Physics of Earthquakes and Buildings

Textbook by

Thomas Heaton

Emeritus Professor of

Engineering Seismology

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

[THOMAS H HEATON]
ORCID: [HTTPS://ORCID.ORG/0000-0003-3363-2197]

Introduction

This document is a combination of class notes for CE/Ge 181, which was a Caltech graduate course about Engineering Seismology, together with several unpublished research papers about the physics of earthquakes.

The science of seismology has its beginnings at the start of the twentieth century when the first seismographs were developed. The early days of seismology focused on understanding the physics of seismic phases that were observed in seismic data. In addition, seismologists sought to understand the characteristics of the seismicity of the Earth. Seismologists focused their efforts on operating seismographic networks that recorded motions that were too small to be felt. Seismologists used the measurements of seismic phases (times and amplitudes) to create models of the Earth's structure and to characterize the occurrence of earthquakes.

Earthquake Engineering is a separate discipline that focuses on designing structures that can survive earthquakes. Whereas seismologists are typically trained in Physics departments (the recent trend is for seismologists to be trained in Earth Science Departments), earthquake engineers are mostly trained in Civil Engineering departments. Earthquake engineering generally began in the 1930's, and the development of strong motion accelerographs resulted in observations of the nature of shaking in damaging earthquakes. Earthquake engineers were mostly interested in the amplitude of ground shaking, including the spectral composition.

At most academic institutions, scientists and engineers are segregated; they typically take different courses, they read different journals, and they are members of different professional societies. I have been studying earthquakes at Caltech since 1972. I received a BS in physics from Indiana University in 1972 after which, I became a graduate student at Caltech's Seismological Laboratory. Prof. Donald Helmberger was my PhD advisor, and I focused on understanding the physics of strong ground motion (Heaton, 1979). Helmberger was one of the pioneers of seismology who developed methodologies to use

computers to simulate the physics of seismic motions. The fact that my research used the theories of seismology to understand the ground motions that were central to earthquake engineering meant that my research career has been at the interface between seismology and earthquake engineering.

Following my PhD in 1978, I briefly worked for Dames and Moore, which was a firm that provided technical advice to companies that needed to design important facilities (e.g., power plants, refineries, ports) to survive natural catastrophes (especially earthquakes). My tenure at Dames and Moore provided a link between my earthquake physics research and practical applications. It also taught me that engineering design firms were necessarily focused on getting things built; there was little time to ponder questions about the fundamental physics of earthquakes.

I joined the USGS as a Research Geophysicist in 1979 and I was stationed at the Pasadena Field Office. Earthquake prediction was the primary task of the USGS Branch that hired me, and a detailed catalog of small earthquakes was the key tool used to search for spatio-temporal patterns that could help to recognize that a major earthquake was about to happen. The 250-station Southern California Seismic Network allowed us to create this seismic catalog. Analog telephone lines were the only practical way to continuously telemeter data from this network. Seismic data was continuously recorded on photographic films that were reviewed by analysts to search for the arrival of seismic phases. Unfortunately, only the arrival time of initial P-waves could be determined from this network; the records of ground motion were typically overdriven in earthquakes.

Following the lead of Don Helmberger, my own research focused on simulating the waveforms recorded by networks. A significant part of my work for the USGS focused on redesigning the Southern California Seismic Network so that it would record ground motion over a broad range of amplitudes and frequencies. Much of what I learned about seismic networks is summarized in Chapter 2 of these class notes.

Caltech hired me to be a Professor of Engineering Seismology in 1995. This meant that I was faculty both in Geophysics and also in Civil Engineering. I taught Engineering

Seismology, which was a graduate course that was co-listed in Geophysics and Civil Engineering. Since I had no formal engineering training, the course was especially focused on classical seismology, which is mostly comprised of deriving the small motions of the Earth with respect to its static equilibrium state. Taylor's Theorem indicates that the governing equations of motion for small perturbations from equilibrium are linear. Solutions of these linear equations can be summed to other linear solutions and they still satisfy the linear equations of motion. There are decades of research to discover solutions to linear wave problems. Chapters 3, 4, and 5 provide examples of a range of these problems.

Teaching classes in classical engineering mechanics was also one of my duties as a Prof. of Civil Engineering. This was challenging since I was not trained as an engineer. Although many of the governing equations are shared between engineering and seismology, there are also many important differences. For example, bending is a key structural engineering concept that is rarely encountered in geophysics. Importantly, engineering structural mechanics emphasizes concepts that allow the analysis of materials that yield inelastically. That is, these problems are typically nonlinear. I discovered that civil engineers and geophysicists used fundamentally different theories to describe the mechanical failure of materials. For example, residual stress is an important subject in engineering, but it is rarely encountered in geophysics. I have been amazed that earth scientists believe that brittle materials are strong and that ductile materials are weak. In contrast, earthquake engineers try hard to achieve designs that yield ductily; brittle behavior is discouraged.

I realized that the deformation of buildings is similar to the deformation of the Earth, but buildings are characterized by effective elastic constants that are very different from those appropriate for the Earth's Crust. Furthermore, seismologists almost always use inertial coordinate frames, whereas structural engineers typically use non inertial coordinates (motions are described relative to the base of a structure). While both coordinate frames can be used to derive identical solutions, the physical intuitions provided by inertial coordinates are often different. For example, ground acceleration is a key concept in engineering, but it does not have much significance in seismological analysis.

Chapter 6 describes the deformation of buildings. I describe classical techniques from structural dynamics (e.g., normal modes and response spectra) and I also show how the same problems can be solved using ray theory developed primarily by seismologists. Chapter 6 also explains why peak ground velocity and displacement are so important to understand building damage.

Chapter 7 focuses on the relationship between slip on faults and the resulting ground motion. There is a comprehensive derivation of equivalent source parameters. If you have ever been confused about why seismologists describe earthquakes with seismic moment tensors, then you should read Chapter 7. You will learn that earthquakes do not release moment and that seismic potency is a more natural parameter than seismic moment. Chapter 7 also describes radiation pattern and rupture directivity, which are critical to understand damaging ground shaking. There is also a description of static displacement changes that are induced by slip on finite fault planes.

Chapter 8 is about the characteristics of earthquake ruptures. I summarize the development of magnitude scales. I discuss the origin of energy magnitude, M_W, and the commonly used moment magnitude, $\mathscr{M}$. Unfortunately, seismic moment is not a physical characteristic of earthquakes, so I introduce potency magnitude, M, which is similar to moment magnitude, but its physics is correct, and it makes the math simpler. I then discuss the Gutenberg-Richter law that describes how the number of earthquakes in a catalog varies with the magnitude of events. I show that the Gutenberg-Richter relationship means that most plate boundary motions occur in the largest events.

Most of Chapter 8 is about the way that earthquake parameters scale with the size of an earthquake, where size is measured as either potency, P, or Energy. I show how total mechanical energy can be decomposed into radiated elastic energy, static strain energy, and inelastic energy. I describe self-similar models of earthquakes, which assume that earthquakes are brittle cracks with slip that is controlled by the stress drop. The most used form of these models is referred to as the Brune ω^{-2} model and I show that it has

fundamental errors. I show that Brune's solution is not the solution to the mechanics problem that he poses. If you use Brune's model, then you should read Chapter 8.

