illustrate the behavior of rate-and-state faults using the geometrically simple model of a rectangular steady-state velocity-weakening region surrounded by a velocity-strengthening area on a planar fault embedded into an elastic half-space (Figure 2), from the study by Lapusta and Liu [83]. The fault is loaded by the plate-like relative velocity of 32 mm/year (or 10^{-9} m/s) on the deeper extension of the fault. The fault response is governed by the standard rate-and-state formulation (4) with the aging form of the state variable evolution (6), regularized at $V = 0$ as in (8). The

details of the model parameters and simulation methodology are given in [83].

As expected from stability properties of rate-and-state interfaces, the velocity-strengthening region steadily slips with velocities comparable to the plate rate, while the velocity-weakening region accumulates most of its slip through earthquakes. This is evident from histories of slip velocity and slip for two representative points (Figure 3). Point P1 from the velocity-weakening region has slip velocity much below the loading plate rate for most of the simulated time, with occasional spikes to values of the order of 1 m/s, typical for observed seismic slip (Figure 3a); this indicates long periods of locking with occasional earthquakes. Slip accumulation of P1 has the corresponding step-like nature (Figure 3c). In contrast, point P2 from the velocity-strengthening region has slip velocity of the order of the plate rate most of the time, with relatively small increases after each earthquake

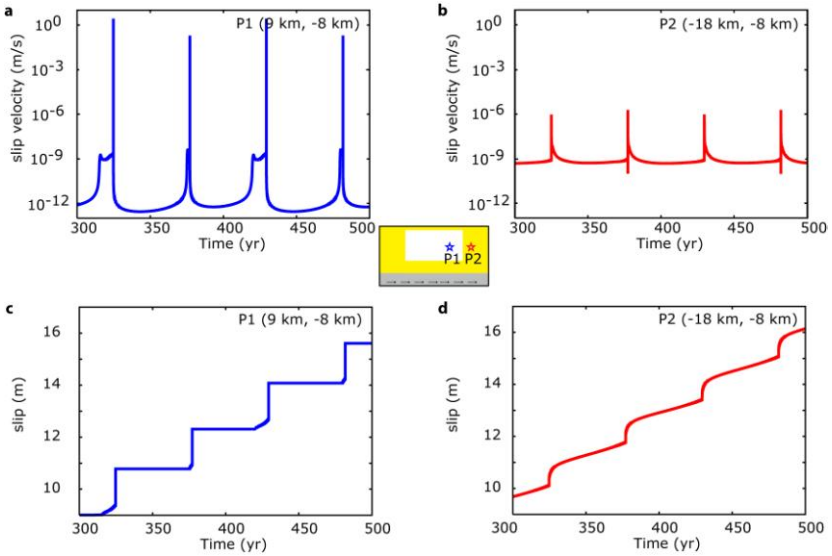


Figure 3. (a)-(b) Slip velocity and (c)-(d) slip of two fault points, one from the velocity-weakening region (P1) and the other from the velocity-strengthening region (P2) (Figure 3 from [83]). Slip velocity is plotted on the logarithmic scale. Point P1 is virtually locked for most of the time, with slip velocity three orders of magnitude below the plate rate of 10^{-9} m/s, but occasionally slips fast, with the maximum slip velocity of the order of 1 m/s. Point P2 moves with near-plate velocity for most of the simulated time; after each dynamic event, it has postseismic slip with the maximum slip velocity of the order of 10^{-6} m/s.

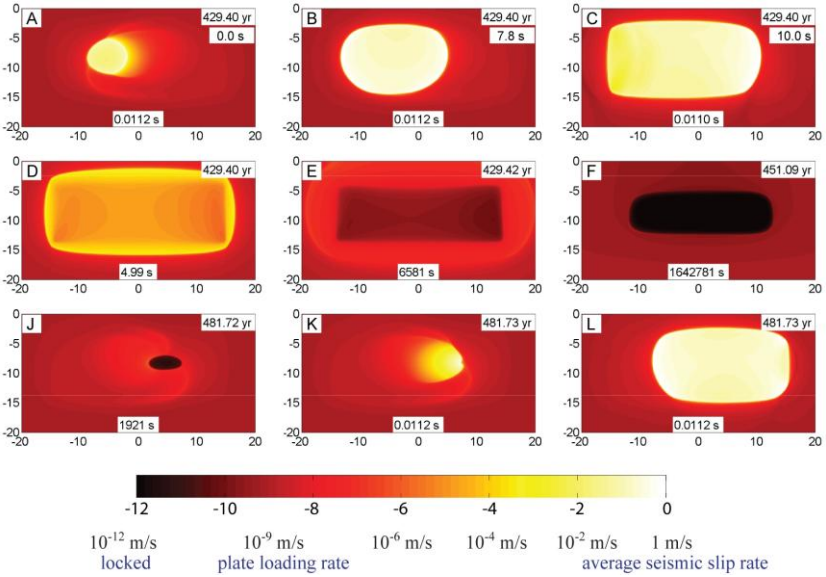


Figure 4. A typical earthquake cycle illustrated through snapshots of spatial slip-velocity distribution (Figure 4 from [83]). Slip history between the 9th and 10th events is illustrated for L = 8 mm and h^*/W_{seis} = 0.8. Colors represent slip velocity on the logarithmic scale. White and bright yellow correspond to seismic slip rates, orange and red correspond to aseismic slip, and black corresponds to locked portions of the fault. Each panel shows the time t of the snapshot in years (in the upper-right corner) and the corresponding time step Δt in seconds (at the bottom of each panel). Panels A-C also show the time in seconds elapsed since the time of panel A. The simulations reproduce dynamic events (panels A-C and K-L), postseismic slip (panels D-E), and the interseismic period (panel F).

corresponding to postseismic slip (Figure 3b). Correspondingly, slip at P2 increases steadily in time, with faster accumulation after each dynamic event (Figure 3d).

Typical earthquake cycles are illustrated in Figures 4 and 5 through snapshots of slip-velocity distribution on the fault. For the case of L = 8 mm, the estimate of the nucleation size (18) is h^* = 9 km, and hence h^*/W = 0.8 where W is the width of the velocity-weakening region. Such a large estimate predicts that most of the velocity-weakening zone should be creeping before an earthquake. This is exactly what the simulation shows (Figure 4). When an earthquake nucleates (the yellow patch in panel A), the rest of the fault, including the velocity-weakening region, moves with the rates comparable to

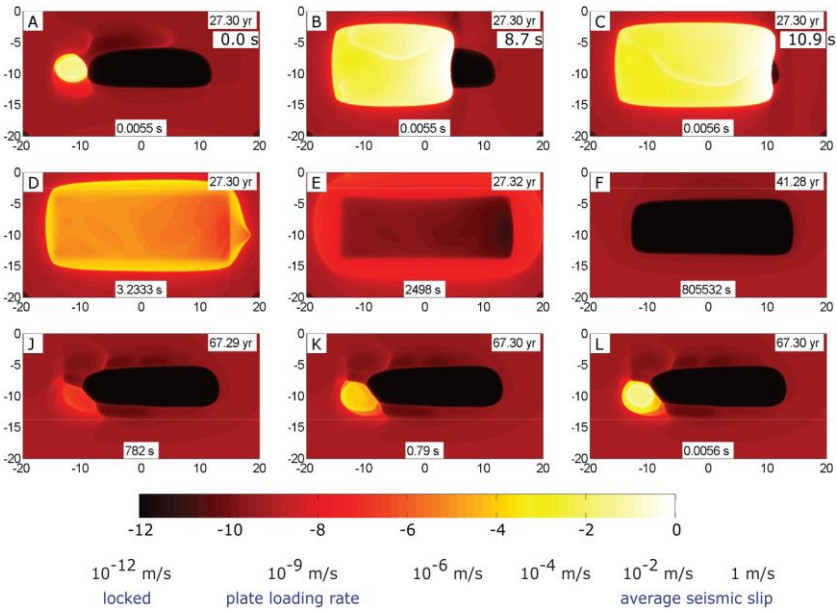


Figure 5. A typical earthquake cycle for L = 4 mm (h^*/W_{seis} = 0.4) (Figure 5 from [83]). Slip history between the 2nd and 3rd events is illustrated. Colors and time markings have the same meaning as in Figure 4. Compared with the case with L = 8 mm (Figure 4), dynamic events in the case with L = 4 mm have smaller nucleation size, nucleate closer to the rheological transition (panels A, L), have more unilateral propagation, and develop faster rupture speeds (panels A-C). Consistently with the smaller value of h^*/W_{seis}, the velocity-weakening region experiences less aseismic slip, with a large portion of the region still locked when a seismic event nucleates (panels A, J-L).

the imposed plate rate. The earthquake propagates bilaterally first and then mostly to the right (panels A-C). The seismic slip causes positive static stress changes in the surrounding velocity-strengthening area, which responds with increased aseismic slip rates that decay over time (panels D-E). This is postseismic slip. During the interseismic period (panel F), the velocity-weakening region is locked, while the surrounding velocity-strengthening region moves with slip velocity of the order of the plate rate. That aseismic slip creates stress concentration at the boundary between the locked and slipping regions, causing slip there and hence continuously moving the boundary into the locked region. For L = 8 mm, the locked region almost disappears (panel J), consistently with the large estimate of the nucleation zone h^* = 9 km which approximates how far slow slip can penetrate into the

velocity-weakening region without nucleating a dynamic event. The next dynamic event nucleates on the right side of the seismogenic region (panel K of Figure 4) and propagates bilaterally first (panel L) and then mainly to the left.

For the smaller value of L = 4 mm, and hence for the smaller h^* = 4.5 km and h^*/W = 0.4, the seismogenic region experiences less aseismic slip in the interseismic period, as expected (Figure 5). Events nucleate closer to the rheological transition (panel A) and propagate more unilaterally (panels B-C). Right after postseismic slip (panels D-E), most of the seismogenic region is locked and the fault behavior for both values of L is quite similar (panels F, Figures 4 and 5). When the next dynamic event nucleates (panels J-L, Figure 5), much of the velocity-weakening region remains locked. The nucleation size of a seismic event, defined as area of accelerating slip right before dynamic rupture propagation, is about 5 km (panels A and K), consistent with the estimate h^* = 4.5 km.

The model produces dynamic events of M_w = 6.6 with a number of realistic features, including stress drops of the order of 3 MPa, maximum slip velocity over the fault exceeding 1 m/s, and rupture speeds reaching 2.5 km/s. Each point which is ruptured dynamically exhibits effective stress-slip dependence that closely resembles linear slip-weakening laws [83, 107]. This is illustrated in Figure 6, which shows the behavior of three velocity-weakening points and

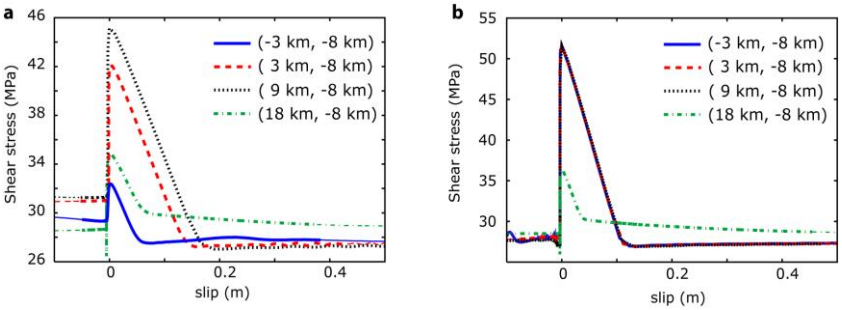


Figure 6. The dependence of shear stress on slip for four locations on the fault with (a) L = 8 mm and (b) L = 4 mm (Figure 8 from [83]). In both cases, dynamic rupture propagates from the left side of the fault to the right side, passing the velocity-weakening locations (-3 km, -8 km), (3 km, -8 km), (9 km, -8 km), and then influencing the velocity-strengthening location (18 km, -8 km) as the rupture arrests in the velocity-strengthening region. Zero slip for each point is chosen as the slip when shear stress at the point reaches its peak during the dynamic event. We see that the effective dependence of stress on slip is similar to linear slip-weakening friction, with the slip-weakening rate $W \approx \sigma b/L$. The velocity-strengthening point has a smaller values of b than the other three points and hence a smaller slope.

one velocity-strengthening point. The velocity-strengthening point is located close to rheological transition. For all curves, the weakening slope is well-approximated by $-\sigma b/L$, as expected theoretically [83,107]. For L = 8 mm, we find that the effective slip-weakening behavior is similar for different points but not identical, with the peak stress and effective slip-weakening distance increasing with the rupture propagation. This is because the rupture accelerates as it propagates along the fault, and the associated increase in peak slip velocity causes increases in the peak stress and effective slip-weakening distance. For L = 4 mm, the dependence of stress on slip is nearly identical for the velocity-weakening points, because the rupture accelerates early in the event and, afterwards, the relatively homogeneous fault properties and conditions ensure that the rupture behavior does not change much as the rupture propagates along the fault.

This example clearly shows that rate-and-state fault models can reproduce a wide range of slip behaviors observed on real faults, including slow slip near the imposed plate rate, accelerated postseismic slip after dynamic events, spontaneous initiation of earthquake-producing ruptures, their dynamic propagation with slip rates and rupture velocities comparable to the ones inferred from observations, and the resulting earthquakes with reasonable slips and slip rates. In [83], this example was used to illustrate the developed methodology, to study the effect of initial prestress, and to compare fully dynamic and quasi-dynamic simulations. Sections 4.2-4.4 give examples of rate-and-state modeling that reproduces various aspects of specific fault behaviors in a remarkably quantitative way.

Note that the adopted values of parameter L of the order of millimeters in this section are larger than the laboratory values of the order of 1-100 microns. This is done for numerical tractability [83], at the expense of increased nucleation zones. Section 4.2 continues this approach to enable simulations of an actual similarly sized fault segment. Sections 4.3-4.4 adopt laboratory values of L and succeed in explaining some intriguing observations about small earthquakes.

4.2. Modeling earthquake cycle at Parkfield

The rate-and-state friction framework and the simulation capabilities discussed in section 4.1 have been used to reproduce a range of observations for the Parkfield segment of the San Andreas Fault in California by Barbot, Lapusta, and Avouac [130]. The Parkfield segment is located between the locked Cholame segment to the south - the site of the 1857 M_w 7.9 Fort Tejon earthquake - and a creeping segment to the north which generates only small magnitude earthquakes. It accommodates about 35mm/yr of the relative displacement between the North American and Pacific plates (Figure 7).

The variety and long period of available observations make this fault segment an ideal location for testing the potential of rate-and-state models to explain consistently fault slip evolution across multiple earthquake cycles. The Parkfield segment has experienced at least six moment-magnitude (M_w) 6 quakes since 1857, and it was the site of the only officially recognized earthquake prediction experiment in the US [135]. In the late 1980s, based on

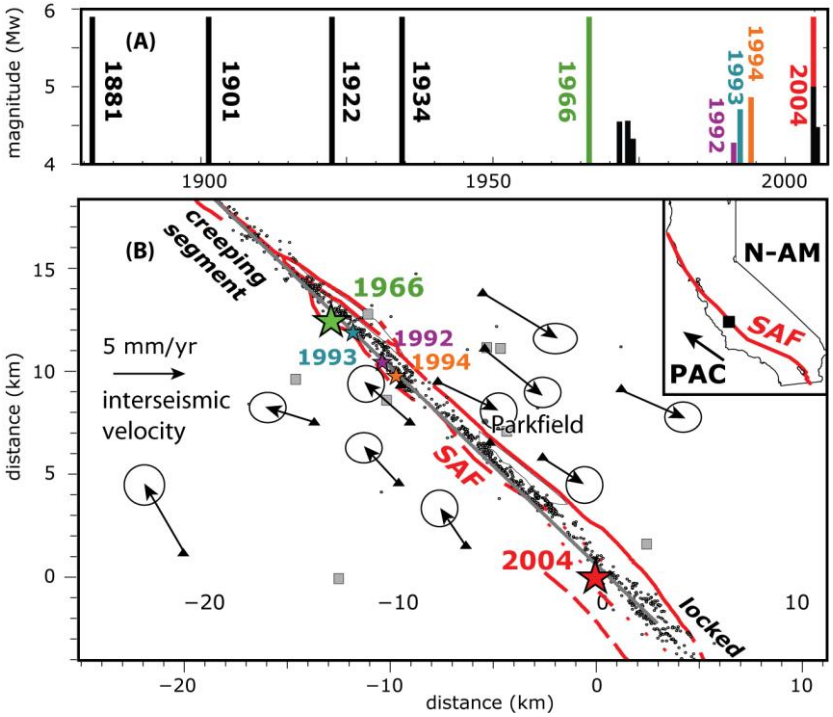


Figure 7. Tectonic setting and observations of earthquake cycle at Parkfield. A) Paleoseismic cycles of the M_w 6 earthquakes since 1881 (with inter-event times between 12 and 38 yr). A seismic crisis occurred in 1992-1993, around the anticipated due date of the latest M_w 6 event, culminating with the M_w 4.6 1992/10 and the M_w 4.4 1993/11 earthquakes. B) Parkfield segment belongs to the San Andreas Fault (SAF) which accommodates most of the relative motion between the Pacific (PAC) and North-American (N-AM) plates in Central California and produces very localized microseismicity [139] (black dots). Deformation during the co-, post-, and interseismic periods is monitored by various arrays of instruments, including GPS (black triangles and pre-2004 velocity vectors) and broad-band seismometers (gray squares). Modified from Figure 1 of [130].

the statistics of the previous recurrence times, the next earthquake was expected to strike in the early 1990s. Instead, a series of M_w 4 to 5 quakes populated the hypocentral area of the previous events at the predicted time. The M_w 6 earthquake took place in 2004, about a decade later than anticipated, and its epicenter was located 20 km away from the preceding event, at the other end of the segment. The prediction experiments prompted the deployment of relatively dense geodetic and seismological arrays in this region, enabling the observational analyses of inter-, co- and postseismic deformation [115,136–142].

The distribution of the rate-and-state properties in the Parkfield model is constructed based on the following observations and considerations [130]:

(i) The microseismicity in the segment exhibits a stable spatio-temporal distribution, with streaks of earthquakes surrounding a domain devoid of microseismicity (Figure 8) [139, 143–145]. Slip in the two recent M_w 6 events is inferred to occur in the domain between the two quasi-horizontal streaks, suggesting that this area has velocity-weakening properties. The surrounding areas tend to creep in the interseismic period (Figure 8), pointing to velocity-strengthening friction. These patterns of creep and seismicity suggest an inherent, long-lasting, structure of friction properties along the segment where a velocity-weakening domain is surrounded by a domain of velocity-strengthening friction (Figure 8). The transition between the two domains generates stress concentration and it is a natural place for microseismicity to occur [81,113,146]. The fact that some of the microseismicity takes the form of repeating earthquakes indicates that the transition may occur over a finite region with an interlaced network of asperities generating smaller earthquakes; such smaller-scale features are not included in the model.

(ii) The model is set up with the following uniform values for the friction parameters found in laboratory experiments to be typical for most rocks: $a = 10^{-2}$, $f_0 = 0.6$, and $V_0 = 10^{-6}$ m/s. The effective normal stress increases from $\bar{\sigma} = 1$ MPa near the surface to $\bar{\sigma} = 130$ MPa at 8 km depth.

(iii) The value of $(a - b)$ in the seismogenic zone is chosen to be compatible with the moment and recurrence times of the M_w 6 quakes, and the size of the seismogenic zone. In particular, the recurrence time T_r can be approximated as

$$T_r = \frac{(b-a)\bar{\sigma}}{\mu}\frac{R}{V_r}\ln\frac{V_{\rm co}}{V_{\rm inter}}\ , \tag{24}$$

where R is the width of the seismogenic zone, μ is the shear modulus, V_{co} is the representative coseismic slip velocity, $V_{\rm inter}$ is the representative

interseismic velocity, and V_r is the plate rate. Guided by expression (24), b = 0.014 is chosen [130], a value that falls within the range of experimentally observed.

Numerical simulations indeed verify that such parameters produce a series of ruptures that can reproduce the moment and average recurrence times of the Parkfield earthquakes, but additional considerations are required to explain other observations. In part, the inclusion of a narrow vertical band of low coupling in the middle of the seismogenic zone may partially explain the observation of two areas of high coseismic slip during the latest earthquake [115, 141, 147].

(iv) To reproduce the coseismic displacement of the near-field GPS stations, some asperities with velocity-weakening properties ($a-b < 0$) are placed at shallow depth. As evidenced by seismic inversions of the 2004 rupture from seismological data [141], the shallow coseismic slip may have been "aseismic", i.e., it did not radiate seismic waves. This is reproduced in the model by ensuring that the dimension of a shallow asperity is smaller than the critical nucleation size (18) at that depth (Figure 9), which prevents slip velocity from reaching seismic speeds (defined as $V > 0.1$ m/s).

(v) The amplitude of afterslip following the main shock is controlled by the area of stable friction and the amplitude of the stress perturbation, which depends on the details of the dynamic rupture.

The duration of the postseismic transient is inversely proportional to $(a - b)\ \bar{\sigma}$, as predicted by simple analytical estimates [115, 117, 130] and observed in numerical simulations [148]. Based on the analytical estimates, a range of $(a - b)$ values is chosen in the stable friction area, with the mean value of $(a - b) = 0.005$. This value is consistent with other estimates for Parkfield [115] and explains the GPS time series of postseismic transient.

(vi) The value of the characteristic slip L is assigned to be the largest one (for numerical tractability) that results in a suitably small nucleation size h^* in equation (18). Values of h^* that are large fractions of the seismogenic depth result in slow slip penetrating far into the seismogenic zone, as in Figure 4 of section 4.1, resulting in accelerated slips closer to the middle of the seismogenic zone. For values of h^* smaller than 2 km, the nucleation of large ruptures spontaneously occurs in the lower corners of the seismogenic zone, near the locations of either the 1966 or the 2004 hypocenters. Such h^* leads to L = 3 mm or lower. Note that the values of L should be much lower than 3 mm to reproduce the microseismicity occurring at the transition. Based on 2D models [81], we hypothesize that adopting smaller values would lead to the complexity of microseismicity in the transition regions coupled with M_w 6 events.