Chapter 8 also has an extensive discussion about the difference between slip-pulse solutions and brittle-crack solutions. I show that slip-pulses lead to chaotic dynamics that can be described with strange attractors, which are states in which the prestress evolves into spatially complex distributions that are described as fractals. That is, I discuss the importance of residual stress (a concept from material sciences) in earthquake physics. I discuss how these concepts may solve the stress paradox, which has been the most important problem in geophysics for at least the past fifty years. I argue that many popular concepts, such as self-similar earthquake scaling, the seismic cycle, and strength of the crust need to be reevaluated. Most importantly, current procedures have underestimated the amplitude of long-period motions in infrequent events. This could lead to catastrophic collapse of tall buildings in future earthquakes.

I wrote Chapter 8 after I stopped teaching, and it is too long and complex to use for a standard course. It consolidates many different studies that I worked on with my graduate students, and I attempt to bring these studies together into a coherent document. The concepts of Chapter 8 are not the current standard, but it is my best attempt to bring these ideas together into a different model of failure dynamics.

Chapter 1 Single-Degree-of-Freedom Linear Oscillator (SDOF)

For most dynamic systems the relationship between restoring force and deflection is approximately linear for small deviations about some equilibrium state, which occurs when the system is motionless (all static forces are balanced). If the system is complex (e.g., a building that requires numerous variables to describe its properties), it is possible to use normal modes to transform the multi-dimensional system into a number of simple 1-dimensional linear oscillator problems (SDOF). The SDOF problem is also fundamental to understanding the principles of seismometers.

Consider the most fundamental of seismometers shown in Figure (1.1). In this case the ground moves in an inertial frame with displacement $u(t)$; a mass m is supported by a linear spring of stiffness k; and there is a viscous damper that resists the relative velocity $\dot{x}$ of the mass with respect to the ground with force $-b\dot{x}$. $x(t)$ is sometimes measured directly using an optical transducer (e.g., a light beam deflected by a mirror on the mass).

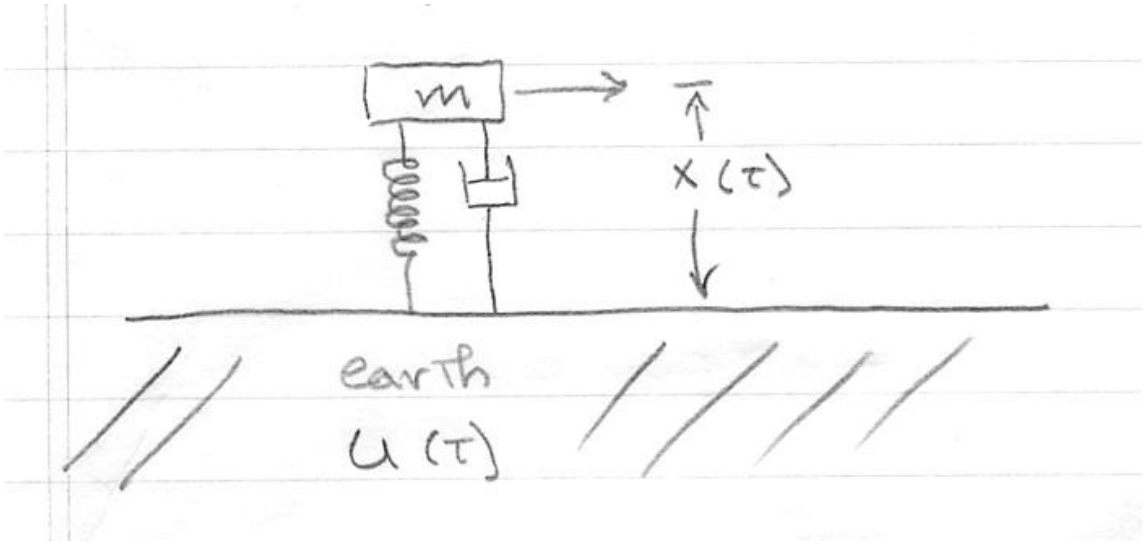


Figure 1.1. Sketch showing the configuration of a simple vertical-component seismometer.

the force on m is $-kx-b\dot{x}$ and the inertial force on m is $m(\ddot{x}+\ddot{u})$. The equation of motion of the system is then.

$$m(\ddot{x}+\ddot{u})+kx+b\dot{x}=0, \tag{1.1}$$

which can be rewritten as

$$\ddot{x}+2\beta\dot{x}+\omega_0^2 x=-\ddot{u}, \tag{1.2}$$

Where,

$$\omega_0 \equiv \sqrt{\frac{k}{m}} = \text{undamped natural frequency} \tag{1.3}$$

$$\beta \equiv \frac{b}{2m} = \text{ damping constant,} \tag{1.4}$$

which is related to the fraction of critical damping ς by

$$\beta = \omega_0 \varsigma. \tag{1.5}$$

Equation (1.2) is a 2nd order linear differential equation and its solution is widely known. In general, the solution is broken into two parts. The homogeneous solution, which solves.

$$\ddot{x}+2\beta\dot{x}+\omega_0^2 x=0 \tag{1.6}$$

Any solutions, $x_n(t)$, of the homogeneous equation (1.6) can be summed and they also solve the homogeneous equation since it is linear. The actual form of the solution to the homogeneous problem is determined by the initial conditions x_0 and $\dot{x}_0$. Solutions to the homogeneous equation can also be summed to solutions to the full inhomogeneous equation (1.2) and they will still solve the inhomogeneous equation. The homogeneous solutions typically represent the transient part of the response of the system. The homogeneous solution that matches the initial conditions is then added to the particular solution that solves equation (1.2). The particular solution often represents the steady-state part of problems.

Frequency Response

It is particularly useful to represent the ground motion $u(t)$ as a Fourier series, or

$$u(t)=\sum_{n=1}^{\infty}\left[A_n\cos(\omega_n t)+B_n\sin(\omega_n t)\right]=\sum_{n=1}^{\infty}C_n\cos(\omega_n t-\theta_n) \tag{1.7}$$

or

$$\ddot{u}(t) = -\omega_n{}^2 \sum_{n=1}^{\infty} \left[A_n \cos(\omega_n t) + B_n \sin(\omega_n t) \right] = -\omega_n{}^2 \sum_{n=1}^{\infty} C_n \cos(\omega_n t - \theta_n) \tag{1.8}$$

This representation is possible for functions that are periodic with a repeat time of T. In this case

$$\omega_n = \frac{2n\pi}{T} \tag{1.9}$$

Since our equation is linear, we can write the solution *x(t)* as the sum of solutions to individual harmonic problems, or

$$x(t) = \sum_{n=1}^{\infty} C_n x_n(t), \tag{1.10}$$

where x_n solves the equation

$$\ddot{x}_n + 2\beta \dot{x}_n + \omega_0^2 x_n = \omega_n^2 \cos(\omega_n t - \theta_n). \tag{1.11}$$

Since the cosine function is truly periodic, it has no beginning or end, there are no initial conditions or transient solutions to deal with. That is, the solution consists of just the particular solution. As it turns out, when a linear system is harmonically forced at one frequency, then the resulting motions (except for transients) are also harmonic at that frequency. Therefore, let us guess that the solution of (1.11) is

$$x_n(t) = D_n \cos(\omega_n t - \delta_n - \theta_n) \tag{1.12}$$

substituting (1.12) into(1.11), we find that

$$D_n \left(\omega_0^2 - \omega_n{}^2 \right) \cos(\omega_n t - \delta_n) - 2 D_n \beta \omega_n \sin(\omega_n t - \delta_n) = \omega_n{}^2 \cos(\omega_n t). \tag{1.13}$$