The model can explain a variety of observations at Parkfield, resulting in a long and rich history of fault slip with spontaneous nucleation and earthquakes of magnitudes ranging from M_w 2 to 6. It reproduces co-, post- and interseismic behavior of the Parkfield segment, with most coseismic slip occurring in the area circumscribed by microseismicity. In particular, it provides a good match to the pre-, co-, and post-seismic GPS recordings for the 2004 event [130]. The sequence of earthquakes includes the nucleation of a rupture near the hypocenter of the 2004 M_w 6 event (Figure 10a), which ruptures the entire seismogenic zone. The rupture propagates northwards, and stops at the creeping segment (Figure 10b), similarly to what occurred during the 2004 Parkfield event. During the rupture, the shallow asperities slip but their velocity remains

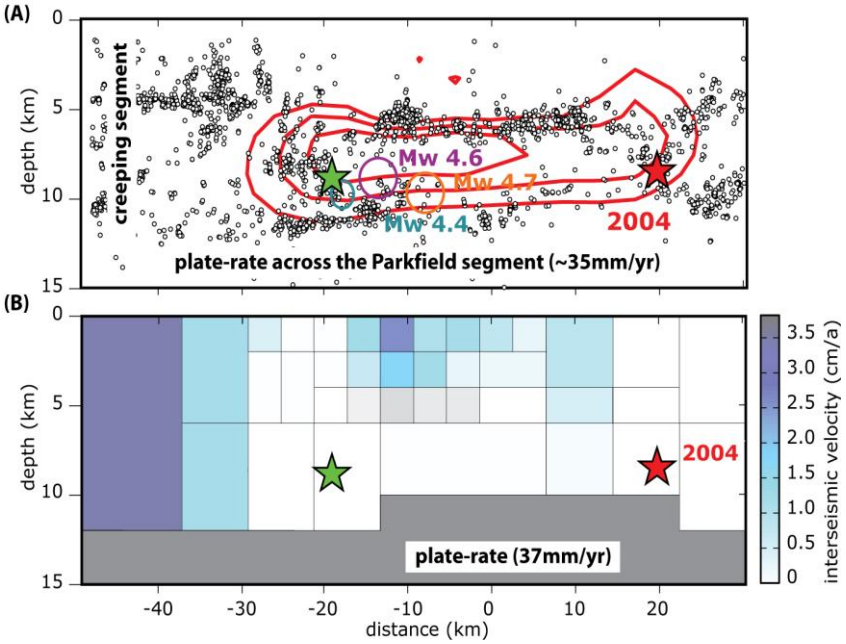


Figure 8. (A) Spatial distribution of fault slip and microseismicity on the SAF during the earthquake cycle inferred from observations. B) Aseismic creep occurs above and below the seismogenic zone before the 2004 Parkfield earthquake based on inversions of GPS data. The area between the microseismic streaks is fully coupled. The coseismic slip distribution of the 2004 main shock (20 cm red contours) may be concentrating in the domain circumscribed by background seismicity. Modified from Figure 1 of [130].

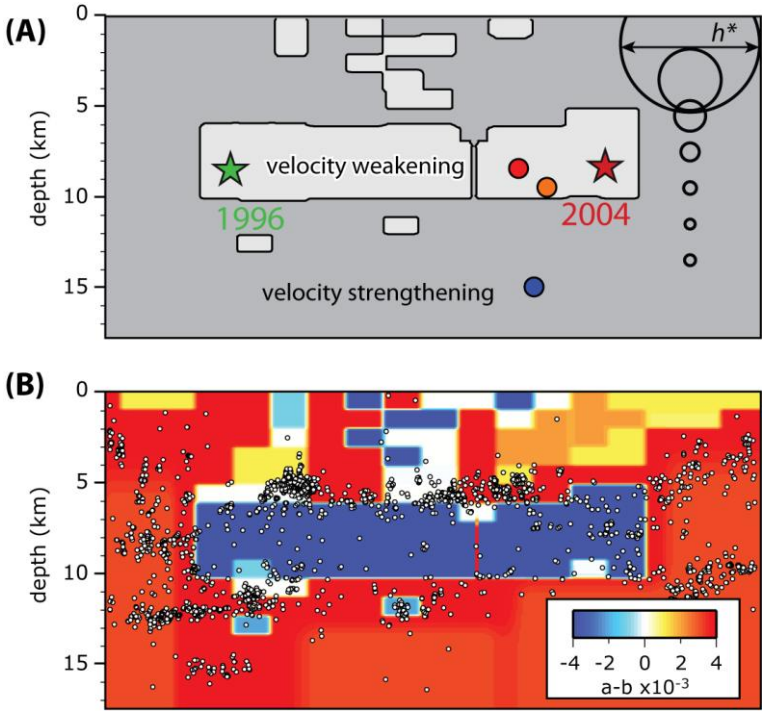


Figure 9. Rate-and-state fault properties of the Parkfield model [130]. A) The distribution of velocity-weakening (lighter grey) and velocity-strengthening (darker grey) friction on the fault, including some shallow asperities with sizes smaller than the nucleation size at that depth. The variation of the nucleation size with depth is indicated by the black circles. Slip velocity at points indicated by small filled circles is plotted in Figure 11. B) The spatial distribution of friction parameter $a - b$, reflecting the conceptual model in (A). The characteristic slip L = 4mm is kept constant in this model.

aseismic, below 0.1 m/s. After the coseismic rupture, the seismogenic zone locks (with velocities much smaller than the plate rate) and afterslip starts to expand around it (Figure 10c). Finally, once afterslip ceases, the stable-friction areas of the fault creep steadily during a phase of interseismic loading. During that period, some creep penetrates into the seismogenic zone and only a part of the seismogenic area stays locked (Figure 10d). The extent of this creep-in is controlled by the assumed parameters, including L, and would be smaller for smaller values of L. Many accelerated creep episodes occur in the velocity-strengthening to velocity-weakening (VS/VW) transition zone, giving rise to a more complex spatio-temporal pattern of slip

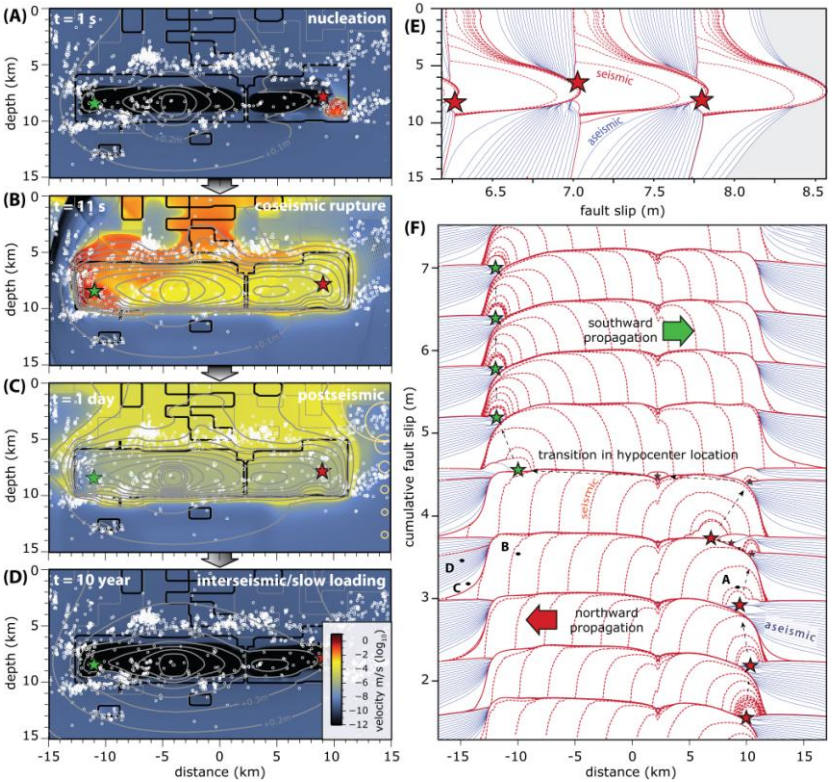


Figure 10. Model response that reproduces the entire seismic cycle at Parkfield (modified from Figures 2 and 4 of [130]. (A)-(D) Slip rate snapshots during a M_w 6.0 seismic cycle, with rupture nucleating spontaneously near the 2004 earthquake hypocenter (A), propagating to the north and rupturing the entire seismogenic zone (B), and followed by a slow postseismic transient (C), with interseismic loading of the partially locked seismogenic zone (D). Another M_w 6 event nucleates 20 years later. Zero time is chosen for plotting convenience. The solid grey profiles indicate the contours of the cumulative slip at 0.1 m intervals. E) Vertical cross section of the slip evolution (red profiles are every 1 s for seismic periods and blue profiles are every 1 yr for aseismic periods). F) Strike-parallel profile of the slip evolution showing a transition in the hypocenter location after 5 M_w 6 seismic events on one side.

evolution in this area during the interseismic period (Figure 11). The cycle repeats with a recurrence time of about 20 years.

The model is capable of qualitatively reproducing the longer seismic cycle and the switch in the hypocenter location similar to the 2004 Parkfield

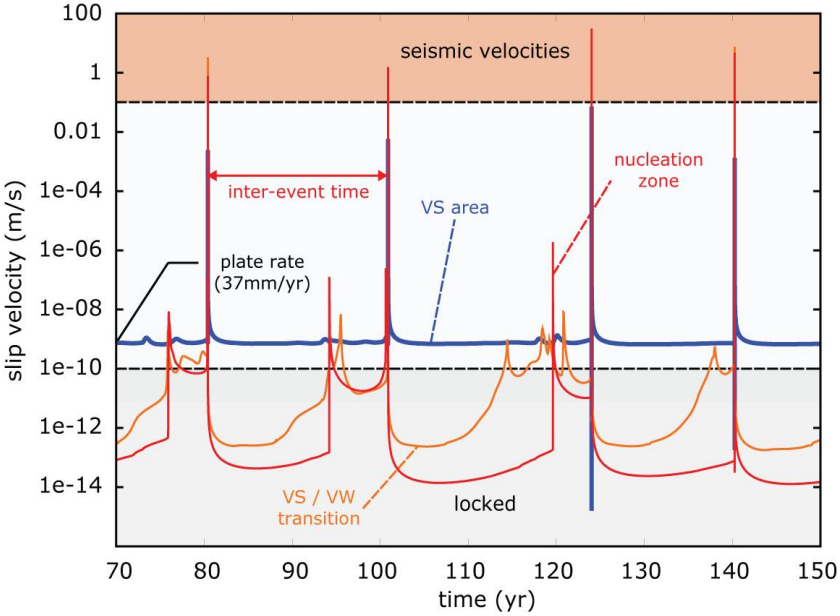


Figure 11. Slip velocity during several earthquake cycles in the middle of the seismogenic velocity-weakening (VW) zone (red line), in the velocity-strengthening (VS) area (blue line), and at the transition between the two domains (orange line). The VW point slips at seismic speed during M_w 6 events, but also experiences slow accelerated transients in the interseismic period. The velocity-strengthening point accelerates after an event, generating afterslip. The point at the transition shows a complex behavior in the second half of the interseismic period. The positions of the points are shown in Figure 9.

event (Figure 10f). This means that such deviations can result from a spontaneous behavior of rate-and-state friction faults, after a complex sequence of foreshocks and smaller interseismic events. The southern and northern corners of the seismogenic zone are two favorable nucleation sites due to their location near a stress concentration at the boundary between two domains of stable and unstable friction. Note that the velocity strengthening region to the south of the seismogenic zone is included as a convenient way to introduce the kind of barrier effect needed to account for the repetition of similar events arresting in that area; the actual nature of the transition from the Parkfield segment to the locked segment further to the south requires further study. In the model, the transition from one nucleation site to the other occurs after a few smaller earthquakes (M_w 2-4) that can be interpreted as

failed nucleations of the main event. The transition is coincidental with the longest simulated recurrence time (($T_r^{\max}$ = 23.1 yr, compared to the smallest $T_r^{\min}$ = 15 yr) of the sequence (Figure 10f).

The rate-and-state modeling of the Parkfield sequence of earthquakes demonstrates the possibility of creating comprehensive physical models of fault zones that integrate geodetic and seismological observations for all stages of the earthquake source cycle. As computational resources and methods improve, more realistic fully dynamic simulations - allowing for a wider range of earthquakes magnitude occurring on a set of interacting faults - will become possible. Such simulations could in principle be used to assess the full range of earthquake patterns that a particular fault system might produce, or assimilate observation about past earthquakes and interseismic loading to assess future seismicity.

4.3. Reproducing scaling of small repeating earthquakes

Rate-and-state models qualitatively similar to that of section 4.1 - but with smaller nucleation sizes - can reproduce several observations about small repeating earthquakes [82]. Repeating earthquakes are seismic events that repeatedly occur in the same location with similar seismic signal. Sequences of small repeating earthquakes have been found on a number of faults [149–155]. Since their recurrence times range from a fraction of a year to several years and their locations are known, small repeating earthquakes are an excellent observation target. This has been exploited in a number of studies, such as the San Andreas Fault Observatory at Depth (SAFOD) drilling project [156] (Figure 12). Repeating earthquakes are used to study an increasingly richer array of problems, from fault creeping velocities and postseismic slip to earthquake interaction and stress drops [149–155,157–161]. To assimilate and properly interpret the wealth of data on small repeating earthquakes, it is important to construct a realistic model of their occurrence.

One of the intriguing observations about small repeating earthquakes is the scaling of their seismic moment M_0 with the recurrence time T as

$$T \propto M_0^{0.17}. \tag{25}$$

This scaling has been first pointed out by [151] for repeating earthquakes along the Parkfield segment of the San Andreas fault, and it has since been confirmed in other tectonic environments [155]. However, a simple conceptual model of these events as circular ruptures, with stress drop $\Delta\tau$

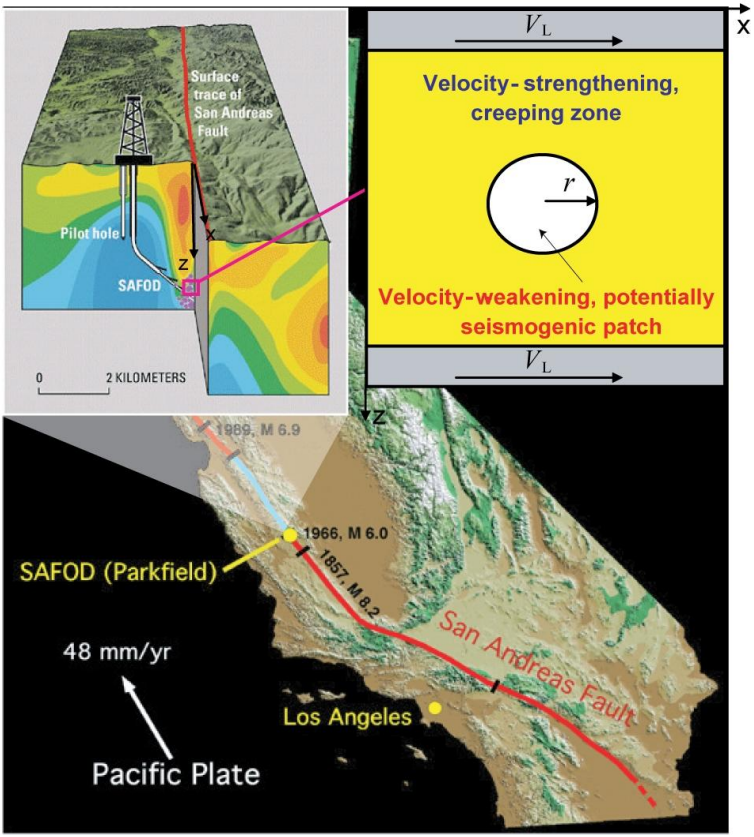


Figure 12. Schematics of the model for simulations of small repeating earthquakes, such as the targets of SAFOD drilling project on the Parkfield segment of the San Andreas fault (Figure 1 from [82]). A segment of a vertical strike-slip fault is considered, which is embedded into an elastic medium and governed by rate-and-state friction laws. On the fault, a small, potentially seismogenic, patch with steady-state velocity-weakening properties (white) is surrounded by a creeping, velocity-strengthening segment (yellow). Outside of the simulated fault segment, steady sliding is imposed with the long-term slip velocity V_L. The creeping (yellow) zone is chosen to be large enough so that the model behavior does not depend on its size. (SAFOD schematics courtesy of Dr. Hickman).

independent of the seismic moment M_0 and slip equal to $V_L T$, where V_L is the long-term slip velocity accommodated by the fault segment, results in [151, 162]

$$T = \Delta\tau^{2/3} M_0^{1/3} / (1.81 \mu V_{\mathrm{L}}) \propto M_0^{1/3}, \tag{26}$$

where μ is the shear modulus of the bulk. Note that this model assumes that all slip at the location of repeating earthquakes is accumulated seismically. The observed and theoretical scaling have different exponents as well as different absolute values of recurrence times, with the observed recurrence times being much larger than the theoretically predicted ones for moment magnitudes typical for small repeating earthquakes. In [151], the discrepancy was interpreted as an indication of the dependence of stress drop on seismic moment, with higher stress drops for smaller events; to fit the observed recurrence times, stress drops for the smallest repeating earthquakes would have to be as high as 2500 MPa [151]. Not only the physical basis for stress drops of such high values is unclear [162], but also recent seismic estimates of stress drops for repeating earthquakes in Parkfield [161, 163] have pointed to values of the order of 1 to 10 MPa, the typical range for earthquakes in general [164].

Based on the rate-and-state view of earthquake physics, a potential explanation for the inadequacy of the theoretical model is its assumption of all slip at the location of small repeating earthquakes being seismic. An obvious model for a repeating earthquake sequence in a creeping segment is that of a steady-state velocity-weakening patch embedded into a larger creeping velocity strengthening region. Since velocity-weakening patches below the critical (nucleation) size should be completely aseismic, it is reasonable to assume that slightly larger patches would still have significant aseismic slip while also producing seismic events. The potential importance of aseismic slip at the location of repeating earthquakes was highlighted in the study by Beeler et al. [162] which used a spring-slider (one-degree-of-freedom) model governed by a constitutive law that incorporated strain hardening in the interseismic period. In the model, part of the accumulated slip was aseismic, due to strain-hardening behavior, and the resulting scaling of the seismic moment with the recurrence time had a trend similar to the one observed. However, it was pointed [162] out that there was no experimental evidence for the strain-hardening law used in the model.

To verify the potential of rate-and-state models to explain the scaling, Chen and Lapusta [82] numerically studied a rate-and-state model of small repeating earthquakes (Figure 12), adopting parameter values typical of laboratory experiments, including the characteristic slip distances L of the order of 10-100 microns. They showed that the model indeed produces repeating sequences of earthquakes with magnitudes from 1 to 4; sequences of different magnitudes can be obtained either by varying the radius of the

velocity-weakening patch, or by adopting different rate-and-state parameters over the patch. Consistently with expectations, significant aseismic slip occurs on patches which are just large enough to produce seismic events (more than 99% of total slip is aseismic in one of the simulations). The ratio of aseismic to total slip decreases as the patch radius increases with respect to the nucleation size.

Remarkably, the model spontaneously reproduces the right exponent of the scaling between the seismic moment and the recurrence time, both (i) for the case of fixed rate-and-state properties and variations in the patch radius (Figure 13a) and (ii) when the rate-and-state parameters *a*, *b*, and the patch radius are all varied (Figure 13b). To reproduce the absolute values of the recurrence times, the loading velocity - which represents the long-term creeping rate in the area of small repeating earthquakes - needs to be of the order of 4.5 mm/yr, within the range of 4 to 35 mm/yr suggested for the portion of the San Andreas fault with repeating earthquakes [165]. Moreover,

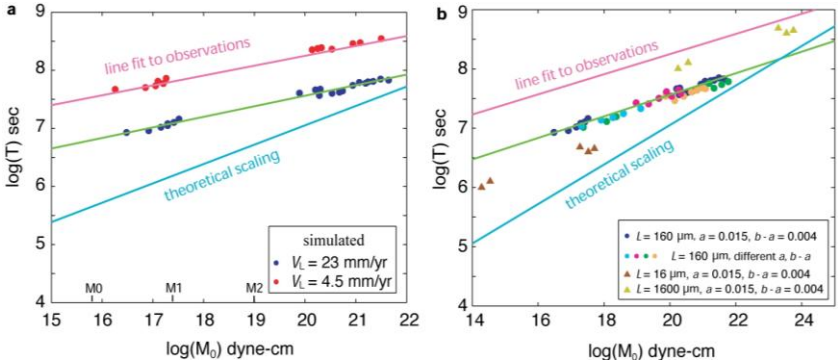


Figure 13. Scaling of seismic moment M_0 with recurrence time T for the simulated sequences of repeating earthquakes (modified from Figure 7 of [82]). (a) Simulation results for long-term slip velocities V_L of 23 mm/yr and 4.5 mm/yr are shown as blue and red dots, respectively. For each V_L, earthquakes of different sizes are obtained by varying the radius of the velocity-weakening patch; all other model parameters are the same. The line fit to the observations in [151] and the scaling in the simple theoretical model (26) are indicated by the magenta and light blue lines, respectively. The best fit to simulations with V_L = 23 mm/yr is shown by the green line. For both values of V_L, the observed scaling exponent is reproduced; the simulations with V_L = 4.5 mm/yr also fit the absolute values of the recurrence times. (b) Simulation results for V_L = 23 mm/yr, L = 160 μm, and several different combinations of rate-and-state parameters *a* and *b* all produce the same scaling, with events of all magnitudes between M_w = 0.3 and M_w = 3.7. V_L = 23 mm/yr was used for computational efficiency. Simulations for different values of L also reproduce the observed scaling exponent.

since repeating earthquakes occur in the transition region next to the partially locked segment that produces occasional M_w 6.0 Parkfield earthquakes, it is indeed possible that the local slip rate there is lower than the overall long-term rate of the entire creeping segment.