We then utilize the following trig identities

$$\cos(\omega_n t - \delta_n) = \cos(\omega_n t)\cos\delta_n + \sin(\omega_n t)\sin\delta_n \tag{1.14}$$

$$\sin(\omega_n t - \delta_n) = \sin(\omega_n t)\cos\delta_n - \cos(\omega_n t)\sin\delta_n \tag{1.15}$$

substituting (1.14) and (1.15) into (1.13), we find that

$$\begin{aligned} &\left\{ \omega_n{}^2 - D_n \left[\left(\omega_0^2 - \omega_n{}^2 \right) \cos\delta_n + 2\omega_n \beta \sin\delta_n \right] \right\} \cos(\omega_n t) \\ &-D_n \left[\left(\omega_0^2 - \omega_n{}^2 \right) \sin\delta_n - 2\omega_n \beta \cos\delta_n \right] \sin(\omega_n t) = 0 \end{aligned} \tag{1.16}$$

Now since $\sin(\omega_n t)$ and $\cos(\omega_n t)$ are linearly independent functions, each term in equation (1.16) must be linearly independent. Setting the second term to zero

$$\left(\omega_0^2 - \omega_n{}^2\right)\sin\delta_n - 2\omega_n\beta\cos\delta_n = 0 \tag{1.17}$$

or

$$\tan\delta_n = \frac{2\omega_n\beta}{\omega_0^2 - \omega_n{}^2} \tag{1.18}$$

Now setting the first term in (1.16) to 0 gives

$$D_n = \frac{\omega_n{}^2}{\left(\omega_0^2 - \omega_n{}^2\right)\cos\delta_n + 2\omega_n\beta\sin\delta_n} \tag{1.19}$$

If we make the clever observation that equation (1.18) can be rewritten as the following two equations

$$\sin\delta_n = \frac{2\omega_n\beta}{\sqrt{\left(\omega_0^2 - \omega_n{}^2\right)^2 + 4\omega_n{}^2\beta^2}} \tag{1.20}$$

$$\cos\delta_n = \frac{\omega_0^2 - \omega_n{}^2}{\sqrt{\left(\omega_0^2 - \omega_n{}^2\right)^2 + 4\omega_n{}^2\beta^2}} \tag{1.21}$$

then we can substitute (1.20) and (1.21) into (1.19) to obtain

$$D_n = \frac{\omega_n{}^2}{\sqrt{\left(\omega_0^2 - \omega_n{}^2\right)^2 + 4\omega_n{}^2\beta^2}} \tag{1.22}$$

thus, the steady-state solution of equation (1.11) is

$$x_n(t) = \frac{\omega_n{}^2}{\sqrt{\left(\omega_0^2 - \omega_n{}^2\right)^2 + 4\omega_n{}^2\beta^2}}\cos\left(\omega_n t - \delta_n - \theta_n\right) \tag{1.23}$$

where,

$$\delta_n = \tan^{-1}\left(\frac{2\omega_n\beta}{\omega_0^2 - \omega_n{}^2}\right) \tag{1.24}$$

Notice that

$$x_n(t) \approx -\cos(\omega_n t) = -U(t) \quad \text{when} \quad \omega_n \gg \omega_0 \tag{1.25}$$

and that

$$x_n(t) \approx -\frac{\omega_n^2}{\omega_0^2}\cos(\omega_n t) = -\frac{m}{k}\ddot{U}(t) \quad \text{when} \quad \omega_n \ll \omega_0 \tag{1.26}$$

That is the motion of the mass with respect to the seismometer case is proportional to ground displacement at high frequencies and it is proportional to acceleration at low frequencies (compared to the natural frequency).

Figure 1.2 shows the amplitude and phase of $x_n(t-\delta)$ as a function of frequency for an SDOF.

Notice that when the damping is $1/\sqrt{2}$, then there is the maximum response without having a peak in the response curve. Most manufacturers of seismometers attempt to achieve this level of damping. Figure 1.2 can be thought of as an amplification factor as a function of frequency for ground acceleration. That is the size of the record x(t) is the ground acceleration times the amplification factor. Notice that the instrument response is proportional to ground acceleration at low frequencies.

The amplification can be determined as a function of the amplitude of the ground displacement by simply multiplying the response by ω^2. In this case the instrument response looks like Figure 1.3.

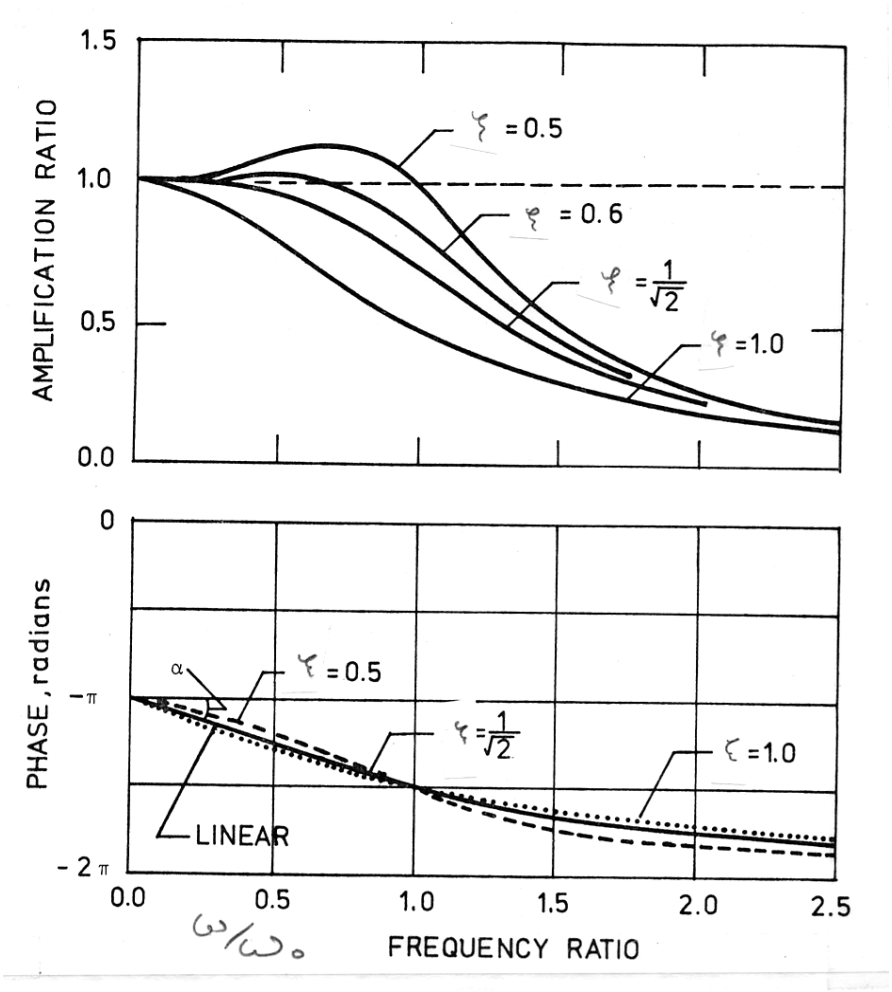


Figure 1.2. Amplification ratio is $X/\ddot{U}$

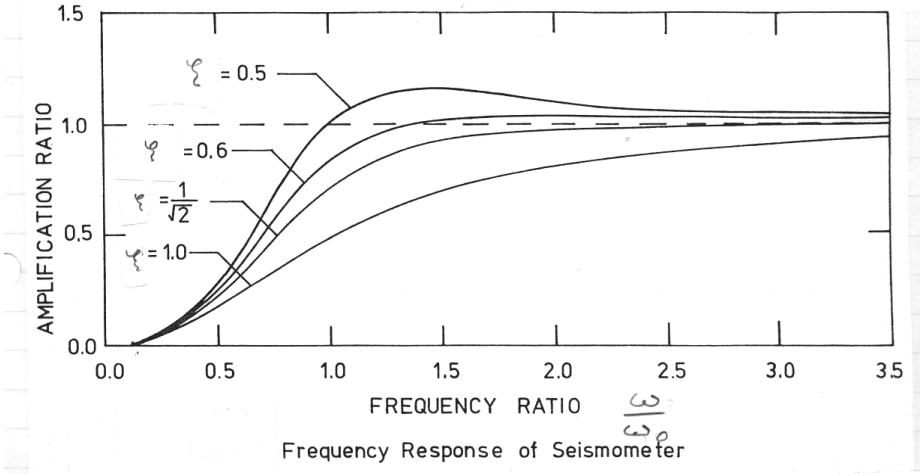


Figure 1.3. Amplification ratio is X/U

We can also find the resonance frequency ω_R at which the SDOF has its maximum amplitude response to a forced vibration by finding when the derivative of the response is zero, or,

$$\left.\frac{dx_n}{d\omega_n}\right|_{\omega_n=\omega_R} = 0 \tag{1.27}$$

Performing this differentiation on equation (1.23) gives

$$\omega_R = \sqrt{\omega_0^2 - 2\beta^2} = \omega_0\sqrt{1-2\zeta^2} \tag{1.28}$$

To add to the things to remember, there is yet another way to describe the system damping called the "quality factor" *Q*, which is defined as

$$Q \equiv \frac{\omega_R}{2\beta} = \frac{\sqrt{1-2\zeta^2}}{2\zeta} \tag{1.29}$$

For lightly damped systems, it can be shown that

$$Q \approx 2\pi\left(\frac{\text{Total Energy in One Cycle}}{\text{Energy Loss During One Cycle}}\right) \approx \frac{1}{2\zeta} \tag{1.30}$$

For lightly damped systems, *Q* can also be determined from the resonance curve of the system.

$$Q \approx \frac{\omega_0}{\Delta\omega} \tag{*1.31*}$$

where $\Delta\omega$ is the frequency interval between the points at which the amplitude of x_n is $\frac{1}{\sqrt{2}}x(\omega_R)$, its maximum. Although I have encountered authors who use Q to describe heavily damped systems, I recommend against it; Q.

Impulse Response

We just saw how we can derive the response to an arbitrary periodic function by decomposing the function into a sum of sinusoids and cosines. However, there are other alternative decompositions that can be quite useful. For example, we can approximate any function by a series of stair steps up and down in time. We define a Heaviside step function as

$$H(t) \equiv \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \tag{1.32}$$

Now suppose that we have the problem of a SDOF subjected to a step function in acceleration with the initial condition that the mass starts at rest. We can write the equation of motion as

$$\ddot{x}+2\beta\dot{x}+\omega_0^2 x=H(t) \tag{1.33}$$

with the initial conditions that

$$x(t=0)=\dot{x}(t=0)=0 \tag{1.34}$$

Unlike the problem of the harmonically driven oscillator, for which the solution was entirely the particular solution, the complementary solution to the homogeneous equation (the transient solution) is very important. The general solution can be written

$$x(t)=x_C(t)+x_P(t) \tag{1.35}$$

If $\beta<\omega_0$ (underdamping), then the complementary solution can be written as

$$x_C(t)=e^{-\beta t}\left(A\cos\omega_1 t+B\sin\omega_1 t\right) \tag{1.36}$$

where

$$\omega_1\equiv\sqrt{\omega_0^2-\beta^2} \tag{1.37}$$

The particular solution, which is the steady-state solution, can be found by inspection.

$$x_P(t)=\frac{H(t)}{\omega_0^2} \tag{1.38}$$

Substituting (1.38) and (1.36) into (1.35), and applying boundary conditions (1.34) leads to

$$x(t)=\frac{1}{\omega_0^2}\left(1-e^{-\beta t}\cos\omega_1 t-\frac{\beta e^{-\beta t}}{\omega_1}\sin\omega_1 t\right)H(t) \tag{1.39}$$

which is shown in figure 1.4.

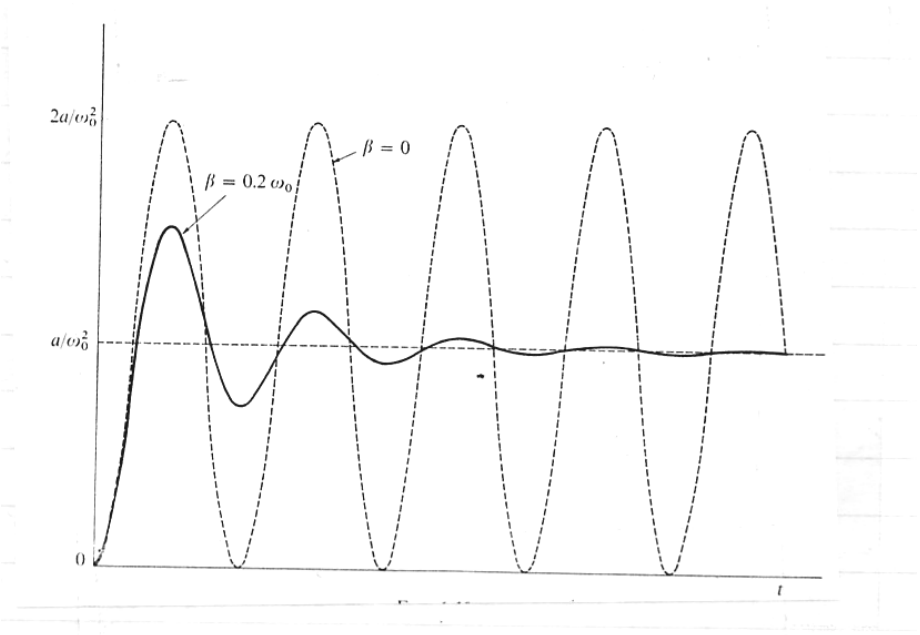


Figure 1.4 shows *x(t)* for both 20% damping and no damping.

We can take the time derivative of the solution for a step in acceleration, equation (1.39), to derive the solution for an impulse in time (delta function, $\delta(t)$).

$$G(t) = H(t)\frac{1}{\omega_1}e^{-\beta t}\sin\omega_1 t \tag{1.40}$$

where *G(t)* is now called a Green's function for this system. The Green's function for 20% damping is shown in Figure 1.5. Although we solved this Green's function problem as a forced vibration problem, we would have obtained the same answer if we had solved a free vibration problem (the homogeneous problem) but with initial conditions of zero displacement and a velocity of unity. Since the integral of the delta function acceleration is a step in velocity, this problem could have been solved in this alternative fashion. In this case, it is easy to see that equation (1.40) consists entirely of the transient response.

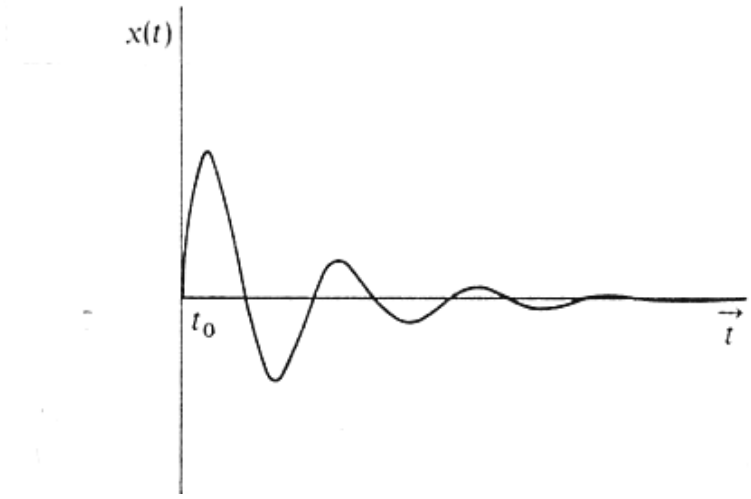


Figure 1.5. The response of an SDOF to an impulse of acceleration.