The model is also capable of reproducing the response of small repeating earthquakes to accelerated loading due to postseismic slip after the M_w 6 2004 Parkfield event [119] (Figure 14). Analysis of 34 repeating earthquake sequences (RES) from 1987 to 2009 at Parkfield shows that, following the 2004 M_w 6 earthquake, many of the repeating events had greatly reduced recurrence intervals T that systematically increased with time after the 2004 event. This expected behavior of the recurrence time is easily reproduced by the model. However, in addition to this change in timing, observations also point to systematic changes in seismic moment: some sequences experience significant increase in seismic moment due to higher loading rates (and hence

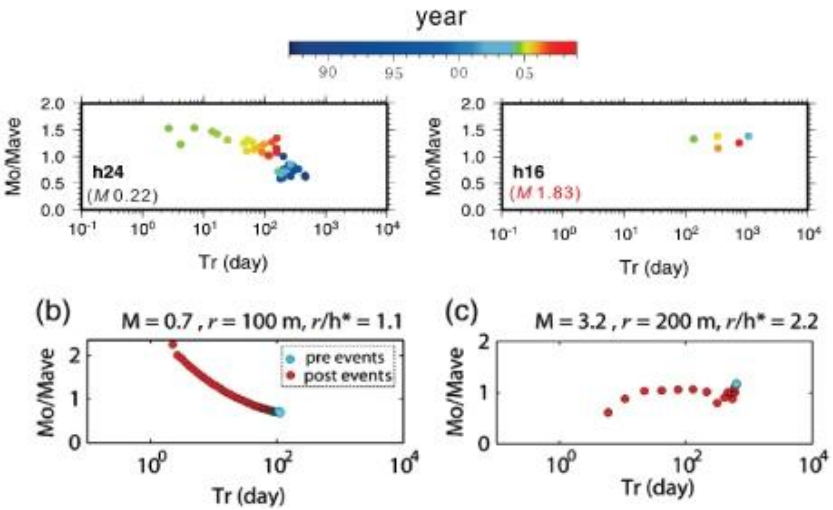


Figure 14. Response of repeating earthquakes to accelerated creep due to the 2004 Parkfield earthquake (modified from Chen et al. [119]). Top: Observations for two repeating sequences. Blue colors indicate repeating earthquakes before the 2004 event, and green to red colors indicate repeating earthquakes after the 2004 event. Notice that sequence h24 has larger moments for shorter recurrence times, whereas sequence h16 has nearly constant moments. Bottom: Results of our simulations with variable loading rate mimicking postseismic slip of the 2004 event. Blue and red indicate events before and after 2004, respectively. The model qualitatively matches the observed behavior, with the different dependence of the moment on the recurrence time for different values of r/h^*.

shorter recurrence times), while other sequences have a small decrease. The latter behavior is what one would expect based on shorter recurrence time and laboratory-based ideas of strengthening in stationary contact (e.g., [10, 27]).

The rate-and-state model is capable of reproducing both behaviors, for different radii r of the earthquake patches as quantified by the ratio r/h^*. Ratios of $r/h^* \sim 1$ produce higher moments M_0 for shorter recurrence times T (or negative M_0-T slopes), whereas larger ratios of r/h^* yield weak positive slopes. Given the same nucleation size h^* (i.e., the same frictional properties and effective normal stress), smaller radii and hence smaller seismic moments result in negative M_0-T slopes, whereas larger radii and hence larger moments lead to weak positive M_0-T slopes, which are consistent with observations. In the models, such variations in seismic moment are caused by the fact that, for $r/h^* \sim 1$, only part of the velocity-weakening patch is ruptured seismically by the event, with most of the slip occurring on the patch aseismically. When the loading rate is increased, the creeping rate on the velocity-weakening patch is also increased, resulting in higher stress levels and enabling seismic events to propagate farther into the patch. This increase in the rupture area creates events with larger moments for shorter recurrence times. For larger ratios of r/h^*, seismic events rupture the entire patch at all loading rates, and hence the rupture area does not vary appreciably. Then the expected effect based on fault strengthening takes over and creates ruptures with smaller moment for smaller recurrence times.

Subsequent studies [166] (manuscript in preparation) have shown that the scaling between the recurrence time and seismic moment is reproduced not only in the models with the aging form of the state variable evolution used in [82], but also in the models with the other common forms including the slip law. At the same time, there are important differences between the results with the different evolution laws, consistent with prior studies of these laws [80,84]. In particular, simulations with the slip law are more prone to unstable slip than those with the aging law, resulting in smaller nucleation sizes and lacking the type of events that rupture only a part of the velocity-weakening patch. The difference in the qualitative features of the model response between the aging and slip laws highlights the importance of using realistic frictional behavior. One of the phenomena not included in either formulation is the potential inelastic shear-induced dilatancy (10) [24, 51, 52, 54]. During nucleation and hence increasing slip rates, such dilatancy may tend to increase the pore space, lower the pore pressure, and hence increase the effective normal stress and frictional resistance, stabilizing fault slip and potentially leading to more aseismic slip.

Since one of the differences between the aging and slip formulations is more aseismic slip during nucleation in the case of the aging law, the aging law may actually be a good proxy for the more complex formulations that involve inelastic dilatancy.

4.4. Rate-and-state models for aftershock sequences

One of the most robust earthquake-related observations is the occurrence of aftershock sequences after seismic events; the decay of aftershocks is well described empirically by Omori's law (e.g., [167])

$$R = \frac{K}{(c+t)^p}, \quad p = 1, \tag{27}$$

where R is the aftershock rate, K, c, and p are empirical constants, with p equal to 1 in the law by Omori and ranging from 0.7 to 1.5 in more recent observations.

Omori's law of the aftershock decay can be explained by a model based on rate-and-state friction, as was discovered by Dieterich [98]. In the model, a pre-existing population of rate-and-state nucleation sites is perturbed by static stress changes due to the mainshock. In the population, each nucleation site is governed by the same rate-and-state nucleation process but time-shifted in such a way that the population would result in a constant background seismicity rate if left unperturbed. After a positive static shear stress step, the nucleation process at each site accelerates, producing an increased seismicity rate (or aftershock rate) that matches Omori's law for a wide range of parameters.

An important ingredient in this aftershock model is the nucleation process and its response to static stress changes. In the model by Dieterich [98], the nucleation process was specified in terms of its slip-velocity evolution. To obtain the evolution, two simplifications in modeling nucleation were used: (i) Elastic interactions were described by a one-degree-of-freedom spring-slider system and (ii) the assumption $V\theta / L \gg 1$ was used to simplify the rate-and-state friction formulation based on a study of earthquake nucleation in a continuum model [79]. These simplifications allowed the derivation of analytical expressions for both slip-velocity evolution during nucleation and the resulting aftershock rate R:

$$\frac{R}{r} = \frac{1}{[\exp(-\Delta\tau/(a\bar{\sigma})) - 1]\exp(-t/t_a) + 1}, \tag{28}$$

where r is the background seismicity rate, $\Delta\tau$ is a positive shear stress change, $t_a = a\bar{\sigma}/\dot{\tau}$, and constant stressing rate $\dot{\tau}$ is assumed before and after the stress step. The work [98] also considered scenarios with variable stress steps and stressing rates; a review of applications is given in [78].

From equation (28), this model has two parameters: $\Delta\tau/(a\bar{\sigma})$ and $t_a = a\bar{\sigma}/\dot{\tau}$. Figure 15 illustrates the resulting aftershock rates and shows that t_a is related to the aftershock duration, since the earthquake rate becomes close to the background rate for $t \geq t_a$. From (28), $R/r = 1$ for $t \gg t_a$. For $t \ll t_a$, $\exp(-t/t_a) \approx (1 - t/t_a)$ and from (28):

$$R = \frac{K}{(c+t)^p}, \quad p = 1, \tag{29}$$

$$K = \frac{r t_a}{1 - \exp\left(-\Delta\tau/(a\bar{\sigma})\right)}, \tag{30}$$

$$c = \frac{t_a}{\exp\left(\Delta\tau/(a\bar{\sigma})\right) - 1}. \tag{31}$$

Hence the model of Dieterich [98] interprets parameters K and c of Omori's law, which were originally introduced as empirical constants. The time interval in which the aftershock rates in this model follow the power law decay of aftershocks depends on the values of $\Delta\tau/(a\bar{\sigma})$ and t_a. For times right after the instability, we have $\exp(-t/t_a) \approx 1$ and $R/r = \exp[\Delta\tau/(a\bar{\sigma})]$. This "plateau" or constant aftershock rate right after the mainshock is shorter for larger values of $\Delta\tau/a\bar{\sigma}$ (Figure 15).

The approach of Dieterich [98] has been further explored in a number of works [168–170]. In particular, aftershock rates based on simulations in spring-slider systems with the full aging rate-and-state formulation were found to follow the results of [98] quite well, validating simplification (ii) for spring-slider models.

Since static stress changes $\Delta\tau$ due to earthquakes are relatively well constrained, aftershock observations can be used to determine the product $a\bar{\sigma}$ in the light of this model (e.g., [105,106,110]). For the model to be consistent with observations, $a\bar{\sigma}$ has to be of the order of 0.01-0.1 MPa (e.g., [106, 171]). Larger values of $a\bar{\sigma}$, of order 1 MPa, are predicted by laboratory values of a (of the order of 0.01) and $\bar{\sigma}$ comparable to overburden minus hydrostatic pore pressure at typical seismogenic depths (of the order of 100 MPa). If

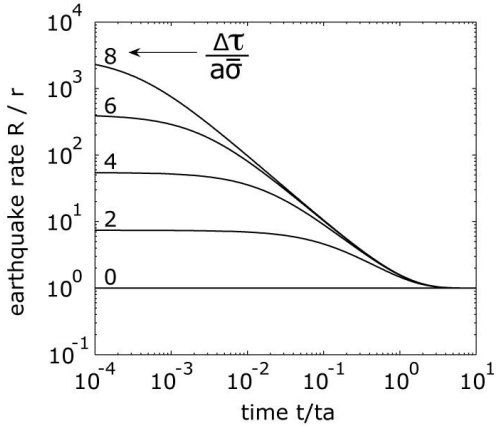


Figure 15. Aftershock rates for the analytical solution of Dieterich [98] given by equation (28) (Figure B1 from [113]. The aftershock rate R is normalized by the background rate r, and the time t after the mainshock is normalized by the aftershock duration t_a. Each curve is computed for a normalized stress step, $\Delta\tau/(a\bar{\sigma})$, with the indicated value. Adapted from Figure 2 of [98].

aftershock production is dominated by static stress triggering of preexisting nucleation sites as described by [98], then, at least on parts of faults where aftershocks nucleate, either the direct effect coefficient a is much smaller than in the laboratory, or effective normal stress $\bar{\sigma}$ is abnormally low.

Kaneko and Lapusta [113] extended the ideas of this approach to continuum models. Instead of perturbing a nucleation process based on the analytical simplified spring-slider solution as done in [98], they created two plausible scenarios of earthquake nucleation in 2D continuum models. The first model incorporates uniform steady-state velocity-weakening friction properties and a weaker patch of slightly (10%) lower effective normal stress where the earthquakes nucleate. The second model contains a rheological transition from steady-state velocity-strengthening to steady-state velocity-weakening friction. Both are relevant to natural faults: Weaker patches may exist for a number of reasons including local fault non-planarity or spatial variations in pore pressure; observations suggest that earthquakes tend to cluster at inferred transitions from locked to creeping regions (e.g., [143,144]) which create stress concentrations that promote earthquake nucleation. Nucleation processes in these models are obtained by simulating spontaneously occurring earthquake sequences under slow, tectonic-like loading following an approach similar to sections 4.1-4.3 but in 2D [34] (Figure 16).

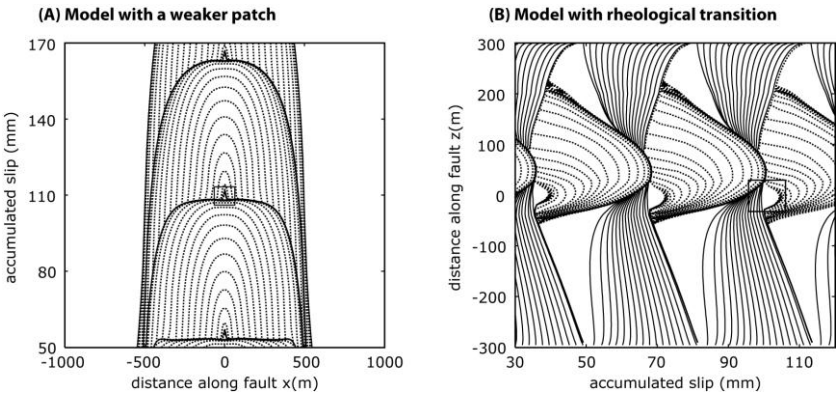


Figure 16. Examples of earthquake sequences simulated (A) in the model with a weaker patch and (B) in the model with rheological transition (Figure 3 from [113]). Solid lines show slip accumulation every 2 years. Dashed lines are intended to capture dynamic events and are plotted every 0.01 seconds during the simulated earthquakes. The nucleation process of a representative earthquake is indicated by a rectangle in both panels.

Numerical study of the response of the obtained nucleation processes to static stress perturbations have revealed several interesting findings. First, the nucleation size of the earthquakes triggered by the static stress change can be much smaller (Figure 17), which implies that the nucleation size depends on the loading history [80]. Second, positive static stress change can *delay* the upcoming earthquake instead of making it occur sooner (Figure 18, panels B, C, I). The positive static stress change does accelerate the slip in the nucleation zone but the acceleration results in an aseismic transient slip event instead of an earthquake. The transient relieves the stress in the nucleation location, slowing the progression of the nucleation and hence delaying the following seismic event. Such a response occurs for the nucleation sites at the rheological transitions.

The results of the numerical perturbation analysis have been converted into aftershock rates following the procedure of Dieterich [98]. By perturbing the simulation at various times before an earthquake (or various times to instability) and considering the resulting response (and, in particular, determining the new, perturbed time to instability), the relation between the unperturbed and perturbed times to instability can be numerically established. From this information, one can construct the aftershock rates for a population of such nucleation sites that would result in a uniform background rate if left unperturbed.

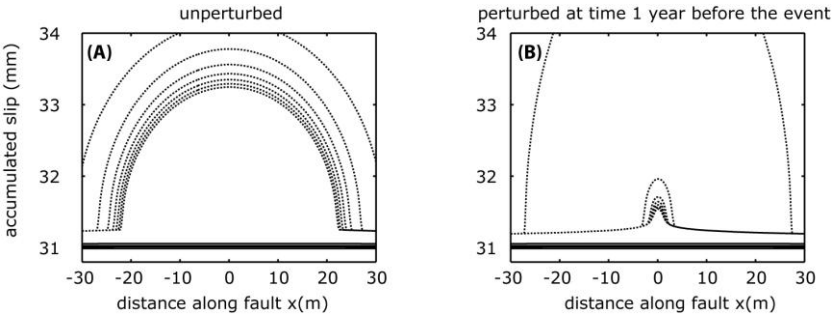


Figure 17. Dependence of nucleation processes on loading history (part of Figure 6 from [113]. (A) Unperturbed nucleation process in a model with a weaker patch. (B) The same nucleation process but perturbed with $\Delta\tau$ = 3.0 MPa one year before the original instability. The dashed lines in (A) and (B) are plotted every 0.01 seconds staring with 0.05 seconds before the onset of instability as defined in [113]. The first 5 lines cluster, indicating the spatial extent of the nucleation zone. The nucleation sizes in these two cases differ by an order of magnitude.

Overall, aftershock rates based on the nucleation processes at weaker patches behave similarly to the theoretical finding (28) based on spring-slider models, with some notable deviations. In particular, aftershock rates are affected by normal-stress heterogeneity in the nucleation zone.

Aftershock rates based on the nucleation processes at rheological transitions behave quite differently, producing pronounced peaks and seismic quiescence (Figure 18, panel E). This is consistent with the complex behavior in which positive stress steps sometimes delay nucleation of seismic events by inducing aseismic transients, as already discussed. Interestingly, superposition of such complex aftershock responses for spatially variable stress changes (Figure 19a) results in Omori's law for a period of time followed by seismic quiescence (Figure 19b). Such behavior was observed at the base of the seismogenic zone near the 1984 Morgan Hill earthquake (Figure 19c). Note that the resulting aftershock rate is much higher than the one based on spring-slider model; such higher rates would result in $a\bar{\sigma}$ closer to the laboratory values.

Kaneko and Lapusta [113] linked the computed aftershock rates to the unperturbed slip-velocity evolution in the nucleation zone and constructed simplified analytical scenarios that explain some features of the response.

The qualitative differences between the two continuum nucleation models suggest that much remains to be learned about the aftershock response of rate-and-state faults to static stress changes. In part, such response

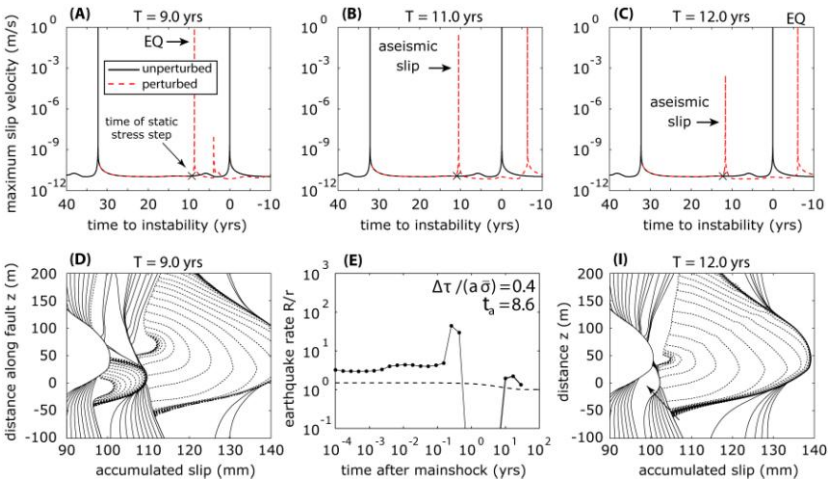


Figure 18. Response to static stress steps in the model with rheological transition (based on Figures 9 and 10 of [113]). (A)-(C): Solid lines show the unperturbed evolution of maximum slip velocity within a part of the velocity-weakening region that contains the nucleation zone. Times of the stress perturbation are given on the top of each panel and marked by '×'. Red dashed lines indicate the resulting perturbed behavior. (D),(I): The effect of stress perturbations on the pattern of slip accumulation. Panel E: Comparison between the aftershock rate based on the continuum model (solid dotted line) and the analytical expression (28) based on spring-slider models (dashed line). Note that a positive shear stress step can delay the timing of the subsequent earthquake by inducing an aseismic transient (panels B, C, I); in panel I, the transient is indicated by the tip of an arrow. Such complex behavior results in the complex aftershock rate response (panel (E).

may depend on the conditions under which nucleation tends to occur on natural faults and may be different from predictions based on spring-slider models [113].

Given that rate and state friction laws have been successfully used to reproduce and explain a number of earthquake phenomena and that mainshocks cause static stress changes, it is reasonable to assume that at least some, and perhaps most, of aftershocks are caused by static triggering of rate-and-state nucleation processes. At the same time, a number of studies have proposed models of aftershocks based entirely on other mechanisms. These include increased loading rate due to aseismic processes such as postseismic slip (e.g., [173, 174]) or relaxation of the viscoelastic lower crust (e.g., [175]); pore fluid motion and induced variations in fault strength (e.g., [176, 177]); triggering due to dynamic stress changes (e.g., [178–180]); and evolution of

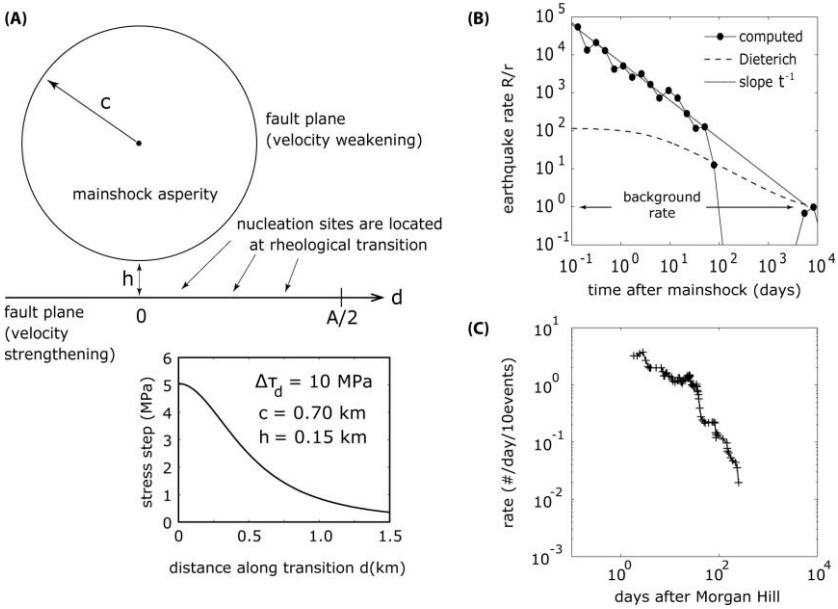


Figure 19. (a) Model for estimating aftershock rates due to a population of nucleation sites located along a segment of rheological transition perturbed by a mainshock asperity (based on Figures 11 and 12 of [113]). Due to the distance from the asperity, the nucleation sites along the segment experience a non-uniform stress step. (b) The resulting aftershock response. It reproduces Omori's law in a limited time period followed by seismic quiescence. Note that this response is a combination of complex non-monotonic responses of the type shown in Figure 18E. The aftershock rate based on [98], for the same friction and stress parameters, is shown by the dashed line. The slope of t^{-1} is plotted for reference. This model produces Omori's law in a limited time period followed by seismic quiescence. For the period of the power-law decay of aftershocks, this model produces much higher aftershock rates than predicted by the model of [98]. (c) From [172], courtesy of Y. Tian. The observed seismicity rate vs. time for a cluster of the 1984 Morgan Hill aftershocks that occurred at a depth appropriate for rheological transition. The multiplet approximately followed Omori's law, but seismicity terminated about one year after the Morgan Hill earthquake. This behavior is qualitatively similar to that of the computed aftershock rates in (b).

viscoelastic damage rheology due to sudden increase in strain (e.g., [181]). The approach of [113] presented here can be used to study the combined effect of two or more mechanisms on aftershock rates. Such combined models would help investigate the relative importance of different aftershock-producing mechanisms.

5. Insights from rate-and-state models and future challenges

One of the main insights emerging from the ability of rate-and-state models to reproduce a broad range of fault slip observations is that laboratory-based rate-and-state formulations truly capture the basics of the fault-zone friction response. The exciting but sobering news is that even that basic response appears to be far from simple. However, this should not be surprising given the complexity of the observed fault behaviors which range from earthquakes, to creep with the plate rate, to many phenomena in between, such as transient slow slip events at the base of many seismogenic zones (e.g., [182]). Furthermore, fault slip is characterized by substantial complexity and variations; e.g., earthquakes in the same area often vary in their characteristics and spatial extent (e.g., [183]).

Much of this complexity can be explained within the standard (Dieterich-Ruina) rate-and-state models due to their rich stability properties (section 3) combined with the interactive nature of long range elastic interactions and inherent nonlinearity of frictional response (sections 4.1-4.4). Models that incorporate spatial variability in rate-and-state properties, even with quite basic patterns, tend to produce complex responses representative of real faults [101, 108, 111, 122, 130]. A study with two velocity-weakening fault segments separated by a velocity-strengthening patch [122] showed that the resulting long-term behavior of the model can be quite complex, with seismic events sometimes rupturing only one of the segments and sometimes both. The probability that the seismic event would propagate through the velocity-strengthening patch and span both patches is linked to a non-dimensional parameter that incorporates properties of velocity-weakening and velocity strengthening patches [122]. Incorporation of a shallow velocity-strengthening fault layer motivated by laboratory measurements leads to shallow coseismic slip deficit [114] which is observed on natural faults [184]; to match the extent of the observed deficit, incorporation of off-fault inelastic processes may be required [128]. Such shallow velocity-strengthening layers also suppress transition of rupture to supershear speeds at the free surface [114] which would otherwise be quite common on weakening interfaces embedded into elastic medium [121]. (Note that supershear rupture propagation have been inferred from observations (e.g., [185]) and observed in experiments [186] but subshear strike-slip ruptures are more common [187]). Models with regions of high characteristic slip L or elevated pore pressure p at the base of the seismogenic zone - both of which result in large values of the nucleation scale (section 3) - are capable of reproducing transient slow slip events [88, 90]. In another model of slow slip events [112], the steady-state dependence of friction on slip rate is varied, changing

from velocity-weakening to velocity-strengthening with increasing slip rates, as motivated by some experiments.

Models of earthquake sequences and fault creep enabled by the rate-and-state formulations have led to a number of general insights. One of them is the importance of assuming self-consistent initial conditions in dynamic rupture simulations, in which features of the simulated earthquakes depend both on the fault friction (or strength) and on the assumed initial stress distribution (e.g., [18,188]). The two distributions - of fault strength and stress - are typically assumed independently, but on natural faults they are linked by stress redistribution due to prior slip. Such redistribution is captured by the long-term simulations of rate-and-state fault models [83, 125]. For example, in a dynamic rupture simulation, one might model a local asperity by higher resistance due to higher normal stress, but assume constant shear prestress; the breakage of such stronger asperity can then lead to a supershear burst [188]. However, simulations of several earthquake cycles in such model reveal that, in the longer history of the fault, the shear stress becomes higher at the asperity as well, substantially reducing its effect on dynamic rupture and eliminating supershear transition for the parameters studied [83].

Another insight is that seismic and aseismic slip can occur in the same region of the fault at different times. This is because velocity-weakening regions can both creep in areas of the order of nucleation sizes and also support seismic slip. Whether a given area of the velocity-weakening region is experiencing seismic or aseismic slip at any given time depends on its current stress and state conditions, which are the result of all prior slip in that and other fault locations and some external factors (such as stress perturbations from neighboring fault segments or fluid flow). In section 4, two interesting consequences of such behavior were discussed: (a) In a model of small repeating earthquakes (section 4.3; [82]), a large fraction of slip in the earthquake-producing patches can be aseismic, resulting in the observed scaling of seismic moment with the recurrence time; (b) When a rate-and-state nucleation site is perturbed by a favorable stress change, of the kind that should speed up the upcoming earthquake, the seismic event can be delayed instead due to the resulting aseismic transient slip (section 4.4; [113]). Such non-obvious effects would be further amplified in models with additional ingredients such as dilatancy.

Rate-and-state models also reveal the special importance of rheological transitions from velocity-weakening to velocity-strengthening behavior, which provide faults with places of stress concentration. Presence of such rheological boundaries on natural faults has ample laboratory and observational evidence (e.g., [10, 25, 28, 143, 144, 189, 190]), suggesting

that slowly moving, velocity-strengthening regions exist above (at shallow depths, 0-3 km) and below locked regions. The stress concentration on such rheological boundaries promotes earthquake nucleation and clustering of small events [34,81,130]. It can help earthquake propagation by providing a highly-stressed region, to the point of promoting supershear transition, especially in the presence of fault heterogeneity [191,192].

Models of earthquake sequences that include full inertial effects during dynamic events [34, 83] have been used to assess the results of the so-called quasi-dynamic models that have been widely used for simulating fault slip (e.g., [88,97,99,102,108,109,111,129]). In the quasi-dynamic methods, the time-dependent process of wave-mediated stress transfers is ignored, with only the static stress changes incorporated at each time step. That leads, in part, to substantially decreased stress concentrations in front of propagating ruptures. Some consequences of using quasi-dynamic models include substantially lower slip rates, rupture speeds, and overall slips during seismic events [34, 83]. While earthquake sequence patterns in rate-and-state models with relatively simple property distributions are similar for fully dynamic and quasi-dynamic models [34, 83], there are qualitative differences between them, including different locations of earthquake nucleation. Finally, quasi-dynamic models cannot reproduce any phenomena that requires accurate representation of inertial effects, such as the formation of slip pulses due to enhanced velocity weakening [49,193] or supershear transition [186].

While the standard, Dieterich-Ruina rate-and-state formulations (4-6) have been quite instructive and useful in understanding fault behavior, more complex laws are needed to account for a fuller range of laboratory observations as discussed in section 2.2. In particular, two effects related to pore pressure variations - dilatancy and thermal pressurization of pore fluids - have been well established in the laboratory (section 2.2) and demonstrated to have important qualitative effects on fault responses in models. Dilatancy during accelerating shear motion - and the associated potential decrease in pore pressure - is a stabilizing factor [51] (section 3) which enables rate-and-state fault models with dilatancy to reproduce a number of observations about transient slow slip [53,54]. Slow slip events in such models can be viewed as protracted nucleation processes, just like in the standard rate-and-state models, and the assumption of low effective normal stress (and hence elevated pore pressure) is still required, but the slow slip events can be produced for small, laboratory-like values of the characteristic slip L of the order of microns [54]. As slip velocities increase, thermal pressurization of pore fluids (section 2.2) may become important at slip rates as low as 10^{-4} m/s [54] and promote instability, provided the shear zone is narrow enough and impermeable enough.

Thermal pressurization of pore fluids, flash heating (section 2.2), and several other potential coseismic weakening mechanisms (e.g., [57]) may dramatically affect the fault resistance during seismic slip and hence affect the entire earthquake cycle. For example, models that incorporate flash heating and thermal pressurization allow fault operation under lower shear stresses than the ones implied by the low-velocity friction, promote short-duration pulse-like modes of earthquake slip, and match the observed increase of the apparent fracture (or breakdown) energy with slip (e.g., [49, 50, 59]). A number of issues related to these processes require further investigation. For example, mechanisms based on pore pressure variations are strongly affected by the permeability of the surrounding medium. If the fluid can flow in and out of the shearing layer sufficiently freely, shear heating (and dilatancy) would not have much of an effect. The permeability values reported for compressive stresses relevant to seismogenic depths are low enough to make pore pressurization quite efficient (e.g., [59, 74]). However, it has not yet been fully explored how permeability would be affected by strong dynamic stress variations at the rupture front which are bound to damage the surrounding material. Spatial variations of poroelastic properties such as permeability can have dominating effect on long-term and short-term model behavior (e.g., [50]), and it is important to explore the effect of plausible variations with time.

In general, off-fault properties and processes have significant effect on fault slip, and the investigation of the full range of such effects is just beginning. Most fault slip models, especially those of earthquake cycles and long-term fault slip, have used linear elastic media for tractability and simplicity. While long-range elastic interactions are certainly key to the realistic model response, inelastic processes such as fault damage (e.g., [116, 181, 194, 195]) may not only change permeability but directly affect rupture propagation by consuming energy and modifying wave-propagation properties in the vicinity of the fault. A related issue is that of local fault non-planarity (e.g., [116, 124, 196]), which interacts with fault slip to potentially produce substantial stress variations that could dramatically affect fault slip and off-fault damage; the extent of such effects would depend on the scale and amplitude of the non-planarity.

Another important issue is the structure and composition of the shearing layers and their variation in space and time. Narrow, near-zero shear zones of the kind that have been found in some studies of exhumed faults [197] support the notion of extreme localization of shear at seismogenic depths and justify the procedure of applying laws based on small-scale laboratory experiments to fault-scale phenomena. However, the degree of localization and hence the width of the shearing zone may vary on faults, most obviously

with depth due to increasing confinement but also perhaps more generally with different rock compositions and slipping histories. A hint of such complexity is given by the fault structure at the San Andreas Fault Observatory at Depth (SAFOD) drilling site, which features two shear zones within a broader zone (of the order of 100 meters) of altered properties with respect to the surrounding medium [198]. Such a complex structure may be due to the relatively shallow depth (~3 km) and the fact that the fault is creeping in that area, but it is intriguing nonetheless. Furthermore, experimental studies on the rock samples from SAFOD suggest that the frictional response of the fault there is dictated by the presence of weak minerals over a very narrow width, and not by the frictional properties of the surrounding material [44]. The structure and composition of the fault zone can depend on chemical processes, which can induce both healing and weakening. For example, chemical decomposition due to shear heating during rapid slips has been shown to release pore fluids such as CO_2 and promote fault weakening due to pore pressurization [70], while leaving an altered material behind. On the flip side, chemical processes during long interseismic periods should result in healing, perhaps in a manner heterogeneous over the fault [199].

The issue of the width and response of the shearing zone is also quite important below the so-called seismogenic zone. Do faults, at least the ones which are mature and represent major plate boundaries, have deeper fault extensions, e.g., relatively narrow zones of shear that extend below the so-called brittle-ductile transition? Several observations suggest that the answer is yes, including localized post-seismic slip [123,142,200] and seismic tremors found below the seismogenic zone [201].

If so, then laboratory studies under the relevant temperature and pressure conditions indicate that such zones should be mostly velocity-strengthening and hence creep under the slow loading. Such creep with the long-term fault slip rate is one of common forms of incorporating loading into simulations of long-term fault slip [34, 50, 83, 94, 97]. However, if much of the deformation below the fault is more broadly distributed [123, 142, 200], the fault loading may have different character which would be important to explore; this may be especially important for models with interacting faults.

In summary, the developments in rate-and-state fault modeling have been quite exciting. The laboratory-based models have succeeded in reproducing a wide range of fault slip phenomena, both qualitatively and quantitatively. Rapid advances in experiments, theories, observations, and computational methods are increasing the number of fault slip studies based on materials science of rock deformation, where each fault constitutive parameter has a physical meaning and can be measured in the lab or in the field, at least in

principle. This opens a tantalizing possibility of statistically predictive physics-based modeling, in which model parameters are informed by all available laboratory and field observations, a suite of simulations is conducted which spans the range of parameter uncertainty, and a range of potential fault behaviors is uncovered. Such modeling output can then contribute to our understanding of seismic hazard. However, much remains to be done toward that goal. The number of the potentially relevant physical phenomena and associated parameters is relatively large as already discussed, including rate-and-state effects, dilatancy, shear heating, fluid flow, a number of potential coseismic weakening mechanisms, local fault non-planarity, and damage processes in the bulk. The current modeling work is moving toward establishing the dominant and/or combined effects of several mechanisms, a must if we are to reduce the number of physically motivated parameters to a manageable number. Most importantly, we need both more measurements related to natural faults at a range of depths (which means drilling) and more laboratory experiments on natural fault samples under relevant temperature/ stress/pore-pressure conditions.

References

1. Terzaghi, K. (1936) The shearing resistance of saturated soils and the angle between the planes of shear. *International Conference of Soil and Foundation Engineering*, Cambridge, Mass, pp. 54-56, Harvard Univ. Press.
2. Dieterich, J. H. (1979) *J. Geophys. Res*., **84**, 2161-2168.
3. Dieterich, J. H. (1981) Constitutive properties of faults with simulated gouge. N. L. Carter, J. M. L., M. Friedman and Stearns, D. W. (eds.), *Monograph 24: Mechanical Behavior of Crustal Rocks*, pp. 103-120, AGU.
4. Ruina, A. L. (1980) *Friction law instabilities: A quasistatic analysis of some dry frictional behavior. Ph.D. Thesis*. Division of Engineering, Brown University.
5. Ruina, A. (1983) *J. Geophys. Res*., **88**, 10,359-10,370.
6. Rice, J. R. (1983) *Pure Appl. Geophys*., **21**, 443-475.
7. Rice, J. R. and Ruina, A. L. (1983) *J. Appl. Mech*., **50**, 343-349.
8. Tullis, T. and Weeks, J. D. (1986) *Pure Appl. Geophys*., **124**, 383-394.
9. Heslot, F., Baumberger, T., Perrin, B., Caroli, B., and Caroli, C. (1994) *Phys. Rev. E*, **49**, 4973-4988.
10. Marone, C. J. (1998) *An. Rev. Earth Planet. Sc*., **26**, 643-696.
11. Scholz, C. H. (1990) *The mechanics of earthquakes and faulting*. 439 pp., Cambidge Univ. Press.
12. Bowden, F. P. and Tabor, D. (1950) *The friction and lubrication of Solids, Part I*. Clarendon Press.
13. Bowden, F. P. and Tabor, D. (1964) *The friction and lubrication of Solids, Part II*. Clarendon Press.
14. Archard, J. F. (1957) *Proc. R. Soc. Lond. A*, **243**, 190-205.

15. Greenwood, J. A. and Williamson, J. (1966) *Proc. R. Soc. London, Ser. A*., **295**, 300-319.
16. Scholz, C. H., Molnar, P., and Johnson, T. (1972) *J. Geophys. Res*., **77**, 6392-6404.
17. Dieterich, J. H. (1978) *Pure Appl. Geophys*., **116**, 790-806.
18. Day, S. M., Dalguer, L. A., Lapusta, N., and Liu, Y. (2005) *J. Geophys. Res*., **110**, 10.1029/2005JB003813.
19. Ida, Y. (1972) *J. Geophys. Res*., **77**, 3796-3805.
20. Palmer, A. C. and Rice, J. (1973) *Proc. R. Soc. London Ser*., **332**, 527-548.
21. Rabinowicz, E. (1951) *J. Appl. Phys*., **22**, 1373-1379.
22. Rabinowicz, E. (1958) *Proc. Phys. Soc. London*, **71**, 668-675.
23. Tullis, T. (1988) *Pure Appl. Geophys*., **126**, 555-588.
24. Marone, C., Raleigh, B. C., and Scholz, C. H. (1990) *J. Geophys. Res*., **95**, 7007-7026.
25. Blanpied, M. L., Lockner, D. A., and Byerlee, J. D. (1991) *Geophys. Res. Lett*., **18**, 609-612.
26. Marone, C. and Kilgore, B. (1993) *Nature*, **362**, 618-620.
27. Beeler, N. M., Tullis, T. E., and Weeks, J. D. (1994) *Geophys. Res. Lett*., **21**, 1987-1990.
28. Blanpied, M. L., Lockner, D. A., and Byerlee, J. D. (1995) *J. Geophys. Res*., **100**, 13045-13064.
29. Beeler, N. M., Tullis, T. E., Blanpied, M. L., and Weeks, J. D. (1996) *J. Geophys. Res*., **101**, 8697-8715.
30. Kato, N. and Tullis, T. E. (2001) *Geophys. Res. Lett*., **28**, 1103-1106.
31. Chester, F. M. (1994) *J. Geophys. Res*., **99**, 7247-7261.
32. Baumberger, T., Berthoud, P., and Caroli, C. (1997) *Phys. Rev. B*, **60**, 3928-3939.
33. Sleep, N. H. (1997) *J. Geophys. Res*., **102**, 2875-2895.
34. Lapusta, N., Rice, J. R., Ben-Zion, Y., and Zheng, G. (2000) *J. Geophys. Res*., **105**, 23765-23789.
35. Nakatani, M. (2001) *J. Geophys. Res*., **106**, 13347-13380.
36. Rice, J. R., Lapusta, N., and Ranjith, K. (2001) *J. Mech. Phys. Solids*, **49**, 1865-1898.
37. Dieterich, J. H. and Kilgore, B. D. (1994) *Pure Appl. Geophys*., **143**, 283-302.
38. Dieterich, J. H. and Kilgore, B. D. (1996) *Tectonophysics*, **256**, 219-239.
39. Dieterich, J. H. (1972) *J. Geophys. Res*., **77**, 3690-3697.
40. Marone, C. J. (1998) *Nature*, **391**, 69-72.
41. Bayart, E., Rubin, A. M., and Marone, C. (2006) *Eos Trans. AGU*, **87(52),** S31A-0180.
42. Reinen, L. A., Weeks, J. D., and Tullis, T. E. (1991) *Geophys. Res. Lett*., **18**, 1921-1924.
43. Carpenter, B. M., Marone, C., and Saffer, D. M. (2011) *Nature Geosci*., **4**, 251-254.
44. Lockner, D. A., Morrow, C., Moore, D., and Hickman, S. (2011) *Nature*, **472**, 82-86.
45. Linker, M. H. and Dieterich, J. H. (1992) *J. Geophys. Res*., **97**, 4923-4940.

46. Prakash, V. and Clifton, R. J. (1993) Time-resolved dynamic friction measurements in pressure-shear. Ramesh, K. T. (ed.), *Experimental Techniques in the Dynamics of Deformable Bodies*, vol. AMD 165, pp. 33-48, Am. Soc. Mech. Eng.
47. Cochard, A. and Rice, J. R. (2000) *J. Geophys. Res*., **105**, 25891-25907.
48. Ranjith, K. and Rice, J. R. (2001) *J. Mech. Phys. Solids*, **49**, 341-361.
49. Noda, H., Dunham, E. M., and Rice, J. R. (2009) *J. Geophys. Res*., **114**, 10.1029/2008JB006143.
50. Noda, H. and Lapusta, N. (2010) *J. Geophys. Res*., **115**, 10.1029/2010JB007780.
51. Segall, P. and Rice, J. R. (1995) *J. Geophys. Res*., **100**, 22155-22171.
52. Segall, P. and Rice, J. R. (2006) *J. Geophys. Res*., **111**, 10.1029/2005JB004129.
53. Rubin, A. M. (2008) *J. Geophys. Res*., **113**, 10.1029/2008JB005642.
54. Segall, P., Rubin, A. M., Bradley, A. M., and Rice, J. R. (2010) *J. Geophys. Res*., **115**, 10.1029/2010JB007449.
55. Santamarina, J. C. and Shin, H. S. (2009) Friction in granular media. Y.H. Hatzor, I. V., J. Sulem (ed.), *Meso-scale Shear Physics in Earthquake and Landslide Mechanics*, pp. 157-188, CRC Press.
56. Andrade, J. E., Chen, Q., Lea, P. H., Avila, C. F., and Evans, T. M. (2012) *J. Mech. Phys. Solids*, **60**, 1122-1136.
57. Tullis, T. E. (2007) Friction of rock at earthquake slip rates. Schubert, G. and Kanamori, H. (eds.), *Treatise in Geophysics*, pp. 131-152, Vol. 4 - Earthquake Seismology, Elsevier. 38.
58. Rice, J. R. (1999) *Eos Trans. AGU*, **80(46)**, F681.
59. Rice, J. R. (2006) *J. Geophys. Res*., **111**, 10.1029/2005JB004006.
60. Beeler, N. M., Tullis, T. E., and Goldsby, D. L. (2008) *J. Geophys. Res*., **113**, 10.1029/2007JB004988.
61. Noda, H. (2008) *J. Geophys. Res*., **113**, 10.1029/2007JB005406.
62. Goldsby, D. L. and Tullis, T. E. (2011) *Science*, 334, 216-218.
63. Sibson, R. H. (1973) *Nature*, **243**, 66-68.
64. Lachenbruch, A. H. (1980) *J. Geophys. Res*., **85**, 6097-6112.
65. Mase, C. W. and Smith, L. (1980) *J. Geophys. Res*., **92**, 6249-6272.
66. Andrews, D. J. (2002) *J. Geophys. Res*., **107**, 10.1029/2002JB001942.
67. Bizzarri, A. and Cocco, M. (2006) *J. Geophys. Res*., **111**, 10.1029/2006JB004759.
68. Bizzarri, A. and Cocco, M. (2006) *J. Geophys. Res*., **111**, 10.1029/2005JB003864.
69. Han, R., Shimamoto, T., Hirose, T., Ree, J.-H., and Ando, J. (2007) *Science*, **316**, 878-881.
70. Sulem, J. and Famin, V. (2009) *J. Geophys. Res*., **114**, 10.1029/2008JB006004.
71. Ferri, F., Di Toro, G., Hirose, T., and Shimamoto, T. (2010) *Terra Nova*, **22**, 347-353.
72. Tsutsumi, A. and Shimamoto, T. (1997) *Geophys. Res. Lett*., **24**, 699-702.
73. Di Toro, G., Hirose, T., Nielsen, S., Pennacchioni, G., and Shimamoto, T. (2006) *Science*, **311**, 647-649.
74. Wibberly, C. A. J. and Shimamoto, T. (2003) *J. Struct. Geol*., **25**, 59-78.