Convolution

With our Green's function in hand, we can find the response to any ground motion using the convolution operator, which is defined as follows.

$$x(t) = \ddot{u}(t) * G(t) \equiv \int_{-\infty}^{\infty} \ddot{u}(\tau) G(t-\tau)\, d\tau \tag{1.41}$$

There are several ways to view the convolution operator. Taken literally, it is the integral of the product of two functions as they slide past each other. This can be shown graphically in a Figure 1.6, which is taken from Bracewell (1965) (Bracewell, 1965). This is an outstanding textbook that describes linear analysis.

Another way to view convolutions is that the Green's function is added to itself many times, but shifted in time τ and with an amplitude that is given by $x(\tau)$. This is the same as saying that the source consists of a continuous sequence of delta functions. Each one excites the response given by (1.40). Then all these delta functions are added together.

We can solve the integral(1.41) as a discrete problem in the following way. Define

$$x_i = x(i\Delta t); \quad i = 1, n \tag{1.42}$$

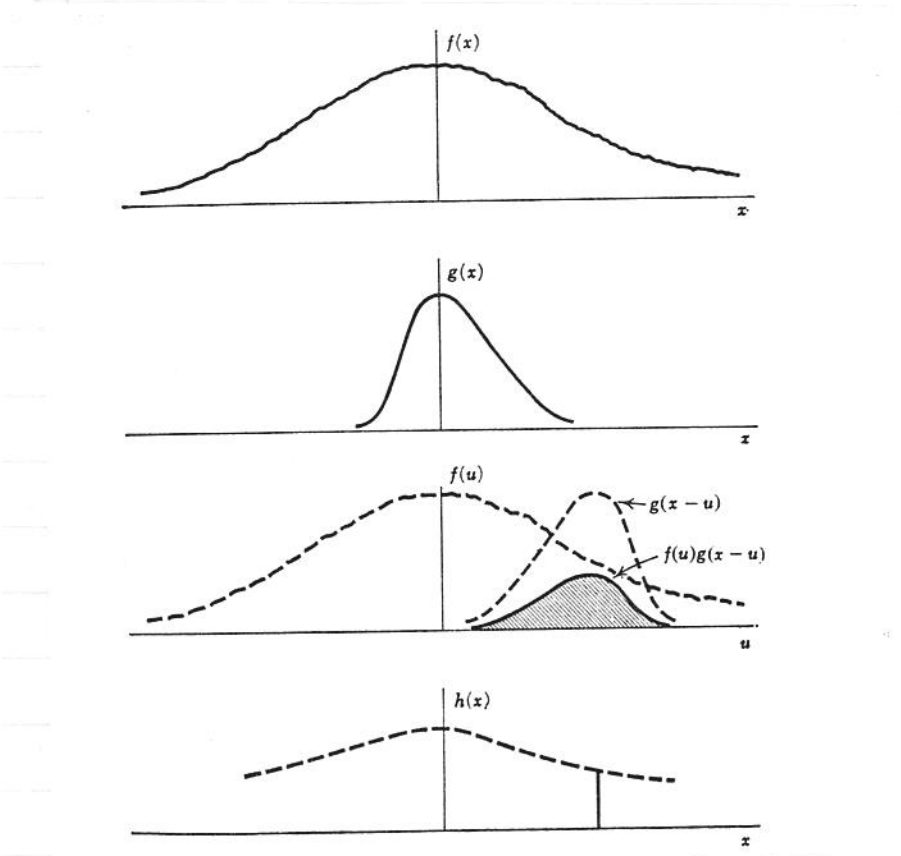
$$\ddot{u}_i = \ddot{u}(i\Delta t); \quad i = 1, n \tag{1.43}$$

$$G_i = G(i\Delta t); \quad i = 1, n \tag{1.44}$$

then

$$x_i = (\ddot{u} * G)_i = \Delta t \sum_{j=1}^{i} \ddot{u}_j G_{i-j} \tag{1.45}$$

which is the **serial product** of $[u_1, ..., u_n]$ and $[G_1, ..., G_n]$. Bracewell (1965) discusses the details.



7. 3.1 *The convolution integral* $h(x) = f(x) * g(x)$ *represented by a shaded area.*

Figure 1.6 A graphical representation of the convolution of two functions (from Bracewell).

Another approach to convolution (and to solving linear differential equations) is to use Fourier transforms. In practice, Finite-Fourier transforms (FFT), which are actually Fourier series, are used in numerical calculations. We define the Fourier transform of *x(t)* as follows.

$$\tilde{x}(\omega) \equiv FT\left[x(t)\right] \equiv \int_{-\infty}^{\infty} x(t) e^{i\omega t} dt \tag{1.46}$$

with an inverse transform given by

$$x(t) = FT^{-1}\left[\tilde{x}(\omega)\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{x}(\omega) e^{i\omega t} d\omega \tag{1.47}$$

An important aspect of Fourier transforms is that differentiation in the time domain is equivalent to multiplying by $i\omega$ in the frequency domain. This allows us to solve linear differential equations algebraically. That is,

$$\tilde{\dot{x}}(\omega) = i\omega\tilde{x}(\omega) \tag{1.48}$$

Another important property of Fourier transforms is the stretch rule.

$$FT\left[x(at)\right] = \frac{1}{|a|}\tilde{x}\left(\frac{\omega}{a}\right) \tag{1.49}$$

Now since the Fourier transform is itself a linear operator, we can take the Fourier transform of our entire differential equation (1.2).

$$-\omega^2\tilde{x}(\omega) + 2i\beta\omega\tilde{x}(\omega) + \omega_0^2\tilde{x}(\omega) = \tilde{\ddot{u}}(\omega) \tag{1.50}$$

or

$$\tilde{x}(\omega) = \frac{\tilde{\ddot{u}}(\omega)}{\omega_0^2 - \omega^2 + 2i\omega\beta} \tag{1.51}$$

or

$$x(t) = FT^{-1}\left[\frac{\tilde{\ddot{u}}(\omega)}{\omega_0^2 - \omega^2 + 2i\omega\beta}\right] \tag{1.52}$$

It just so happens that the Fourier transform of *G(t)* , which is given by equation (1.40), is

$$\tilde{G}(\omega) = FT\left[H(t)\frac{1}{\omega_1}e^{-\beta t}\sin\omega_1 t\right] = \left[\frac{1}{\omega_0^2 - \omega^2 + 2i\omega\beta}\right] \tag{1.53}$$

Therefore, we can now write

$$x(t) = \ddot{u}(t) * G(t) = FT^{-1}\left[\tilde{\ddot{u}}(\omega)\cdot\tilde{G}(\omega)\right] \tag{1.54}$$

That is, convolution in the time domain is equivalent to simple multiplication in the frequency domain.

In practice, the discrete values of *x(t)* are usually calculated by taking the inverse FFT of the product of the FFT's of *G(t)* and $\ddot{u}(t)$.