75. Gu, J., Rice, J. R., Ruina, A. L., and Tse, S. T. (1984) *J. Mech. Phys. Sol.*, **32**, 167-196.
76. Blanpied, M. L. and Tullis, T. E. (1986) *Pure Appl. Geophys*., **124**, 415-444.
77. Samuelson, J., Elsworth, D., and Marone, C. (2011) *J. Geophys. Res.*, 116, 10.1029/2011JB008556.
78. Dieterich, J. (2007) Application of rate-and-state-dependent friction to models of fault slip and earthquake occurrence. Schubert, G. and Kanamori, H. (eds.), *Treatise in Geophysics*, pp. 107-129, Vol. 4 - Earthquake Seismology, Elsevier.
79. Dieterich, J. H. (1992) *Tectonophysics*, **211**, 115-134.
80. Rubin, A. M. and Ampuero, J.-P. (2005) *J. Geophys. Res*., **110**, 10.1029/2005JB003686.
81. Lapusta, N. and Rice, J. R. (2003) *J. Geophys. Res*., **108**, 10.1029/2001JB000793.
82. Chen, T. and Lapusta, N. (2009) *J. Geophys. Res*., **114**, 10.1029/2008JB005749.
83. Lapusta, N. and Liu, Y. (2009) *J. Geophys. Res*., **114**, 10.1029/2008JB005934.
84. Ampuero, J.-P. and Rubin, A. M. (2008) *J. Geophys. Res*., **113**, 10.1029/2007JB005082.
85. Richardson, E. and Jordan, T. H. (2002) *Bull. Seism. Soc. Am*., **92**, 1766-1782.
86. Kanamori, H. and Anderson, D. L. (1975) *Bull. Seism. Soc. Am*., **65**, 1073-1095.
87. Hanks, T. C. and Kanamori, H. (1979) *J. Geophys. Res*., **84**, 23482350.
88. Liu, Y. and Rice, J. R. (2005) *J. Geophys. Res*., **110**, 10.1029/2004JB003424.
89. Suppe, J. (2007) Geology, 35, 1127-1130.
90. Liu, Y. and Rice, J. R. (2007) *J. Geophys. Res*., **112**, 10.1029/2007JB004930.
91. Dieterich, J. H. and Linker, M. F. (1992) *Pure Appl. Geophys*., **19**, 1691-1694.
92. Schmitt, S. V., Segall, P., and Matsuzawa, T. (2011) *J. Geophys. Res*., **116**, 10.1029/2010JB008035.
93. Noda, H. and Lapusta, N. (2011) *AGU Fall Meeting*, T42C-03.
94. Tse, S. T. and Rice, J. R. (1986) *J. Geophys. Res*., **91**, 9452-9472.
95. Okubo, P. G. (1989) *J. Geophys. Res*., **94**, 12321-12335.
96. Horowitz, F. G. and Ruina, A. (1989) *J. Geophys. Res*., **94**, 10279-10298.
97. Rice, J. R. (1993) *J. Geophys. Res*., **98**, 9885-9907.
98. Dieterich, J. (1994) *J. Geophys. Res*., **99**, 2601-2618.
99. Ben-Zion, Y. and Rice, J. (1995) *J. Geophys. Res*., **100**, 12959-12983.
100. Perrin, G., Rice, J. R., and Zheng, G. (1995) *J. Mech. Phys. Sol*., **43**, 1461-1495.
101. Boatwright, J. and Cocco, M. (1996) *J. Geophys. Res*., **101**, 13,895-13,909.
102. Rice, J. R. and Ben-Zion, Y. (1996) *Proc. Nat. Natl. Acad. Sci*., **93**, 3811-3818.
103. Tullis, T. E. (1996) *Proc. Nat. Acad. Sci*., **93**, 3803-3810.
104. Ben-Zion, Y. and Rice, J. R. (1997) *J. Geophys. Res*., **102**, 17771-17784.
105. Gross, S. and Bürgmann, R. (1998) *J. Geophys. Res*., **103**, 4915-4927.
106. Toda, S., Stein, R. S., Reasenberg, P. A., Dieterich, J. H., and Yoshida, A. (1998) *J. Geophys. Res*., **103**, 24543-24565.
107. Cocco, M. and Bizzarri, A. (2002) *Geophys. Res. Lett*., **29**, 10.1029/2001GL013999.
108. Hori, T., Kato, N., Hirahara, K., Baba, T., and Kaneda, Y. (2004) *Earth Plan. Sci. Lett*., **228**, 215-226.

109. Kato, N. (2004) *J. Geophys. Res*., *109*, 10.1029/2004JB003001.
110. Toda, S., Stein, R. S., Richards-Dinger, K., and Bozkurt, S. B. (2005) *J. Geophys. Res*., **110**, 17.
111. Hillers, G., Ben-Zion, Y., and Mai, P. M. (2006) *J. Geophys. Res*., **111**, 10.1029/2005JB003859.
112. Shibazaki, B. and Shimamoto, T. (2007) *Geophys. J. Int*., **171**, 191-205.
113. Kaneko, Y. and Lapusta, N. (2008) *J. Geophys. Res*., **113**, 10.1029/2007JB005154.
114. Kaneko, Y., Lapusta, N., and Ampuero, J.-P. (2008) *J. Geophys. Res*., **113**, 10.1029/2007JB005553.
115. Barbot, S., Fialko, Y., and Bock, Y. (2009) *J. Geophys. Res*., **114**, 10.1029/2008JB005748.
116. Dieterich, J. and Smith, D. E. (2009) *Pure Appl. Geophys*., **166**, 1799-1815.
117. Barbot, S. and Fialko, Y. (2010) *Geophys. J. Int*., **182**, 1124-1140.
118. Cocco, M., Hainzl, S., Catalli, F., Enesco, B., Lombardi, A. M., and Woessner, J. (2010) *J. Geophys. Res*., **115**, 10.1029/2009JB006838.
119. Chen, K. H., Bürgmann, R., Nadeau, R. M., Chen, T., and Lapusta, N. (2010) *Earth Planet. Sci. Lett.*, **299**, 118-125.
120. Dieterich, J. H. and Richards-Dinger, K. B. (2010) *Pure Appl. Geophys*., **167**, 1087-1104.
121. Kaneko, Y. and Lapusta, N. (2010) *Tectonophysics*, **493**, 272-284.
122. Kaneko, Y., Avouac, J.-P., and Lapusta, N. (2010) *Nature Geoscience*, **3**, 363-369.
123. Segall, P. (2010) *Earthquake and volcano deformation*. Princeton University Press.
124. Dunham, E. M., Belanger, D., Cong, L., and Kozdon, J. (2011) *Bull. Seism. Soc. Am*., 101, 2308-2322.
125. Jiang, J. and Lapusta, N. (2011) *AGU Fall Meeting*, T23C-2408.
126. Kaneko, Y. and Ampuero, J.-P. (2011) *Geophys. Res. Lett*., **38**, doi:10.1029/2011GL049953.
127. Kaneko, Y., Ampuero, J.-P., and Lapusta, N. (2011) *J. Geophys. Res*., **116**, 10.1029/2011JB008395.
128. Kaneko, Y. and Fialko, Y. (2011) *Geophys. J. Int*., **186**, 1389-1403.
129. Shibazaki, B., Matsuzawa, T., and Tsutsumi, A. (2011) *Geophys. Res. Lett*., **38**, 10.1029/2011GL049308.
130. Barbot, S., Lapusta, N., and Avouac, J. P. (2012) *Science*, **336**, 707-710.
131. Geubelle, P. H. and Rice, J. R. (1995) *J. Mech. Phys. Solids*, **43**, 1791-1824.
132. Shibazaki, B. and Matsu'ura, M. (1992) *Geophys. Res. Lett*., **19**, 1189-1192.
133. Cochard, A. and Madariaga, R. (1996) *J. Geophys. Res*., **101**, 25321-25336.
134. Duan, B. and Oglesby, D. D. (2005) *Bull. Seism. Soc. Am*., **95**, 1623-1636.
135. Bakun, W. H. and Lindh, A. G. (1985) *Science*, **229**, 619-624.
136. Segall, P. and Du, Y. (1993) *J. Geophys. Res*., **98**, 45274538.
137. Bakun, W. H., et al. (2005) *Nature*, **437**, 969-974.
138. Custódio, S., Liu, P., and Archuleta, R. J. (2005) *Geophys. Res. Let*., **32**, 4 10.1029/2005GL024417.

139. Thurber, C., Zhang, H., Waldhauser, F., Hardebeck, J., Michael, A., and Eberhat-Phillips, D. (2006) *Bull. Seism. Soc. Am*., **96**, S38-S49.
140. Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M., and Larose, E. (2008) *Science*, **321**, 1478-1481.
141. Uchide, T., Ide, S., and Beroza, G. C. (2009) *Geophys. Res. Let*., **36**, 10.1029/2008GL036824.
142. Bruhat, L., Barbot, S., and Avouac, J. P. (2011) *J. Geophys. Res*., **116**, 10.1029/2010JB008073.
143. Schaff, D. P., Bokelmann, G. H. R., Beroza, G. C., Waldhauser, F., and Ellsworth, W. L. (2002) *J. Geophys. Res*., **107**, 10.1029/2001JB000633.
144. Waldhauser, F., Ellsworth, W. L., Schaff, D. P., and Cole, A. (2004) *Geophys. Res. Lett*., **31**, 10.1029/2004GL020649.
145. Peng, Z. and Zhao, P. (2009) *Nature Geosc*., **2**, 877-881.
146. Sammis, C. G. and Rice, J. R. (2001) *Bull. Seism. Soc. Am*., **91**, 532-537.
147. Allmann, B. P. and Shearer, P. M. (2007) Science, 318, 1279-1283.
148. Kato, N. (2007) *Geophys. J. Int*., **168**, 797-808.
149. Ellsworth, W. L. and Dietz, L. D. (1990) Repeating earthquakes: characteristics and implications. *Proc. of Workshop 46, the 7th U.S.-Japan Seminar on Earthquake prediction*, pp. 226-245, Open File Report, USGS.
150. Vidale, J. E., Ellsworth, W. L., Cole, A., and Marone, C. (1994) *Nature*, **368**, 624-626.
151. Nadeau, R. M. and Johnson, L. R. (1998) *Bull. Seism. Soc. Am*., **88**, 790-814.
152. Bürgmann, R., Schmidt, D., Nadeau, R. M., d'Alessio, M., Fielding, E., Manaker, D., McEvilly, T. V., and Murray, M. H. (2000) *Science*, **289**, 1178-1182.
153. Igarashi, T., Matsuzawa, T., and Hasegawa, A. (2003) *J. Geophys. Res*., **108**, 10.1029/2002JB001920.
154. Peng, Z.-G. and Ben-Zion, Y. (2005) *Geophys. J. Int*., **160**, 1027-1043.
155. Chen, K. H., Nadeau, R. M., and Rau, R. J. (2007) *Geophys. Res. Lett*., **34**, 10.1029/2007GL030554.
156. Hickman, S., Zoback, M., and Ellsworth, W. (2004) *Geophys. Res. Lett*., **31**, 10.1029/2004GL020688.
157. Marone, C., Vidale, J. E., and Ellsworth, W. (1995) *Geophys. Res. Lett*., **22**, 3095-3098.
158. Schaff, D. P., Beroza, G. C., and Shaw, B. E. (1998) Geophys. Res. Lett., 25, 4549-4552.
159. Nadeau, R. M. and McEvilly, T. V. (1999) Science, 285, 718-721.
160. Matsubara, M., Yagi, Y., and Obara, K. (2005) *Geophys. Res. Lett*., **32**, 10.1029/2004GL022310.
161. Allmann, B. P. and Shearer, P. M. (2007) *J. Geophys. Res*., **112**, 17.
162. Beeler, N. M., Lockner, D. L., and Hickman, S. H. (2001) *Bull. Seism. Soc. Am*., **91**, 1797-1804.
163. Imanishi, K., Ellsworth, W. L., and Prejean, S. G. (2004) *Geophys. Res. Lett*., **31**, 10.1029/2004GL019420.
164. Abercrombie, R. E. (1995) *J. Geophys. Res*., **100**, 24015-24036.
165. Harris, R. A. and Segall, P. (1987) *J. Geophys. Res*., **92**, 7945-7962.

166. Chen, T. (2012) *Part I: Structure of central and southern Mexico from velocity and attenuation tomography; Part II: Physics of small repeating earthquakes. Ph.D. Thesis*. California Institute of Technology.
167. Utsu, T., Ogata, Y., and Matsu'ura, R. S. (1995) *J. Phys. Earth*, **43**, 1-33.
168. Gomberg, J., Beeler, N. M., Blanpied, M. L., and Bodin, P. (1998) *J. Geophys. Res.*, **103**, 24411-24426.
169. Gomberg, J., Beeler, N. M., and Blanpied, M. L. (2000) *J. Geophys. Res.*, **105**, 7857-7871.
170. Gomberg, J., Reasenberg, P., Cocco, M., and Belardinelli, M. E. (2005) J. Geophys. Res., 110, 10.1029/2004JB003404.
171. Belardinelli, M. E., Cocco, M., Coutant, O., and Cotton, F. (1999) *J. Geophys. Res.*, **104**, 14925-14945.
172. Tian, Y. and Rubin, A. M. (2005) *AGU Chapman Conference on Radiated Energy and the Physics of Earthquake Faulting*.
173. Benioff, H. (1951) *Bull. Seism. Soc. Am.*, **41**, 31-62.
174. Perfettini, H. and Avouac, J.-P. (2004) *J. Geophys. Res.*, **109**, 10.1029/2003JB002488.
175. Freed, A. M. and Lin, J. (2001) *Nature*, **441**, 180-183.
176. Nur, A. and Booker, J. R. (1972) *Science*, **175**, 885-887.
177. Bosl, W. J. and Nur, A. (2002) *J. Geophys. Res.*, **107**, 10.1029/2001JB000155.
178. Hill, D. P., et al. (1993) *Science*, **260**, 1617-1623.
179. Gomberg, J. P., Bodin, P., and Reasenberg, P. A. (2003) *Bull. Seism. Soc. Am.*, **93**, 118-138.
180. Felzer, K. R. and Brodsky, E. E. (2006) *Nature*, **441**, 735-738.
181. Ben-Zion, Y. and Lyakhovsky, V. (2006) *Geophys. J. Int.*, **165**, 197-210.
182. Schwartz, S. Y. and Rokosky, J. M. (2007) *Rev. Geophys.*, **45**, 10.1029/2006RG000208.
183. Chlieh, M., Avouac, J.-P., Sieh, K., Natawidjaja, D. H., and Galetzka, J. (2008) *J. Geophys. Res.*, **113**.
184. Fialko, Y., Sandwell, D., Simons, M., and Rosen, P. (2005) *Nature*, **435**, 295-299.
185. Ellsworth, W. L., Celebi, M., Evans, J. R., Jensen, E. G., Kayen, R., Metz, M. C., Nyman, D. J., Roddick, J. W., Spudich, P., and Stephens, C. D. (2004) *Earthquake Spectra*, **20**, 597-615.
186. Rosakis, A. J., Xia, K.-W., and Lykotrafitis, G., K. (2007) Dynamic shear rupture in frictional interfaces: Speeds, directionality and modes. Schubert, G. and Kanamori, H. (eds.), *Treatise in Geophysics*, pp. 153-192, Vol. 4 - Earthquake Seismology, Elsevier.
187. Heaton, T. H. (1990) *Phys. Earth Planet. Inter.*, **64**, 1-20.
188. Dunham, M., Favreau, P., and Carlson, J. (2003) *Science*, **299**, 1557-1559.
189. Marone, C., Scholz, C. H., and Bilham, R. (1991) *J. Geophys. Res.*, **96**, 8441-8452.
190. Shearer, P., Hauksson, E., and Lin, G. (2005) *Bull. Seism. Soc. Am.*, **95**, 904-915.
191. Liu, Y. and Lapusta, N. (2008) *J. Mech. Phys. Solids*, **56**, 25-50.

192. Liu, Y. (2009) *Three-dimensional elastodynamic modeling of frictional sliding with application to intersonic transition. Dissertation (Ph.D.).* California Institute of Technology.
193. Thomas, M., Lapusta, N., Noda, H., and Avouac, J.-P. (2010) *GSA Annual Meeting*.
194. Poliakov, A. N. B., Dmowska, R., and Rice, J. R. (2002) *J. Geophys. Res*., **107**, 10.1029/2001JB000572.
195. Bhat, H. S., Sammis, C. G., and Rosakis, A. J. (2011) *Pure Appl. Geophys*., **168**, 2181-2198.
196. Sagy, A., Brodsky, E. E., and Axen, G. J. (2007) *Geology*, **35**, 283-286.
197. Chester, F. M. and Chester, J. S. (1998) *Tectonophysics*, **295**, 199-221.
198. Jeppson, T. N., Bradbury, K. K., and Evans, J. P. (2010) *J. Geophys. Res*., **115**, 10.1029/2010JB007563.
199. Hillers, G., Carlson, J. M., and Archuleta, R. J. (2009) *Geophys. J. Int*., **178**, 1363-1383.
200. Bürgmann, R. and Dresen, G. (2008) *Ann. Rev. Earth Plan. Sc*., **36**, 531-567.
201. Shelly, D. R. (2010) *Nature*, **463**, 648-652.

Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

The Mechanics of Faulting: From Laboratory to Real Earthquakes, 2012: 209-236
ISBN: 978-81-308-0502-3 Editors: Andrea Bizzarri and Harsha S. Bhat

7. Significance of high velocity friction in dynamic rupture process

Hiroyuki Noda
Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology Yokohama, Kanagawa, Japan

Abstract. There are multiple lines of evidences discovered over the last couple of decades for the coseismic weakening of a fault which is much more drastic than what is predicted from the conventional friction laws verified at low slip rates. Such weakening undoubtedly affects the dynamic rupture propagation in various ways. For example, coseismic weakening considered in a framework of rate-weakening has been shown to play an important role in determining the manner of dynamic rupture propagation (crack-like versus pulse-like rupture) given a pre-stress level. Moreover in the sequence of earthquakes, the pre-stress is affected by the coseismic frictional resistance. In this article, some of the recent studies on the significance of high velocity friction shall be reviewed briefly.

1. Introduction

The frictional constitutive law of fault rocks is one of the main ingredients in considering dynamic rupture propagation during an earthquake, and thus actively being studied experimentally, theoretically, numerically,

Correspondence/Reprint request: Dr. Hiroyuki Noda, Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology Yokohama, Kanagawa, Japan. Email: hnoda@jamstec.go.jp

and by field observations. The fault behavior is called "frictional" if the shear resistance τ is roughly proportional to the effective normal stress σ_e:

$$\tau = f\sigma_e = f\ \ \sigma - p \tag{1}$$

where f is the friction coefficient, σ is the total compressional normal stress on the fault, and p is the pore pressure. There are so many studies on how τ varies during an earthquake, and one of the recent prominent experimental discoveries include ramarkable reduction of τ (weakening) at coseismic slip rates (> 0.1 m/s), sometimes referred to as "high slip rates" or "high velocity" [e.g., 1]. This article presents how such weakening affects the dynamic rupture propagation and its sequences.

Section 2 is devoted to describe the significance of the weakening at high slip rates in comparison with the typical observation of τ at lower slip rates. In order to say if the weakening is significance or not in terms of elastodynamics, one should compare the weakening rate, decrease in τ per unit increase in the slip rate V, of a fault with a characteristic weakening rate from elastodynamics. Section 3 describes such a comparison based on an assumption that the weakening is considered as a rate-weakening; there is a decreasing function $\tau(V)$. The effect of the significant rate-weakening is demonstrated by explaining some of the proposed criteria [2, 3] which determine the manner (crack-like vs. pulse-like) of rupture propagation [4].

The weakening at the coseismic slip rate affects not only coseismic, but also interseismic fault behaviors such as the level of the shear stress at which a fault operates. Section 4 presents a brief review of a recent study [5] which explicitly present this point by considering a coseismic increase in p due to frictional heating (thermal pressurization of pore fluid) in a simulation of sequence of earthquakes.

2. Evidences of weakening of a fault at coseismic slip rates

Byerlee [6] compiled the data of friction experiments for variety of rocks, and concluded that the peak value of the frictional resistance τ_{peak}, which is attained near the onset of sliding, is independent of the rock type with exception of clay minerals. At a compressional effective normal stress σ_e below 200 MPa, τ_{peak} is given by $0.85\sigma_e$ with significant scattering of the data points, and at a higher normal stress, τ_{peak} is given by 50MPa + $0.6\sigma_e$. Thus, the peak friction coefficient f_{peak} ranges from 0.6 to 0.85 independently of the rock type. This is so-called Byerlee's law. This notion agrees with field observations such as the state of stress measured at boreholes [e.g., 7].