Properties of Convolution

Convolution is ubiquitous to linear vibrational problems. In this section I summarize some of the useful properties of the convolution of functions. In the following, *f, g,* and *h* are all arbitrary functions of the same real variable (usually time for our problems).

$$FT(f*g) = FT(f)\cdot FT(g) \tag{1.55}$$

$$f*g = g*f \tag{1.56}$$

$$(f*g)*h = f*(g*h) \tag{1.57}$$

$$f*(g+h) = f*g + f*h \tag{1.58}$$

$$\frac{\partial(f*g)}{\partial t} = \dot{f}*g = f*\dot{g} \tag{1.59}$$

$$f(t)*g(t-\tau) = f(t-\tau)*g(t) \tag{1.60}$$

$$f*\delta = f, \quad \text{where } \delta \text{ is a dirac-delta function} \tag{1.61}$$

$$f*H = \int f dt, \quad \text{where } H \text{ is a Heaviside step function} \tag{1.62}$$

Pay special attention to differentiation (1.59); **the normal chain rule does not apply.**

Convolution is the operator that is used in linear filtering. For example, a Rectangle function (sometimes called a boxcar) is defined as

$$\Pi(t) = \begin{cases} 0 & t < -1/2 \\ 1 & -1/2 \le t \le 1/2 \\ 0 & t > 1/2 \end{cases} \tag{1.63}$$

and its Fourier transform is

$$\tilde{\Pi}(\omega) = \frac{2\sin\frac{\omega}{2}}{\omega} \equiv \operatorname{sinc}\left(\frac{\omega}{2}\right) \tag{1.64}$$

Convolution with this function is identical to taking the running mean of a function where the width of the running mean is unity in the time domain. It is also the same thing as multiplying by a sinc function in the frequency domain. A sinc function is shown in figure1.7. The sinc function has the property that its amplitude decays as ω^{-1} at high frequencies and it approaches a value of 1 at low frequencies. Thus, filtering with a rectangle function (a running mean) causes the signal to decay as ω^{-1} at high frequencies compared to the unfiltered signal.

Figure 1.8 shows a sinc function plotted on a log-log scale. Log-log plots are useful for recognizing power-law relationships since if

$$y = x^{\alpha} \tag{1.65}$$

then

$$\log y = \alpha \log x \tag{1.66}$$

which is a linear relationship on a log-log plot. The slope of the relationship is the exponent of the power law. Notice the linear slope in Figure 1.8 that corresponds to a spectral amplitude decay of ω^{-1}. This decay rate is sometimes referred to as 6 dB per octave, since each octave refers to a factor of 2 in frequency and 6 dB refers to a factor of 2 in amplitude (there's more about dB's in Chapter 2). A filter with an ω^{-2} spectral decay (e.g., sinc^2) has a spectral decay of 12 db per octave.

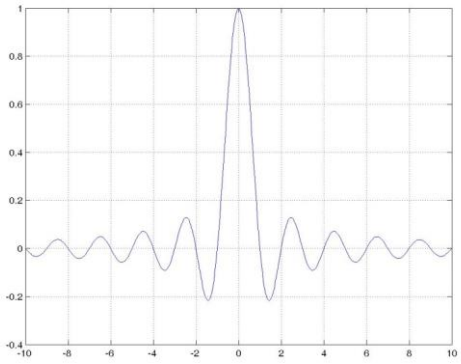


Fig. 1.7 plot of sinc(ω).

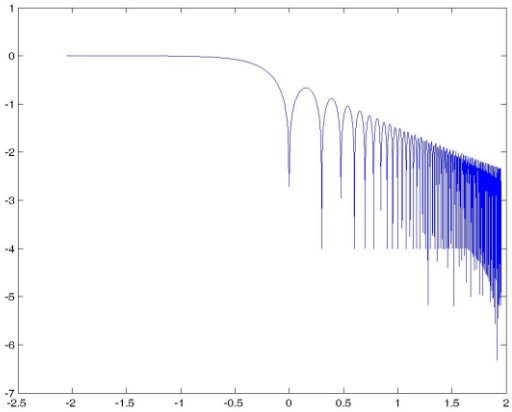


Fig. 1.8 log-log plot of |sinc(ω)|.

Deconvolution

A common problem in seismology is that of determining the ground motion *u(t)* that produced a particular seismogram *x(t)*. We can invert the convolution process. That is, if

$$\tilde{x}(\omega) = \tilde{\ddot{u}}(\omega)\tilde{G}(\omega) \tag{1.67}$$

then

$$\ddot{u}(t) = FT^{-1}\left[\frac{\tilde{x}(\omega)}{\tilde{G}(\omega)}\right] \tag{1.68}$$

This operation is typically performed using an FFT. However, as was the case with convolution, it is possible to do the calculation recursively in the time domain using serial division. That is the inverse of equation (1.45) is

$$\ddot{u}_i = \frac{1}{\Delta t G_0}\left(x_i - \sum_{j=0}^{i-1} \ddot{u}_j G_{i-j}\right) \tag{1.69}$$

In practice, it is often not feasible to perform an exact deconvolution since it often involves a division by zero or a very small number in the frequency domain. That is, most deconvolutions are only valid over a limited frequency band.

One approach to deconvolution is pose it as a linear inverse problem. That is, if we are dealing with the discrete problem, then we can write $x(t) = \ddot{u}(t) * G(t)$ as the following matrix equation (this is a serial product).

$$\mathbf{G}\cdot\ddot{\mathbf{u}}=\mathbf{x} \tag{1.70}$$

Where

$$\mathbf{G}\equiv\begin{bmatrix} G_1 & G_n & G_{n-1} & G_{n-2} & . & G_2 \\ G_2 & G_1 & G_n & G_{n-1} & . & G_3 \\ G_3 & G_2 & G_1 & G_n & . & G_4 \\ . & . & G_2 & G_1 & . & . \\ G_{n-1} & G_{n-2} & . & . & . & G_n \\ G_n & G_{n-1} & G_{n-2} & . & G_2 & G_1 \end{bmatrix} \tag{1.71}$$

$$\mathbf{x}\equiv\left(x_1,x_2,x_3,...,x_n\right) \tag{1.72}$$

$$\ddot{\mathbf{u}}\equiv\left(u_1,u_2,u_3,...,u_n\right) \tag{1.73}$$

This means that if we know $\mathbf{G}$ and $\mathbf{x}$, then we can determine $\ddot{\mathbf{u}}$ as

$$\ddot{\mathbf{u}}=\mathbf{G}^{-1}\cdot\mathbf{x} \tag{1.74}$$

Where

$$\mathbf{G}^{-1}\cdot\mathbf{G}=\mathbf{I} \tag{1.75}$$

While this formulation is exactly equivalent to using an FFT and division in the frequency domain, there are stabilization techniques (e.g., singular value decomposition) in linear inverse problems that can help to reduce instabilities due to ill-conditioned decovolutions (i.e. division by small numbers in the frequency domain). Since the SDOF oscillator problem is linear in all aspects, it is not surprising that it is equivalent to a linear algebra problem, namely the inversion of the matrix **G** can be done by finding eigenvectors and eigenvalues.

Direct Solution

The solution techniques that we have shown explicitly, or implicitly, assume that we are dealing with periodic functions. That is, we are usually assuming that the record time series repeats itself indefinitely. This can be seen directly in the structure of the matrix **G** in (1.71). This can cause complications if there is a discontinuous jump from the end of a record and the beginning. That is, there is often a step change in the value at the beginning/end of the repeating function. Of course, this repeating jump is unphysical and is simply the result of truncating our analysis after some finite time. One simple way to deconvolve the record that avoids this problem is direct integration of equation(1.2).

$$\ddot{u}=-\ddot{x}-2\beta\dot{x}-\omega_0^2 x \tag{1.76}$$

Integrating once, we obtain

$$\begin{aligned}\dot{u}(t) &= -\int_0^t \ddot{x}dt - 2\beta\int_o^t \dot{x}dt - \omega_0^2\int_0^t xdt \\ &= -\dot{x} - 2\beta x - \omega_0^2\int_0^t xdt + C\end{aligned} \tag{1.77}$$

Integrating a 2nd time, we obtain

$$u(t) = -x - 2\beta\int_0^t xdt - \omega_0^2\int_0^t\int_0^t xdt + Ct + D \tag{1.78}$$

Where C and D are constants of integration.

If $u(t=0) = \dot{u}(t=0) = x(t=0) = \dot{x}(t=0) = 0$, then $C = D = 0$.

Chapter 1. Homework

Problem 1.1 Derive equation (1.30).

Problem 1.2 Calculate and sketch the function given by $\Pi\left(t/T_1\right) * \Pi\left(t/T_2\right)$. What does the Fourier amplitude spectrum look like?

Problem 1.3 Derive the response of an SDOF to ground motion described by

$$U(t) = \begin{Bmatrix} 0, & t < 0 \\ \dfrac{t}{T_1}, & 0 < t < T_1 \\ 1, & t > T_1 \end{Bmatrix}$$

It is common to separate seismographic systems into "seismographs" and "strong-motion accelerographs." Seismographs have generally been developed by geophysicists and seismographs are typically designed to record ground motions that are far too small to be felt. Strong-motion accelerographs have usually been designed by earthquake engineers to record the ground acceleration during severe earthquake shaking. In actuality, though, there is nothing fundamentally different in the physics of these two systems.

Seismographs (including strong-motion accelerometers) consist of (at least) a sensing unit and a recording unit. I will begin by discussing the recording system. Current state-of-the-art is to record output voltages from a seismometer (the sensing system) with a digital data logger, which typically consists of an analog to digital converter (ADC) and some type of digital computer for processing, storage, and communications.

Dynamic Range

The most critical specification of a data logger is its dynamic range, which is defined as the ratio of the largest on-scale voltage V_{max} divided by the smallest resolvable voltage V_{min}. That is,

$$DR = \frac{V_{\max}}{V_{\min}} \tag{2.1}$$

Traditionally, dynamic range is given in the somewhat obscure units of decibels dB (a tenth of a Bell). This nomenclature was originated with measuring the relative intensity of sound, which is proportional to the power of acoustic waves. Bells are a base 10 logarithmic measure of energy per unit time, and 1 Bell corresponds to a factor of 10 increase in energy per unit time. However, most of our discussions concern the amplitude of some signal as opposed to its power. Since power is proportional to the square of the amplitude, a factor of 10 in amplitude corresponds to a factor of 100 in power, which is 2 Bells or 20dB. Therefore, we can define dynamic range in units of dB as

$$DR = 20 \cdot \log\left(\frac{V_{\max}}{V_{\min}}\right) \text{dB} \qquad (2.2)$$

The dynamic range of a digitizer is typically determined by the number of bits that it uses to characterize voltage. Each bit represents a factor of 2 in dynamic range, so the dynamic range is $2^{\#\text{bits}}$. Since $\log 2 = 0.301$, the dynamic range in dB is #bits times 6.02.

Early digitizers were typically 8-bit or 12-bit units. 16-bit units were common by the mid 1980's, and 20-bit or 24-bit units had become the standard by the mid 1990's.

Some examples of dynamic range are given in Table 2.1.

Table 2.1

# of bits	Dynamic range	dB
8	256	48
12	4,096	72
16	65,456	96
20	1,048,560	120
24	16,777,216	144

Paper or film recording devices were the most common system prior to introduction of digital systems. These older analog systems typically had a dynamic range of 50 to 60 dB, depending on how well the trace could be measured. As we shall see later, the total dynamic range of motions encountered in the Earth is on the order of 200 dB, and current digitizers do not come close to having the dynamic range to record both the strongest motions in earthquakes and the smallest motions that occur at quiet sites.

Most electro-mechanical seismometers have an effective dynamic range that is about 100 dB, which is much larger than the range of optical recording systems. However, electronic feedback seismometers (described below) typically have dynamic ranges of about 140 dB.

Seismometers

Seismometers are the sensors that produce the signal to be recorded. Modern seismometers produce some voltage that is related to the ground motion by the instrument response. The earliest seismometers (circa 1900) consisted of a

mass, a spring, and sometimes a damper. The mass was usually very large since its motion was typically measured by a series of levers that caused a needle stylus to move over a rotating drum covered with smoked paper. Thus, it was necessary for the small motions of the ground to cause enough momentum in the mass to overcome the friction of the recording system.

In 1922, Harry Wood (a seismologist) and John Anderson (an astronomer) collaborated to build a simple system known as the Wood-Anderson torsion seismometer. They developed a system that illuminated a mirror on a mass suspended by a vertical wire that served as a torsion spring. When the ground moved horizontally, the wire would twist, causing a deflection of the reflected light. The reflected light was focused onto a rotating drum covered with photographic paper. The motion of the mass was damped electro-magnetically. The Wood –Anderson has an undamped natural period of 0.8 s, its gain is 2,800, and its damping is 70% of critical. Several dozen Wood-Anderson seismometers were operated in southern California until about 1980. This system was the standard that was used by Richter (1957) (Richter, 1957) in the definition of earthquake magnitude in the 1930's (see Chapter 8). The response of this instrument is that of a simple single degree-of-freedom oscillator.

Many strong-motion accelerographs were also simple optical SDOF's. More than 10,000 SMA-1 series of accelerographs were manufactured by Kinemetrics in Pasadena from the late 1960's to the mid 1990's. This instrument also has a mirror that deflects in torsion. It's natural frequency is about 30 Hz and it also is 70% damped. A sketch of this instrument is shown in Figure 2.1. At frequencies lower than 30 Hz, the records from this instrument are proportional to ground acceleration (see Figure 1.2). These instruments record on 70 mm film and they only record when triggered by vertical accelerations that exceed about 1% g. The clip level on an SMA1 is about 1.5 g. When the frequency of the signal exceeds 15 Hz, it is necessary to deconvolve the instrument response to obtain true ground acceleration from this instrument.
Figure 2.2 shows a schematic of another common strong motion accelerometer (SMAC) used in Japan from the 1960's through the 1980's. This is a purely mechanical instrument with air damping. Unfortunately, static friction in the system made this a poor system for recording motions at periods exceeding 2 seconds.

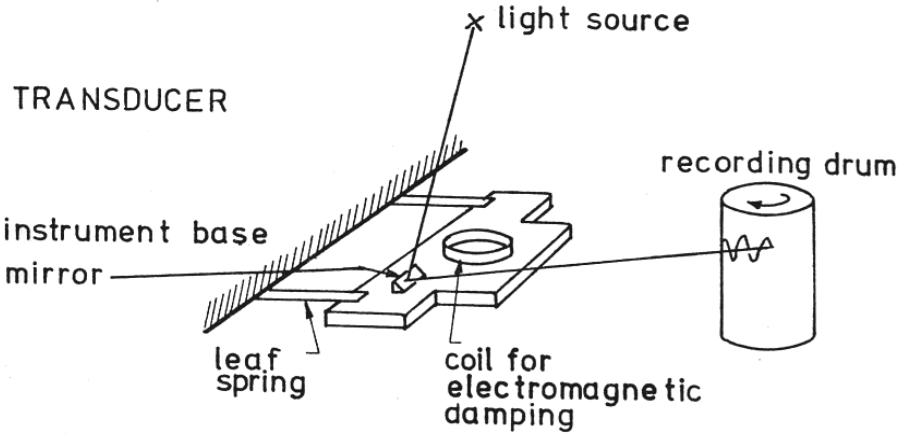


Figure 2.1 Schematic of the mechanics of an SMA-1 strong motion accelerometer.