An earthquake occurs as a result of acceleration of relative motion (slip) on an active fault and its propagation. Then, the dependency of the sliding frictional resistance τ or the friction coefficient f (= τ / σ_e) on the speed of the relative motion (slip rate) V or, more generally, the governing equation of τ is of great importance in considering the earthquake generation process. Dieterich [8, 9] investigated the dependency of f_{peak} on the time of the stationary contact and the evolution of f after a sudden change in V. f_{peak} and f respectively depend on the contact time and V rather modestly; an e-fold increase in the contact time or V causes a change in f by on the order of 0.01 or smaller. Those experiments are typically carried out below 1 mm/s. Therefore, Byerlee's law holds for the sliding frictional resistance τ as well as for τ_{peak} for a wide range of σ_e and V for variety of rocks. The value of f from 0.6 to 0.85 (typically 0.7) has been referred to as a standard value of the friction coefficient of the rock. Note that during a large earthquake, an active fault slips at a slip rate on the order of 0.1 to 10 m/s. The experimental dataset which bases Byerlee's law does not include experiments at such high slip rates.

Given a characteristic value of f (~0.7) and a long-term slip rate of an active fault, one can calculate the frictional heating and the distribution of the heat flux at the surface. Lachenbruch and Sass [10] compared such theoretical prediction with measured heat flux along the San Andreas fault, California, and concluded that there is no detectable evidence of frictional heating. This observation suggests that the rate of frictional heating is much smaller than what is expected from the Byerlee's law if σ_e at the depth is assumed to be derived from lithostatic and hydrostatic stress condition. This issue is sometimes called as "heat flow paradox", and the explanation for it has been one of the major tasks during last decades.

There are multiple lines of evidences for low frictional heating during coseismic slip. After 1999 Chi-Chi, Taiwan, Earthquake (Mw 7.6), for example, 2 direct measurements of temperature anomaly in bore holes which penetrate the Chelungpu fault, a source fault of Chi-Chi Earthquake, at about 300 m and 1100 m in depth revealed that f during coseismic slip is about 0.1 if σ_e is given by the lithostatic and hydrostatic condition [11, 12]. d'Alessio et al. [13] studied exhumed San Gabriel fault, California, and concluded that there is no evidence for frictional heating recorded as thermal healing of fission tracks, and estimated that f must be below 0.4 for an earthquake which produces 4 m of slip.

Ancient earthquakes accompanied by high slip rates and resulting frictional heating can be recorded geologically as pseudotachylyte, melted and quenched glassy rock [e.g., 14]. Similar fused rocks are sometimes found in large landslides, and Erismann et al. [15] explained the occurrence of such

rocks as a result of frictional melting by producing pumice by friction experiments at a slip rate about 10 m/s. They reported that the averaged friction coefficient decreases with increasing normal stress from about 1 to as low as 0.17. There are many following studies producing pseudotachlyte in high velocity friction experiments [1, 16-25] as well as recent theoretical and numerical studies trying to formulate the shear resistance during fault motion [26-28]. Most of them reported much smaller f than characteristic value of Byerlee's law especially at high normal stress, suggesting that melt lubrication is one of the strong candidates explaining the low frictional heating to some extent.

Pseudotachylyte is, however, rather rarely observed along exhumed faults which used to be located at a seismically active depth. Sibson [29] explained the scarcity of pseudotachylyte by hydro-thermal effects. Because the thermal expansivity of water is much larger than that of rocks, frictional heating causes an increase in the pore pressure in the shear zone and thus a decrease in σ_e and τ (see Eq. (1)) if the surrounding rock is impermeable enough so that it can confine the pressurized fluid efficiently. This mechanism is called thermal pressurization of pore fluid. It has been extensively studied mainly theoretically and numerically in terms of fault constitutive law, frictional instability, dynamic rupture propagation, and sequence of earthquakes [e.g., 3, 5, 29-46]. If a fault slips at a constant V forever with a finite value of f and if the onset of melting is neglected, then the shear resistance ultimately decays to zero with a length scale for slip depending on the hydrothermal properties and slip rate. Rempel and Rice [41] derived a condition for impossibility of onset of melting.

Tsutsumi and Shimamoto [1] conducted friction experiments for gabbro at V which ranges from about 5 cm/s to 1.3 m/s at σ_e = 1.5 MPa, and revealed that f dramatically decreases even without the generation of a melt layer and without any confinement of pore fluid (air); f is around 0.8 at 5 cm/s and 0.5 at 0.55 m/s. Later experimental studies have reconfirmed this observation for many different kinds of rocks in variety of experimental conditions (Figure 1, modified from Wibberley et al. [47]). The weakening is explained by different mechanisms for different rock types and different experimental conditions, such as localized temperature rise at the true area of contact (flash heating) [40, 48], macroscopic temperature rise and intrinsic temperature-weakening effect [49], formation of silica gel for SiO_2-rich rocks [50], thermal decomposition of minerals and associated increase in pore pressure and/or generation of weak material (for example for coal [51], calcite [52], siderite [53], kaolinite [54], and gypsum [55]), and so on. Di Toro et al. [56] showed that the heat-induced weakening mechanisms can be distinguished by plotting the steady-state friction as a function of frictional power density.

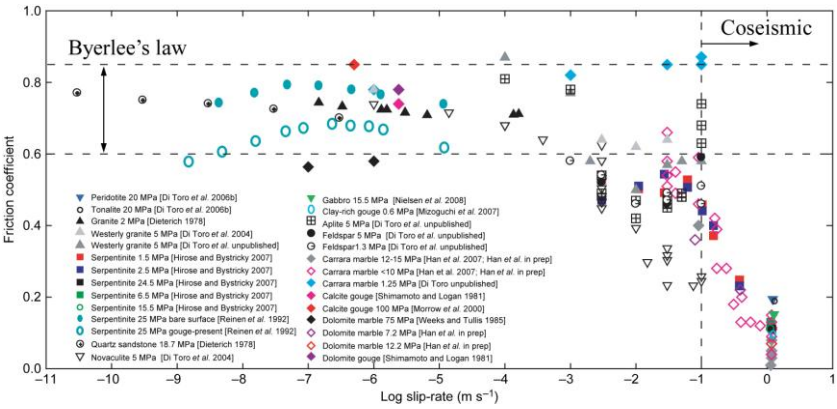


Figure 1. Experimentally observed friction coefficient f at low to high slip rates. Figure from Wibberley et al. [47].

Natural faults are not simple in terms of mineral composition, structure, and temperature-pressure and chemical conditions. Thus it is difficult to determine a physico-chemical mechanism which dominantly affects τ during dynamic rupture in a general case. But whatever the mechanism might be, it is likely that the shear resistance of a fault during coseismic slip is much smaller than what is predicted by Byerlee's law and an ambient effective normal stress based on lithostatic and hydrostatic stress state. With some exceptional materials which has low friction coefficient at low slip rates (e.g., graphite [58]), there is a large difference between the shear resistances at low (< 1 mm/s) and coseismic (> 0.1 m/s) slip rates. This difference is associated with a much more remarkable rate-weakening than what is observed at the low slip rates (see Figure 1) if the assumption that the shear resistance τ is a function of the slip rate V is appropriate.

It should be emphasized that at this point, the significance of the rate-weakening is defined in comparison with the conventional notion on the frictional resistance at low slip rates, and not relevant to the dynamic rupture process. In the next section, significance of the rate-weakening is discussed from the point of view of elastodynamics.

3. Significant rate-weakening in terms of elastodynamics

In the last section, significance of the rate-weakening at high slip rates is discussed in comparison with the frictional behavior at low slip rates (< 1 mm/s). In order to say that the absolute value of the slope of rate-

weakening [stress/velocity] is large or small in a context of dynamic rupture propagation, we need to consider nondimensionalization and normalization in a proper manner. In this section, a measure of rate-weakening is discussed in a dynamic rupture process.

Let us consider a rate-weakening frictional constitutive law of a fault,

$$\tau = \tau(V) \text{ for } V > 0 \text{ ; } d\tau / dV < 0 \text{ .} \tag{2}$$

Cochard and Magariaga [58] conducted numerical simulations of dynamic rupture propagation for a rate-weakening fault. Many following studies [e.g., 2, 3, 58-60] employed remarkable rate-weakening by embedding it into friction laws which have one or more internal variables (state variables) and regularize Eq. (2). In those cases, the friction law can be written as

$$\tau = \tau(V, \boldsymbol{\theta}) \tag{3}$$

where $\boldsymbol{\theta}$ is the state variables which is a vector in general and evolves to a steady-state value $\boldsymbol{\theta}_{ss}(V)$ if V is fixed. The rate-weakening then means

$$\tau_{ss}(V) = \tau(V, \boldsymbol{\theta}_{ss}(V)) \text{ ; } d\tau_{ss} / dV < 0 \text{ .} \tag{4}$$

Note that Rice et al. [61] proved that purely rate-weakening law can cause mathematical ill-posedness; there is no solution when Eq. (2) is coupled with elastodynamics if the rate-weakening is significant in a sense discussed in this section. Even though the regularization using a state variable is required in solving dynamic rupture problems, the notion of rate-weakening is still useful in discussing the results from those numerical studies as shown later.

$\tau_{ss}(V)$ is usually assumed to be rather simple, having a value consistent with Byerlee's law at V close to zero, much lower values at coseismic value of V, and a transition between them (see Figure 1) corresponding to a maximum value of $-d\tau_{ss}/dV$,

$$\chi \equiv \max(-d\tau_{ss} / dV) \text{ .} \tag{5}$$

For example, Rice [40] explained the experimentally observed weakening before the onset of melting by flash heating, and proposed a friction law,

$$\tau_{ss} = f_{ss}\sigma_e = \begin{cases} f_{LV}\sigma_e & V \le V_w \\ \left(\dfrac{V_w}{V} \left(f_{LV} - f_w \right) + f_w \right) \sigma_e & V \ge V_w \end{cases} \tag{6}$$

where f_{ss} is the steady-state friction coefficient, f_{LV} is the steady-state friction coefficient at low slip rates, V_w is the slip rate at which the weakening at high slip rate becomes efficient, and f_w is the friction coefficient at high enough slip rates (Figure 2). In this case, the maximum slope of the rate weakening is achieved at $V = V_w$:

$$\chi = \frac{f_{LV}(V_w) - f_w}{V_w}\sigma_e \ . \tag{7}$$

A question is how to evaluate χ in terms of elastodynamics.

The dynamic rupture process is often considered as a problem dealing with an interaction between a boundary (fault) on which a rupture propagates and surrounding elastic medium with inertial effects. χ is a quantity having a dimension of [stress/velocity], and there is a scale for it given by elastodynamics. Elastodynamics has 3 physical properties, shear modulus μ, Poisson's ratio ν, and density ρ. Then $(\mu\rho)^{1/2} = \rho c_s = \mu/c_s$ (acoustic impedance of s-wave) is a characteristic value of the medium having a dimension of [stress/velocity], where c_s is the s-wave speed which is equal to $(\mu/\rho)^{1/2}$. Therefore, it is reasonable to measure χ nondimensionalized by a quantity which is proportional to the impedance.

Let us consider a planer fault embedded in a linearly elastic full space which is homogeneous. The comparison of χ and the shear acoustic impedance is most visible in a boundary integral expression for the traction on the fault, $\boldsymbol{\tau}$ [e.g., 58, 62-65],

$$\boldsymbol{\tau}(\mathbf{r},t) = \boldsymbol{\tau}_0(\mathbf{r},t) + \boldsymbol{\varphi}[\mathbf{V};\mathbf{r},t] - \boldsymbol{\eta}\cdot\mathbf{V}(\mathbf{r},t) \ , \tag{8}$$

where $\mathbf{r}$ is the position vector which spans on the fault, t is time, $\boldsymbol{\tau}_0$ is the traction on the fault if there is no slip on the fault, $\boldsymbol{\varphi}$ is a convolution term which depends on the history of the jump in the particle velocity across the fault $\mathbf{V}$. If a unit normal vector to the fault is denoted as $\mathbf{n}$ and the other Cartesian basis vectors are as $\mathbf{p}_1$ and $\mathbf{p}_2$, then a second rank tensor $\boldsymbol{\eta}$ can be written as,

$$\boldsymbol{\eta} = \frac{\rho c_p}{2}\mathbf{nn} + \frac{\rho c_s}{2}\mathbf{p}_1\mathbf{p}_1 + \frac{\rho c_s}{2}\mathbf{p}_2\mathbf{p}_2 = \frac{\mu c_p}{2c_s^{\,2}}\mathbf{nn} + \frac{\mu}{2c_s}\mathbf{p}_1\mathbf{p}_1 + \frac{\mu}{2c_s}\mathbf{p}_2\mathbf{p}_2 \ , \tag{9}$$

where c_p is the p-wave speed and ρc_p is the acoustic impedance of p-wave. The third term in Eq. (8) is called as the radiation damping term [66]. In a 2-dimensional problem without allowing opening of a fault, we have only to consider one component of Eq. (8) in the shear direction,

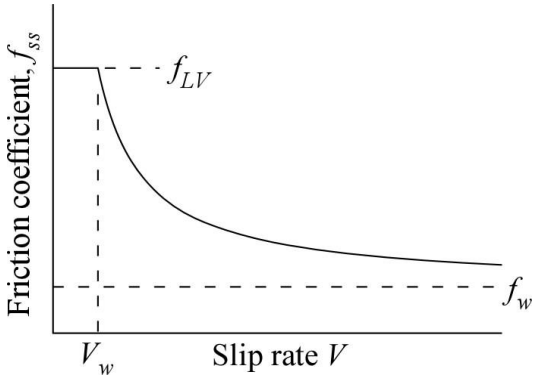


Figure 2. Friction coefficient f_{ss} as a function of slip rate V predicted by Eq. (6) which accounts for flash heating of microscopic asperities [40].

$$\tau(x,t) = \tau_0(x,t) + \varphi[V;x,t] - \eta V(x,t)\ ;\ \eta = \frac{\mu}{2c_s}\ , \tag{10}$$

where x is the spatial coordinate along the fault. η shall be chosen as the scale having the dimension of [stress/velocity]. The radiation damping term takes care of the instantaneous effect between V and τ. This term can be understood as a linear interpolation between stress-free and glued boundary conditions. Note that if the fault is a surface of constant stress (i.e., $\tau = \tau_0$ where τ_0 is constant with time) and the "stress wave" φ is propagates on the fault, then the propagating distribution of the slip rate V is proportional to φ as discussed by Brune [67].

In order to satisfy the boundary condition on the fault, Eqs. (2) or (3) and (10) must be satisfied simultaneously. For simplicity, we use the concept of pure rate-weakening in the discussion here assuming that the state evolution is rapid enough compared with the evolution of φ. The significance of χ/η is evident as indicated in Figure 3. Let us consider a loading history $\varphi(t)$ at a certain point on a fault such that the point on the fault is initially locked ($V = 0$), ruptured and slipped at coseismic slip rate, and decelerated and re-locked at the final condition. For illustration purpose, the initial and final value of φ us chosen to be equal to each other in Figure 3. As φ varies continuously with time t, $\chi/\eta < 1$ results in continuous change in V (Figure 3a) while $\chi/\eta < 1$ causes abrupt jumps in V (Figure 3b) as discussed by Cochard and Madariaga [58]. When a point on the fault stop sliding as φ decreases, $\chi/\eta < 1$ causes an efficient brake which is turned on at $V = V_{pulse}$,

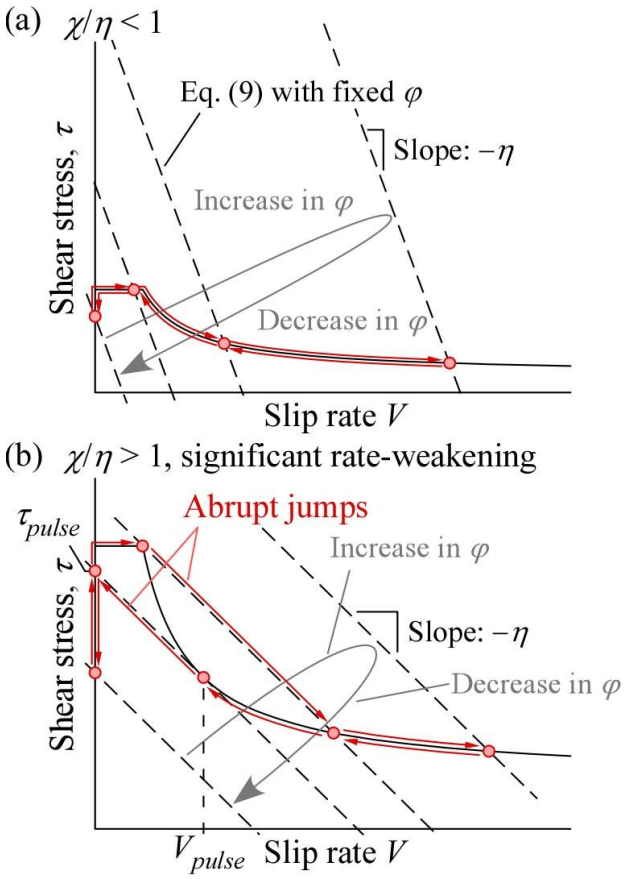


Figure 3. Trajectories of (V, τ) at a point on a fault which is governed by a purely rate-weakening friction law. The convolutional term φ increases to a peak value and decreases to the initial value for simplicity. (a) The case with $\chi/\eta < 1$; the rate-weakening is not significant in terms of elastodynamics. (b) The case with $\chi/\eta > 1$; the rate-weakening is significant in terms of elastodynamics. Significant rate-weakening defined by $\chi/\eta > 1$ causes efficient break at the deceleration at $V = V_{pulse}$.

the slip rate at which $\tau_{ss}(V)$ and the line which has a slope of $-\eta$ fit tangentially as shown in Figure 3b. With a rate- and state-dependent friction law, at this point, the trajectory of (V, τ) diverges from $\tau_{ss}(V)$, and V starts decreasing towards zero at τ much smaller than $\tau_{ss}(V)$ unless a loading is applied by incoming wave to this point (i.e., an increase in φ) [e.g., 3]. This

behavior causes a qualitative difference in the manner of dynamic rupture propagation.

3.1. Self-healing pulse-like rupture due to significant rate-weakening

Heaton [4] have revealed that many earthquakes occur in a self-healing pulse-like manner rather than a crack-like manner; slip on a fault propagates as a localized pulse of slip rate which has a rupture front and a healing front (Figure 4). The efficient brake due to the significant rate-weakening in terms of elastodynamics, $\chi/\eta > 1$, plays an important role in determining whether a rupture propagates in crack-like or a pulse-like manners [2,3,58-60, 68-70]. Note that there are other mechanisms proposed which cause the generation of the pulse-like rupture propagation such as a contrast in material properties across the fault [e.g., 71], and arrest waves from fault edges or from heterogeneity along the fault [e.g., 72-75], but those mechanisms will not be discussed here. In following subsections, some of the proposed criteria determining the manner of rupture propagation shall be presented which help understanding of the transition between crack-like and pulse-like ruptures and significance of the coseismic weakening.

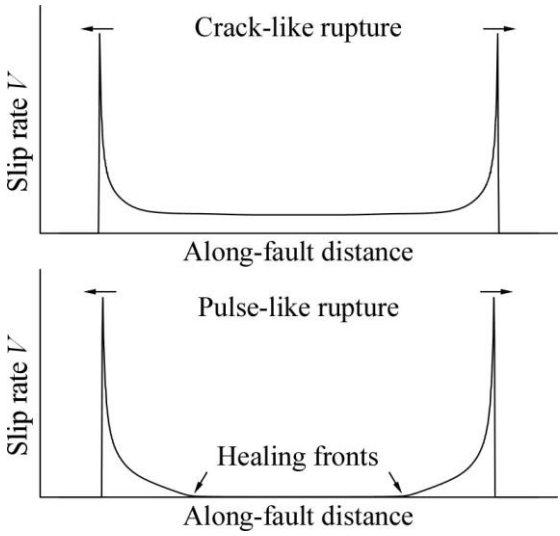


Figure 4. A schematic diagram showing a crack-like (upper) a pulse-like (lower) ruptures.

3.1.1. Impossibility of crack-like rupture at low background shear stress

Zheng and Rice [2] derived a sufficient condition for non-existence of crack-like ruptures. Please see their paper for detailed derivation. For a planer fault, the functional term φ in Eq. (10) satisfies

$$\int_{-\infty}^{\infty} \varphi(x,t)dx = 0 \cdot \tag{11}$$

Note that the rupture is a process of re-distribution of shear stress on a fault, and not a process of stress drop if one considers the entire system such as an infinitely long fault in a two-dimensional problem. Zheng and Rice [2] mathematically proved the following. Suppose there is an expanding crack-like rupture $S_{rupt}(t)$ which has a symmetry around $x = 0$. Then the shear stress is concentrated outside it $S_{out}(t)$,

$$\int_{S_{out}(t)} \varphi(x,t)dx \geq 0 \cdot \tag{12}$$

Eqs. (11) and (12) immediately yield

$$\int_{S_{in}(t)} \varphi(x,t)dx < 0 \cdot \tag{13}$$

The spatial integral of Eq. (10) inside $S_{rupt}(t)$ is then expressed as

$$\int_{S_{rupt}(t)} \tau - (\tau_0 - \eta V)dx = \int_{S_{rupt}(t)} \varphi dx < 0 \cdot \tag{14}$$

Assuming that $\tau = \tau_{ss}(V)$, if

$$\tau_{ss}(V) - (\tau_b - \eta V) \geq 0 \text{ fot all } V > 0, \tag{15}$$

where τ_b is background shear stress, then it is impossible for Eq. (14) to be satisfied. Therefore, a crack-like rupture can not exist at such a low τ_b. Note that τ_b is the spatial average of τ_0 inside S_{rupt}, and approaches to the spatial average of τ_0 over the entire fault as the hypothetical crack-like rupture expands.

This theorem rigorously gives a sufficient condition for the non-existence of the crack-like rupture. Let us define a critical shear stress value, τ_{pulse},

which is defined by the intersection of $V = 0$ and a line having the slope of $-\eta$ which tangentially fit to $\tau_{ss}(V)$ (Figure 3b). Then Eq. (15) is equivalent to

$$\tau_b < \tau_{pulse} . \tag{16}$$

If the background shear stress τ_b is smaller than τ_{pulse}, then an expanding crack-like rupture is impossible.

If a rupture is initiated by a compact over-stressed region which has high τ_0, the ruptures is crack-like just after its initiation, and it may undergo a transition to a pulse-like rupture or be arrested. Those transitions can understood through this theorem. τ_b, a spatial average of τ_0 over S_{rupt}, will decrease with S_{rupt} expands. Even if τ_b is larger than τ_{pulse} while S_{rupt} is small and thus a crack-like rupture is possible while the rupture is small, τ_b can become smaller than τ_{pulse} for large enough hypothetical ruptured area. In this case, the rupture must become pulse-like or arrested before it expands to such a size.

Zheng and Rice [2] conducted numerical simulations with a rate- and state-dependent friction law and verified that this theorem works even if there are process zones within which the purely rate-weakening law is not a good approximation. Also, they showed that pulse-like ruptures exist at a τ_b larger than, but close enough to τ_{pulse}. Similar results have been obtained by following studies [e.g., 3, 70, 71].

3.1.2. Condition in terms of dynamic stress drop

Noda et al., [3] have conducted dynamic rupture simulations accounting for flash heating (Eq. (6)) for two-dimensional anti-plane cases, and explained the critical value of the background shear stress τ_b between pulse-like and crack-like ruptures by considering a first-order approximation of the friction law in a necessary and sufficient manner.

The idealization of the friction law is the following. The shear resistance during coseismic sliding, $V > V_{pulse}$, is not significantly dependent on V. Therefore, the frictional resistance at this slip rate regime could be regarded as constant at τ_d (Figure 5a). On the other hand, the value of τ_{ss} in the slip rate regime where the rate-weakening is significant does not matter much. In fact, in the purely rate-weakening case, τ_{ss} at slip rates where $|d\tau_{ss}/dV| > \eta$ does not affect the solution at all (Figure 3b). There is a slip rate (V_{pulse}) at which the efficient braking on the fault slip takes place (Figure 5a) between those two regimes. In this idealized friction law, $0 < V < V_{pulse}$ is a forbidden range for the slip rate.

Let us hypothetically consider a singular anti-plane crack expanding self-similar mannerly. The shear stress is initially uniform at τ_b, and a ruptured

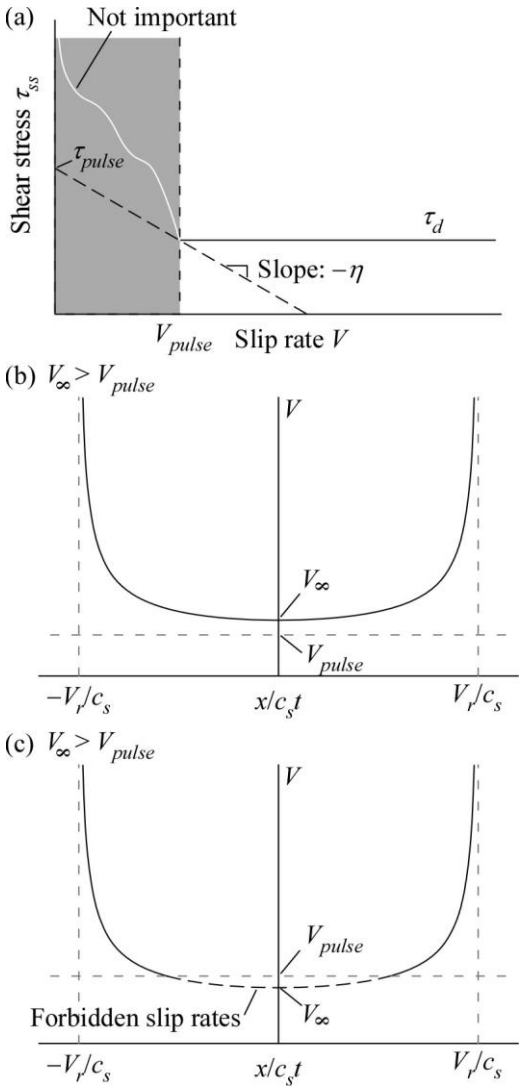


Figure 5. (a) Idealized friction law used in discussing the criterion between crack-like and pulse-like ruptures. At high enough slip rate $V > V_{pulse}$, the rate-weakening is not significant and thus the shear stress is idealized as a constant. (b) A possible self-similar crack-like rupture. $V_\infty > V_{pulse}$. (c) A self-similar crack-like rupture which is impossible because of the forbidden slip rate range, $0 < V < V_{pulse}$.

region propagates bilaterally by a constant rupture speed V_r in which the shear stress is uniformly equal to the kinetic frictional resistance, $\tau_d = \tau_b + \Delta\tau$. The corresponding slip rate distribution is [76, 77]

$$V = \mathrm{Re}\left[F\left(\frac{V_r}{c_s}\right)\frac{-\Delta\tau}{\mu}\frac{V_r}{\sqrt{1-(x/V_r t)^2}} \right], \tag{17}$$

where $F(V_r/c_s)$ is a nondimensional factor on the order of unity for realistic V_r/c_s (Figure 5b). This solution is valid if the range of V does not intersect with the forbidden slip rate range (0, V_{pulse}). At a point on the fault except $x = 0$, V is zero before the rupture reaches there, infinitely large at the rupture front, and decrease towards a certain value V_∞ which is given by

$$V_\infty = \lim_{t\to\infty} V = F\left(\frac{V_r}{c_s}\right)\frac{-\Delta\tau}{\mu}V_r \cdot \tag{18}$$

If the fault is governed by the idealized friction law (Figure 5a), this hypothetical crack-like rupture is possible if and only if $V_{pulse}, \leq V_\infty$ (Figure 5b and c). Noda et al. [3] discussed their numerical results along this line after a rough estimate of $F(V_r/c_s)\cdot V_r/c_s$ and $\Delta\tau$.

Let us consider the comparison with the condition by Zheng and Rice [2] in the idealized friction law discussed in this section (Figure 5a). The threshold of τ_b for impossibility of the crack-like rupture, τ_{pulse}, is given by

$$\tau_{pulse} = \tau_d + \eta V_{pulse} \ , \tag{19}$$

where τ_d is the shear resistance at $V > V_{pulse}$. The critical value of τ_b, τ_{bc} obtained from $V_{pulse} = V_\infty$ is given by

$$\tau_{bc} = \tau_d + \eta V_{pulse} \times 2\left(F\left(\frac{V_r}{c_s}\right)\frac{V_r}{c_s} \right)^{-1} \equiv \tau_d + \eta V_{pulse}\Xi\left(\frac{V_r}{c_s}\right) \cdot \tag{20}$$

The critical value of the dynamic stress drop $-\Delta\tau$ is given by $\eta V_{pulse}\Xi(V_r/c_s)$. Figure 6 shows $\Xi(V_r/c_s) > 1$, indicating $\tau_{bc} > \tau_{pulse}$. Therefore, the criterion by Noda et al. [3], which is necessary and sufficient, is consistent with and more restrictive than that by Zheng and Rice [2] in this specific problem although the former requires idealization in order to be applied to a general case which causes ambiguity.

It should be emphasized that the critical value of $-\Delta\tau$, $\eta V_{pulse}\Xi(V_r/c_s)$ could be constrained from physical properties of the fault and surrounding

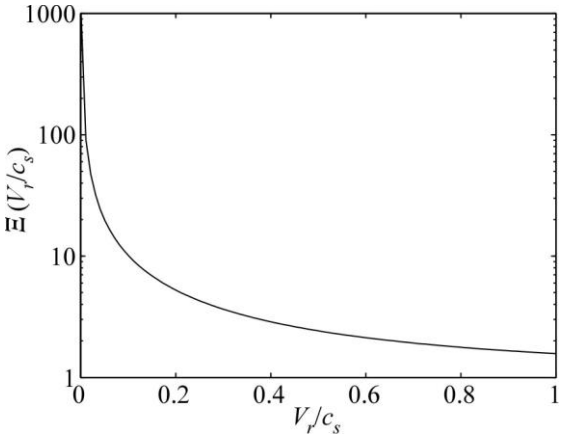


Figure 6. Plot of a function Ξ (Eq. 20) as a function of normalized rupture speed V_r/c_s.

medium, based on an assumption on the rupture speed. In the case numerically studied by Noda et al. [3], V_{pulse} = 1.5 m/s (at σ_e = 126 MPa corresponding to 7 km depth) and $\eta = \mu/2c_s$ = 5 MPa(m/s)$^{-1}$. For a rupture which propagates at a speed about $V_r/c_s \sim 0.8$, $\Xi(V_r/c_s)$ is about 2 (Figure 6) and then the critical value of $-\Delta\tau$ becomes about 15 MPa. This estimation explains their numerical results very well (see appendix C in their paper).

4. Effect of coseismic weakening on the long-tem fault behavior

In the previous section, the effect of coseismic weakening on dynamic rupture is discussed with the background shear stress as one of the parameters. However, the background shear stress or a pre-stress before each dynamic event is determined through the history of the fault behavior which includes sequence of earthquakes and interseismic fault motion. Undoubtedly, the coseismic weakening affects the shear stress right after an earthquake, during the following interseismic period, and the pre-stress before the next dynamic event. Therefore, the drastic coseismic weakening probably affects the characteristics of the long-term fault behavior such as the overall stress level at which the active fault operates. This section presents a review of a recent study by Noda and Lapusta [5] which investigated the effect of coseismic increase in the pore pressure p due to frictional heating (thermal pressurization of pore fluid) in a sequence of earthquakes.

4.1. Hydrothermal effects on frictional resistance

In Section 3, the weakening at coseismic slip rates has been discussed in a framework of rate-weakening, a decreasing function $\tau(V)$ or $\tau_{ss}(V)$. There are many possible fault constitutive laws which do not have those straightforward approximations by functions of V but produce weakening at coseismic slip rates. Among them, thermal pressurization of pore fluid is discussed in this section. This mechanism is one of the best studied weakening mechanisms [e.g., 3, 5, 29-46].

Figure 7 is a schematic diagram of thermal pressurization of pore fluid. Rapid sliding of a fault during an earthquake produces frictional heating which causes an increase in the temperature T by the fault. If the surrounding fault rock which is porous material is saturated with water, both rock and water try to expand thermally. The thermal expansion coefficient of water is much larger than that of rocks [e.g., 40, 41]. If the rock is not permeable enough, the pressurized water is confined near the shear zone and fluid pressure p increases locally, causing a decrease in the effective normal stress σ_e and thus dynamic weakening during an earthquake (see Eq. (1)).

Thermal pressurization of pore fluid is often modeled by considering diffusion of T and p normal to the fault with source terms corresponding to frictional heating [e.g., 30],

$$\frac{\partial T}{\partial t} = \alpha_{th} \frac{\partial^2 T}{\partial y^2} + \frac{\omega}{\rho c} \ , \tag{21}$$

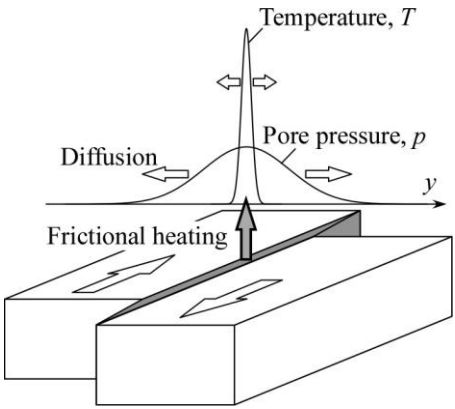


Figure 7. Schematic diagram showing the process of thermal pressurization of pore fluid. This figure is from Noda and Lapusta [5].

$$\frac{\partial p}{\partial t} = \alpha_{hy} \frac{\partial^2 p}{\partial y^2} + \Lambda \frac{\partial T}{\partial t} \ , \tag{22}$$

where y is the spatial coordinate normal to the fault, α_{th} and α_{hy} are thermal and hydraulic diffusivities, respectively, ρc is the specific heat capacity of the medium, Λ is the increase in p per unit increase in T under undrained condition, and ω is distribution of frictional heat generation which satisfies

$$\int_{-\infty}^{\infty} \omega dy = \tau V = (\sigma - p|_{y=0}) fV \tag{23}$$

if all of the dissipated energy by frictional resistance turns into heat. The hydraulic diffusivity α_{hy} is given by $k/\eta_w\beta$ where k is permeability of the rock, η_w is water viscosity, and β is storage capacity. Note that the nonlinear terms such as convection and heat generation due to pressurizing the fluid are neglected here. For the full formulation, see Mase and Smith [32]. For materials with low enough permeability ($k < 10^{-16}$ m^2), the convective term is not important [30, 32, 33]. Vredevoogd et al. [78] conducted numerical simulations that included all terms in the conservation equations [32], and demonstrated that the nonlinear terms may be safely neglected.

Among the physical properties which appear in Eqs. (21) and (22), the hydraulic diffusivity α_{hy} has the largest variation for orders of magnitude, depending on the rock types, effective mean stress, chemistry of pore water, and so on [e.g., 36, 79-88]. Thermal pressurization is efficient for mature faults with well-developed fault core which has low k and thus α_{hy}. Figure 8 (Figure 11 in Wibberley and Shimamoto [85]) shows an example of the internal and permeability structures of mature faults (Median Tectonic Line, Southwest Japan, as an example studied by Wibberley and Shimamoto [85]). Fracturing of an intact rock (e.g., mylonites and metapelitic schist in Figure 8) remarkably increases permeability k and thus α_{hy} locally. As cataclastic deformation accumulates towards the fault core, the fault rock becomes granulated (fault gouge), and then grain size reduction due to further cataclastic deformation causes a decrease in the permeability near the central slipping plane in Figure 8 which is also called as “principal slip surface” [40]. Mature faults which host large earthquakes have material with low permeability by the principal slip surface where coseismic shear strain is localized.

The width of the distribution of shear strain rate w plays an important role in thermal pressurization. The amplitude of heat generation density is

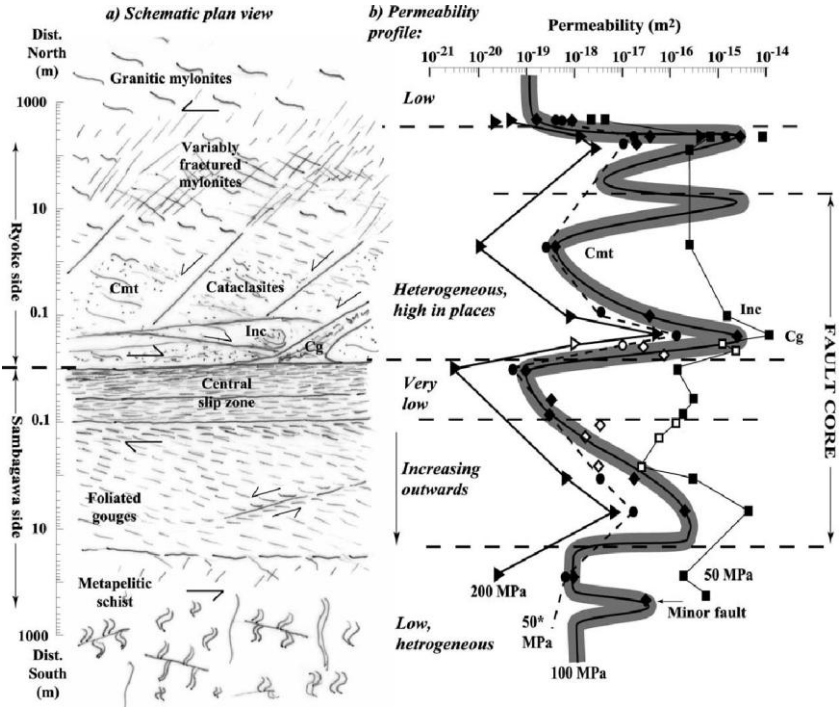


Figure 8. An example of (a) internal and (b) permeability structure in the mature fault zone (Median Tectonic Line, Southwest Japan as an example). This figure is Figure 11 in Wibberley and Shimamoto [85]. Permeability was measured with nitrogen as a pore fluid at different confining pressure, 50 MPa, 100 MPa, and 200 MPa, with 20 MPa of pore pressure.

inversely proportional to w, and the system can be approximated by adiabatic and/or undrained limits if the diffusion lengths of T and/or p are much smaller than w. Observation of the exhumed fault and drilled fault core [89-92], and samples after rotary shear friction experiments [92, 93] suggest that the shear deformation localizes within a very thin layer typically sub-millimeters thick which are often recognized as the layer of preferred orientation of platy minerals.

Rice [40] derived an analytic solution to Eqs. (21) and (22) at fixed V and f for a shear zone localized to a mathematical plane,

$$\omega = (\sigma - p) f V \delta_D(y) \tag{24}$$

where δ_D is Dirac's delta function. The analytic solution for the frictional resistance is written as

$$\tau = (\sigma - p_0) f \exp\left(\frac{\delta}{L_*}\right) \mathrm{erfc}\left(\sqrt{\frac{\delta}{L_*}}\right), \tag{25}$$

where σ is the total normal stress, p_0 is the initial pore pressure, δ is slip on the fault, and L_* is a length scale given by

$$L_* = \frac{4}{f^2}\left(\frac{\rho c}{\Lambda}\right)^2 \frac{\left(\sqrt{\alpha_{hy}} + \sqrt{\alpha_{th}}\right)^2}{V}. \tag{26}$$

For a range of realistic set of parameters, Rice [40] estimated L_* as 2 to 50 mm at coseismic slip rates. Figure 9 shows the evolution of τ predicted by Eq. (25) (modified from Figure 3 in Rice [40]). It should be emphasized that the apparent length scale of the displacement required for weakening of the fault depends on the final slip of the event. Rice [40] successfully explained the dependency of the seismic fracture energy, the area below τ as a function of slip and above the final value of τ, on the size of the earthquake [94] using those stress-reduction curves in Figure 9. The frictional resistance keeps decreasing towards zero as the fault slips at a constant V. Although it is true that the weakening due to thermal pressurization is more efficient at higher V (i.e., L_* decreases with V), the concept of rate-weakening becomes ambiguous.

Andrews [34] conducted dynamic rupture simulation with slip-weakening friction law for f and thermal pressurization which produced crack-like ruptures, and pointed out that efficient thermal pressurization produces nearly complete stress drop.

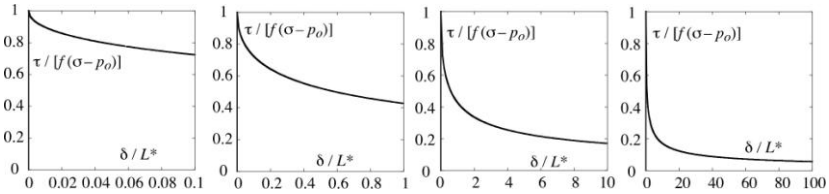


Figure 9. Figure 3 in Rice [40]. Analytic solution to thermal pressurization with a deformation localized on a mathematical plane, Eq (25), is plotted with different horizontal scales. Apparently, the slip required for the weakening of a fault is always a good fraction of the total slip.

4.2. Effect of coseismic weakening on the sequence of earthquakes

Noda and Lapusta [5] developed a suitable methodology to implement Eq. (21) and (22) in a calculation of sequence of earthquakes which accounts for full inertial effect in coseismic periods and long-term tectonic loading [64, 65], and examined the effect of coseismic weakening. This subsection presents a brief review of their study. For detailed methodology and results, please refer to their paper.

They considered a fault which is governed by a rate- and state-dependent law,

$$f = a\sinh^{-1}\left[\frac{V}{2V_0}\exp\left(\frac{f_0 + b\ln(V_0\theta/L)}{a}\right)\right] \approx f_0 + a\ln(V/V_0) + b\ln(V_0\theta/L)\cdot \quad (27)$$

The $\sinh^{-1}$ regularization is important only if the shear stress on the fault becomes so small that f is close to or smaller than a. For the physical basis of this regularization, see Nakatani [95] and Rice et al. [61]. f_0 is the steady-state friction coefficient at a reference slip rate $V = V_0$, a and b are nondimensional constants representing the magnitude of the direct and evolutional effect, and L is the characteristic slip displacement of the state evolution,

$$\frac{d\theta}{dt} = 1 - \frac{V\theta}{L} = \frac{V}{L}\ \theta_{ss}(V) - \theta \quad ; \quad \theta_{ss}(V) = \frac{L}{V}\cdot \quad (28)$$

The geometry of the fault studied by Noda and Lapusta [5] is shown in Figure 10. 2 square patches of potentially different hydraulic diffusivity α_{hy} are placed in a rate-weakening (a = 0.01, b = 0.014) region embedded in a rate-strengthening (a = 0.01, b = 0) region. The fault is loaded by a creeping at a constant rate V_{pl} near the periodic boundaries. α_{hy} can be heterogeneous on natural faults depending on the local lithology as shown by experimental measurements by Tanikawa and Shimamoto [88] for the Chelungpu fault, Taiwan, a source fault of 1999 Chi-Chi earthquake. Because of restricted computational resources, Noda and Lapusta [5] used rather thick (w = 1 cm) shear zone to make the rupture front numerically resolvable.

Distributions of cumulative slip and shear stress along the mid-depth (z = 0 in Figure 10) is shown in Figures 11 and 12, respectively. Figure 11a, 11b, 12a, 12b represent cases with same α_{hy} in both of the patches and the region between them. In these uniform cases, the earthquakes span the whole seismogenic region. This is partly because the nucleation size is not very small compared with the size of the seismogenic zone. The size of the nucleation can

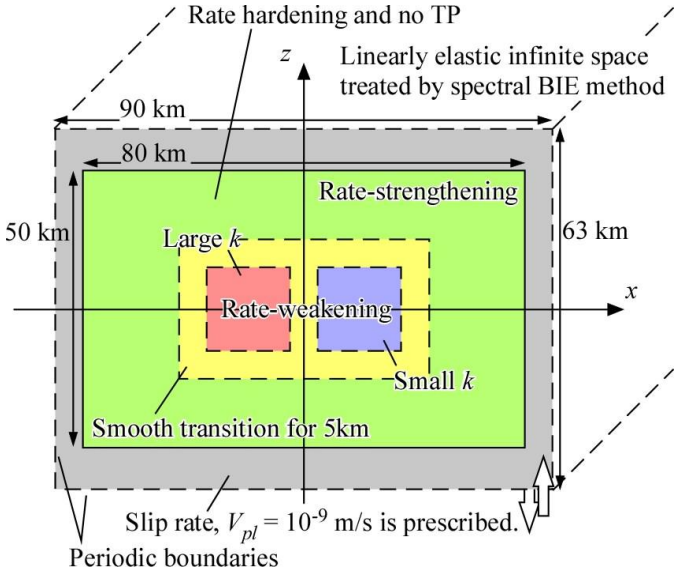


Figure 10. A geometrical setting of the fault studied by Noda and Lapusta [5]. This figure is modified from Figure 3 in Noda and Lapusta [5].

be recognized by the length of penetration of a creeping region which has high shear stress in Figure 12. Lapusta and Rice [96] showed that small nucleation size causes occurrence of frequent small earthquakes near the transition between seismogenic and aseismic regions. If the earthquake cycle is simple having only one event which spans the model, more efficient thermal pressurization causes longer recurrence interval and larger slip for each event as indicated by a spring-slider-dashpot model [97].

If there is spatial heterogeneity in the efficiency of thermal pressurization, the sequence of the earthquake becomes complex. The region with efficient dynamic weakening slips a lot when it ruptures. Thus, such a region cannot rupture in every event, and the slip deficit in the other patch is filled by more frequent and smaller events (Figure 11c). The vertical orange streaks in Figure 12c around x = 5 to 10 km is the stress concentration in front of the arrested earthquakes in the middle of low pre-stress region which is a result of previous earthquakes.

What is striking in Figure 12 is that the interseismic shear stress distribution is determined by the distribution of coseismic shear resistance except in the region of nucleation and around arrested rupture fronts. High velocity friction plays an important role in dynamic rupture propagation and,

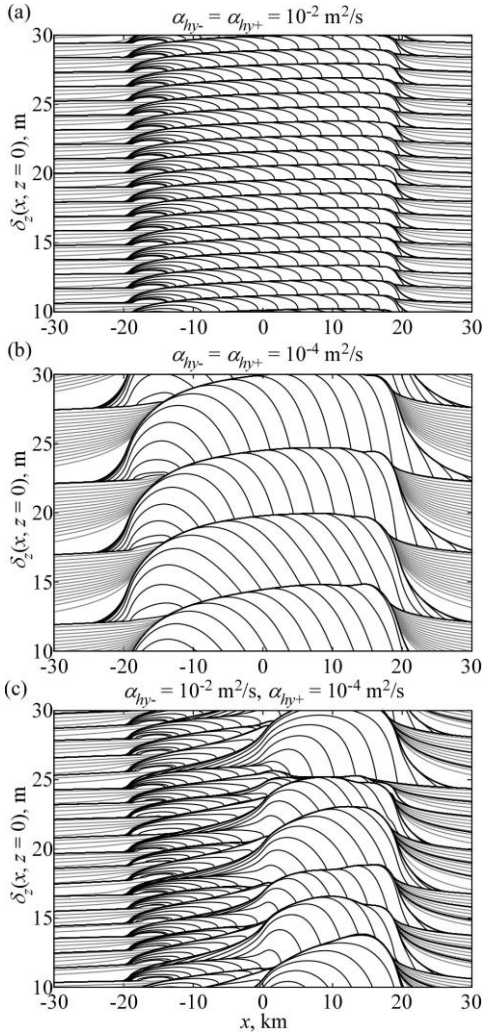


Figure 11. Cumulative slip distribution at the mid-depth (z = 0 in Figure 10). Gray lines are plotted every 10 years, showing interseismic slip accumulation mainly outside the seismogenic region. Black lines are plottded every 1 sec during earthquakes. α_{hy-} and α_{hy+} are hydraulic diffusivities in the left and right patches, respectively. Thermal pressurization is (a) not efficient in both of the patches, (b) efficient in both of the patches, and (c) efficient only in the right patch. This figure is modified from Figure 5 and 7b in Noda and Lapusta [5].

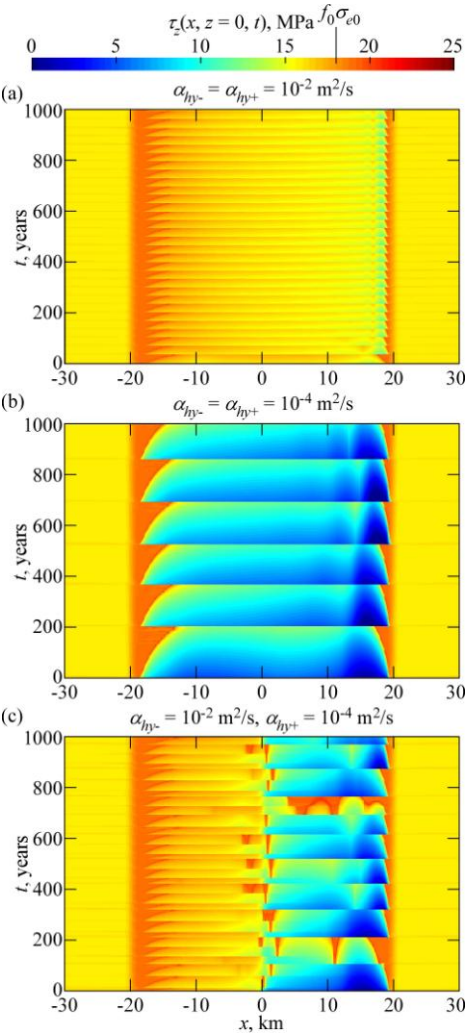


Figure 12. Spatio-temporal distribution of shear stress at the mid-depth (z = 0 in Figure 10). Ambient effective normal stress is 30 MPa and f_0 = 0.6 with $V_0 = 10^{-6}$ m/s so that yellow to orange color corresponds to the steady-state frictional resistance at low slip rates. Interseismic shear stress is controlled by the coseismic frictional resistance of the fault. Thermal pressurization is (a) not efficient in both of the patches, (b) efficient in both of the patches, and (c) efficient only in the right patch. This figure is modified from Figure 9 and 10b in Noda and Lapusta [5].

as its result, in determining the distribution of interseismic shear stress and thus pre-stress before the following dynamic events.

5. Summary

In this article, recent studies on the significance of high velocity friction on the dynamic rupture and its sequence are reviewed. There are multiple lines of evidences which support that the frictional resistance of a fault is much smaller than what is predicted by the Byerlee's law and lithostatic and hydrostatic stress condition. Especially, many experimental studies suggest a remarkable weakening of a fault at coseismic slip rates. In section 3, the meaning of significant rate-weakening is clarified in terms of elastodynamics, and reviewed some of the studies on the manner (crack-like versus pulse-like) of dynamic rupture propagation in which the significance of rate-weakening plays a central role. Coseismic weakening undoubtedly affects the shear stress distribution right after an earthquake, interseismic shear stress distribution, and then the pre-stress before the following events. In section 4, a study on the effect of coseismic weakening on the sequence of earthquakes is presented which employs thermal pressurization of pore fluid. Fault constitutive law at coseismic slip rates is important not only in considering the characteristics of individual earthquake event, but also in understanding the long-term fault behavior such as the long-term shear stress under which a fault operates.

References

1. Tsutsumi, A., and Shimamoto, T. 1997, Geophys. Res. Lett., 24(6), 699-702, doi:10.1029/97GL00503.
2. Zheng, G., and Rice, J. R. 1998, Bull. Seism. Soc. Am., 88, 1,466-1,483.
3. Noda H., Dunham, E. M., and Rice, J. R. 2009, J. Geophys. Res., 114, B07302, doi:10.1029/2008JB006143.
4. Heaton, T. H. 1990, Phys. Earth. Planet. Inter., 64, 1-20.
5. Noda H., and Lapusta, N 2010, J. Geophys. Res., 115, B12314, doi:10.1029/2010JB007780.
6. Byerlee, J. 1987, Pure Appl. Geophys., 116, 615-626.
7. Townend, J., and Zoback, M. D. 2000, Geology, 28, 5, 399-402.
8. Dieterich, J. H. 1972, J. Geiphys. Res., 77, 20, 3690-3697.
9. Dieterich, J. H. 1978, Pure. Appl. Geophys., 116, 790-806.
10. Lachenbruch, A. H. and Sass, J. H. 1980, J. Geophys. Res., 85, 6185-6223.
11. Kano, Y., Mori, J., Fujio, R., Ito, H., Yanagidani, T., Nakao, S., and Ma, K. F. 2006, Geophys. Res. Lett., 33, L14306, doi:10.1029/2006GL026733.

12. Tanaka, H., Chen, W. M., Wang, C. Y., Ma, K. F., Urata, N., Mori, J., and Ando, M. 2006, Geophys. Res. Lett., 33, L16316, doi:10.1029/2006GL026673.
13. d'Alessio, M. A., Blythe, A. E., and Bürgmann, R. 2003, Geology, 31(6), 541-544, doi:10.1130/0091-7613(2003)031<0541:NFHATS>2.0.CO;2.
14. Sibson, R. H. 1975, Geophys. J. R. Astr. Soc., 43, 775-794.
15. Erismann, T. L., Heuberger, H., and Preuss, E. 1977, Tschermaks Mineral. Prtrogr. Mitt., 24, 67-119.
16. Spray, J. G. 1987, J. Struct. Geol. 9, 1, 49-60.
17. Spray, J. G. 1992, Tectonophys., 204, 205-221.
18. Spray, J. G. 1993, J. Geophys. Res., 98(B5), 8053-8068.
19. Spray, J. G. 1995, Geology, 23, 12, 1119-1122.
20. Hirose, T., and Shimamoto, T. 2003, J. Struct. Geol., 25, 1569-1574.
21. Hirose, T., and Shimamoto, T. 2005, Bull. Seism. Soc. Am., 95(5), 1666-1673.
22. Hirose, T. and Shimamoto, T. 2005, J. Geophys. Res., 110(B5), B05202, 10.1029/2004JB003207.
23. Di Toro, G., Hirose, T., Nielsen, S., Pennacchioni, G., and Shimamoto, T. 2006, Science, 311, 647-649.
24. Di Toro, G., Hirose, T., Nielsen, S., and Shimamoto, T. 2006, Radiated Energy and the Physics of Earthquake Faulting, (Ed.) Abercrombie, R., McGarr, A., Kanamori, H., and Di Toro, G., AGU Geophysical Monograph, 170, 121-134, Washington, D. C.
25. Del Gaudio, P., Di Toro, G., Han, R., Hirose, T., Nielsen, S., Shimamoto, T., and Cavallo, A. 2009, J. Geophys. Res., 114, B06306, doi:10.1029/2008JB005990.
26. Sirono, S., Satomi, K., and Watanabe, S 2006, J. Geophys. Res., 111, B06309, doi:10.1029/2005JB003858.
27. Nielsen, S., Di Toro, G., Hirose, T., and Shimamoto, T. 2008, J. Geophys. Res., 113, B01308, doi:10.1029/2007JB005122.
28. Nielsen, S., Mosca, P., Giberti, G., Di Toro, G., Hirose, T., and Shimamoto, T. 2010, J. Geophys. Res., 115, B10301, doi:10.1029/2009JB007020.
29. Sibson, R. H. 1973, Nature, 243, 66-68.
30. Lachenbruch, A. H. 1980, J. Geophys. Res., 85(B11), 6097-6112, doi:10.1029/JB085iB11p06097.
31. Mase, C. W., and Smith, L. 1985, Pure Appl. Geophys., 122, 583-607.
32. Mase, C. W., and Smith, L. 1987, J. Geophys. Res., 92(B7), 6249-6272, doi:10.1029/JB092iB07p06249.
33. Lee, T. C., and Delaney, P. T. 1987, Geophys. J. Roy. Astr. Soc., 88(3), 569-591.
34. Andrews, D. J. 2002, J. Geophys. Res., 107(B12), 2363, doi:10.1029/2002JB001942.
35. Wibberley, C. A. J. 2002, Earth Plant. Space, 54(11), 1153- 1171.
36. Noda, H., and Shimamoto, T. 2005, Bull. Seismol. Soc. Am., 95(4), 1224-1233, doi:10.1785/0120040089.
37. Sulem, J., Vardoulakis, I., Ouffroukh, H., and Perdikatsis, V. 2005, Soils Found., 45(2), 97- 108.
38. Bizzarri, A., and Cocco, M. 2006, J. Geophys. Res., 111, B05303, doi:10.1029/2005JB003862.

39. Bizzarri, A., and Cocco, M. 2006, J. Geophys. Res., 111, B05304, doi:10.1029/2005JB003864.
40. Rice, J. R. 2006, J. Geophys. Res., 111(B5), B05311, doi:10.1029/2005JB004006.
41. Rempel, A. W., and Rice, J. R. 2006, J. Geophys. Res., 111, B09314, doi:10.1029/2006JB004314.
42. Segall, P., and Rice, J. R. 2006, J. Geophys. Res., 111, B09316, doi:10.1029/2005JB004129.
43. Suzuki, T., and Yamashita, T. 2006, J. Geophys. Res., 111, B03307, doi:10.1029/2005JB003810.
44. Suzuki, T., and Yamashita, T. 2008, J. Geophys. Res., 113, B07304, doi:10.1029/2008JB005581.
45. Suzuki, T., and Yamashita, T. 2009, J. Geophys. Res., 114, B00A04, doi:10.1029/2008JB006042.
46. Vredevoogd, M., Oglesby, D., and Park, S. 2007, Geophys. Res. Lett., 34, L18304, doi:10.1029/2007GL030754.
47. Wibberley, C. A. J., Yielding, G., and Di Toro, G. 2008, Geological Society, London, Special Publications 2008, 299, 5-33, doi:10.1144/SP299.2.
48. Beeler, N. M., Tullis, T. E., and Goldsby, D. L. 2008, J. Geophys. Res., 113, B01401, doi:10.1029/2007JB004988.
49. Noda, H. 2008, J. Geophys. Res., 113, B09302, doi:10.1029/2007JB005406.
50. Di Toro, G., Goldsby, D. L., and Tullis, T. E. 2004, Nature, 427, 436-439.
51. O'Hara, K., Mizoguchi, K., Shimamoto, T., and Hower, J. C. 2006, Techtonophys., 424, 109-118.
52. Han, R., Shimamoto, T., Hirose, T., Ree, J. H. and Ando, J. 2007, Science, 316, 878-881.
53. Han, R., Shimamoto, T., Ando, J., and Ree, J. H. 2007, Geology, 35, 12, 1131-1134, doi: 10.1130/G24106A.1.
54. Brantut N., Schubnel, A., Rouzaud, J.-N., Brunet, F., and Shimamoto, T. 2008, J. Geophys. Res., 113, B10401, doi:10.1029/2007JB005551.
55. Brantut N., Han, R., Shimamoto, T., Findling, N., and Schubnel, A. 2011, Geology, 39(1), p. 59-62, doi: 10.1130/G31424.1.
56. Di Toro, G., R. Han, T. Hirose, N. De Paola, S. Nielsen, K. Mizoguchi, F. Ferri, M. Cocco, and T. Shimamoto, Nature, 471, doi:10.1038/nature09838.
57. Oohashi, K., Hirose, T., and Shimamoto, T. 2011, J. Struct. Geol., 33, 6, 1122-1134, doi:10.1016/j.jsg.2011.01.007.
58. Cochard, A., and Madariaga, R. 1994, Pure Appl. Geophys., 142, 419-445.
59. Cochard, A., and Madariaga, R. 1996, J. Geophys. Res., 101, 25,321-25,336.
60. Ampuero, J.-P. and Ben-Zion, Y. 2008, Geophys. J. Int., 173, 674-692, doi: 10.1111/j.1365-246X.2008.03736.x.
61. Rice, J. R., Lapusta, N., and Ranjith, K. 2001, J. Mech. Phys. Solids, 49, 1865-1898.
62. Geubelle, P. H., and Rice, J. R. 1995, J. Mech. Phys. Solids, 43, 1,791-1,824.
63. Fukuyama, E. and Madariaga, R. 1998, Bull. Seismol. Soc. Am., 88, 1-17.

64. Lapusta, N., Rice, J. R., Ben-Zion, Y., and Zheng, G. 2000, J. Geophys. Res., 105(B10), 23,765-23,789, doi:10.1029/2000JB900250.
65. Lapusta, N. and Liu, Y. 2009, J. Geophys. Res., 114, B09303, doi:10.1029/2008JB005934.
66. Rice, J. R. 1993, J. Geophys. Res., 111, 9,885-9,907.
67. Brune, J. N. 1970, J. Geophys. Res., 75, 26, 4997-5009.
68. Perrin, G., Rice, J. R., and Zheng, G. 1995, J. Mech. Phys. Solids, 43, 1461-1495.
69. Beeler, N. M., and Tullis, T. E. 1996, Bull. Seism. Soc. Am., 86, 1130-1148.
70. Nielsen, S. B. and Carlson, J. M. 2000, Bull. Seismol. Soc. Am., 90(6), 1480-1497.
71. Andrews, D. J., and Ben-Zion, Y. 1997, J. Geophys. Res., 102, 553-571.
72. Day, S. M. 1982, Bull. Seism. Soc. Am., 72, 705-727.
73. Johnson, E. 1990, Geophys. J. Int., 101, 125-132.
74. Beroza, G. C., and Mikumo, T. 1996, J. Geophys. Res., 101, 22,449-22,460.
75. Bizzarri, A., Cocco, M., Andrews, D. J., and Boschi, E. 2001, Geophys. J. Int., 144, 656-678.
76. Kostrov, B. V. 1964, Applied Mathematics and Mechanics, 28, 1,077-1,087. English translation from PMM, 28 (1964), 644-652.
77. Nielsen, S., and Madariaga, R. 2003, Bull. Seism. Soc. Am., 93(6), 2375-2388.
78. Vredevoogd, M., Oglesby, D., and Park, S. 2007, SCEC Annual Meeting Proceedings and Abstracts, vol. XVII, p. 188, South. Calif. Earthquake Cent., Los Angeles.
79. Morrow, C.A., Shi, L. Q., and Byerlee, J. D. 1981, Geophys. Res. Lett. 8, 325-329.
80. Morrow, C.A., Shi, L. Q., and Byerlee, J. D. 1984, J. Geophys. Res., 89, 3193-3200.
81. Morrow, C.A., and Byerlee, J. D. 1988, Geophys. Res. Lett. 15, 1033-1036.
82. Morrow, C. A., and Byerlee, J. D. 1992, J. Geophys. Res., 97, 5145-5151.
83. Faulkner, D. R., and Rutter, E. H. 2000, J. Geophys. Res., 105(B7), 16, 415- 16,426.
84. Faulkner, D. R., and Rutter, E. H. 2003, J. Geophys. Res., 108(B5), 2227, doi:10.1029/2001JB001581.
85. Wibberley, C. A. J. and Shimamoto, T. 2003, J. Struct. Geol., 25, 59-78.
86. Tsutsumi, A., Nishino, S., Mizoguchi, K., Hirose, T., Uehara, S., Sato, K., Tanikawa, W., and Shimamoto, T. 2004, Tectonophys., 379, 93–108.
87. Mizoguchi, K., Hirose, T., Shimamoto, T., and Fukuyama, E. 2008, J. Struct. Geol., 30, 513-524, doi:10.1016/j.jsg.2007.12.002.
88. Tanikawa, W., and Shimamoto, T. 2009, J. Geophys. Res., 114, B01402, doi:10.1029/2008JB005750.
89. Chester, F. M., and Chester, J. S. 1998, Tectonophysics, 295, 199–221, doi:10.1016/S0040-1951(98)00121-8.
90. Chester, F. M., Chester, J. S., Kirschner, D. L., Schulz, S. E., and Evans, J. P. 2004, in Rheology and Deformation in the Lithosphere at Continental Margins, (Ed) G. D. Karner et al., pp. 223-260, Columbia Univ. Press, New York.

91. Heermance, R., Shipton, Z. K., and Evans, J. P. 2003, Bull. Seismol. Soc. Am., 93(3), 1034–1050, doi:10.1785/0120010230.
92. Mizoguchi, K., Hirose, T., Shimamoto, T., and Fukuyama, E. 2009, Tectonophysics, 471, 285-296, doi:10.1016/J.Tecto.2009.02.
93. Kitajima H., Chester, J. S., Chester, F. M., and Shimamoto, T. 2010, J. Geophys. Res., 115, B08408, doi:10.1029/2009JB007038
94. Abercrombie, R. E and Rice, J. R. 2005, J. Geophys. Int., 162, 406-424, doi:10.111/j.1365-246X.2005.02579.x.
95. Nakatani, M. 2001, J. Geophys. Res., 106, 13,347-13,380.
96. Lapusta, N. and Rice, J. R. 2003, J Geophys. Res., 108, B4, 2205, doi:10.1029/2001JB000793.
97. Mitsui, Y., and Hirahara, K. 2009, J. Geophys. Res., 114, B09304, doi:10.1029/2008JB006220.