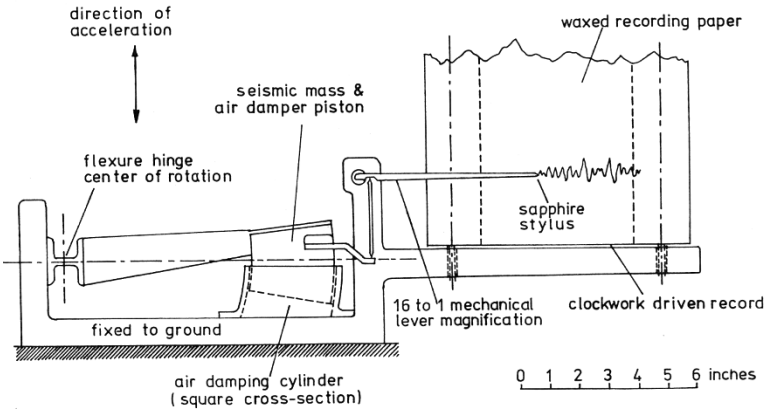


Fig. 2.8 Schematic Diagram of SMAC Accelerograph (from Hudson, 1970 [15])

Figure 2.2. Schematic of a Japanese SMAC strong motion acclerograph.

Table 2.1 gives the instrument constants of a number of strong-motion seismographs.

Instrument	Undamped Natural Frequency (Hz)	Damping Ratio	Damping Type	Accelerograph Sensitivity (cm/g)	Full Scale (g)	Country	Manufacturer or Supplier
USCGS Standard	11.8-23.3	0.6/1.0	Magnetic	5.5-19.7	1	U.S.A.	USCGS
AR-240	15.4-25.0	0.55-0.65	Electro-magnetic	5-7.5	1	U.S.A.	Teledyne, Inc., Pasadena
RFT-250	20	0.6	Electro-magnetic	1.9	1	U.S.A.	Teledyne, Inc., Pasadena
RMT-280	20	0.6	Electro-magnetic	±200Hz FM deviation/±1g	1	U.S.A.	Teledyne, Inc., Pasadena
SMA-1	20-25	0.6	Electro-magnetic	1.9	1	U.S.A.	Kinemetrics, Inc.
SMAC-A	10	1	Air Piston	4	1	Japan	Akashi Seisakusho, Ltd., Tokyo
SMAC-B(B2)	10(7.14)	1	Air Piston	4(8)	1(0.5)	Japan	Akashi Seisakusho, Ltd., Tokyo
SMAC-C(C2)	10(7.14)	1	Air Piston	4(8)	1(0.5)	Japan	Akashi Seisakusho, Ltd., Tokyo
SMAC-D	20	1	Air Piston	1	0.5	Japan	Akashi Seisakusho, Ltd., Tokyo
SMAC-E(E2)	20	0.6	Air Piston	0.5(1)	1(0.5)	Japan	Akashi Seisakusho, Ltd., Tokyo
DC(DC3C)	10	1	Air Piston (Oil Piston)	8(4)	0.5(0.1)	Japan	Hosaka Seismograph Co.
SMAC-Q	20	0.6	Air Piston	0.5	1	Japan	Akashi Seisakusho, Ltd., Tokyo
MO-2	33	0.6	Oil Paddle	1.5 Horizontal 2.2 Vertical	1	New Zealand	Victoria Engineering Ltd.
UAR	20	0.7	Electro-magnetic	1.6	1	USSR	Earth Physics Institute

Seismometers with electromagnetic pickups were developed by Galitzen in about 1910. Currents generated by the seismometer were used to drive a galvanometer that deflected a beam of light. A number of different types of these instruments became popular in the 1930's. These instruments gave new flexibility since the signals could be electronically amplified and filtered. Velocity transducers are the most common type of pickup and they typically consist of a magnetized mass that moves through a conducting coil. The voltage generated in the coil is proportional to the velocity of the mass with respect to the coil, and hence the term, velocity transducer. Benioff short-period seismometers with a 1-second free period, 70% damping, and velocity transducers were important standards in seismology.

By the 1970's more compact 1-second velocity seismometers were manufactured in large numbers for use in exploring for petroleum. Over a thousand L4-C (Mark Products, Inc) 1-second seismometers were employed by regional seismic U.S. networks from the 1970's through the 1990's.

The electrical output from these seismometers have been recorded several ways. In many important seismographic systems, the electrical current from the seismometer was used to drive galvanometers. These galvanometers consisted of a mirror suspended on a torsion wire. Deflection of the mirror was measured photographically as in other direct seismometers. The

galvanometer was itself a linear SDOF whose forcing was the output from the seismometer. The galvanometer/seismometer system constitutes two coupled linear oscillators. The exact solution to this problem is a normal mode problem that is described in Chapter 6. However, the momentum of the seismometer mass greatly exceeds that of the galvanometer. Therefore, the response of these galvanometer seismometer systems is approximately given by.

$$\ddot{x}(t)+2\beta\dot{x}(t)+\omega_0^2 x(t)=\ddot{u}(t) \tag{2.3}$$

$$V(t)=C\dot{x}(t) \tag{2.4}$$

$$\ddot{y}(t)+2\beta_G\dot{y}(t)+\omega_G^2 y(t)=DV(t) \tag{2.5}$$

where *x(t)* and *y(t)* correspond to motion of the seismometer and galvanometer masses, respectively, C and D are constants, $V(t)$ is voltage, β_G and ω_G are the damping and free period of the galvanometer. Note that the driving term in the second equation is not a 2nd derivative with respect to time.

Since the damping of both systems is 70% of critical, and since older seismometer masses were quite large, the feedback from the motion of the galvanometer back into the seismometer is minimized. Therefore we can approximately solve this problem as if the solution to equation (2.3) is used as the input to equation (2.5). As was the case for the simple SDOF, we can write the solution for these equations as a convolution with the Green's functions of the seismometer *G(t)* and the galvanometer $G_G(t)$.

$$\begin{aligned} y(t) &= G_G(t)*V(t) \\ &= G_G(t)*D\frac{d}{dt}\left[CG(t)*\ddot{u}(t)\right] \\ &= CDG_G(t)*G(t)*\ddot{u}(t) \end{aligned} \tag{2.6}$$

where,

$$G(t)=\frac{H(t)}{\sqrt{\omega_0^2-\beta^2}}e^{-\beta t}\sin\left(t\sqrt{\omega_0^2-\beta^2}\right) \tag{2.7}$$

and

$$G_G(t)=\frac{H(t)}{\sqrt{\omega_G^2-\beta_G{}^2}}e^{-\beta_G t}\sin\left(t\sqrt{\omega_G^2-\beta_G{}^2}\right) \tag{2.8}$$

The dynamic range of these electromagnetic seismometers is typically in the range of 80 to 100 dB. This is far greater than the range of the film or paper systems that were used to record the data from them.

Just as before, we can take the Fourier transform of (2.6) to obtain

$$\tilde{y}(\omega) = CD\tilde{G}_G(\omega)\tilde{G}(\omega)\left(-i\omega^3\right)\tilde{u}(\omega) \tag{2.9}$$

where

$$\tilde{G}(\omega) = \frac{1}{\omega_0^2 - \omega^2 + 2i\omega\beta} \tag{2.10}$$

$$\tilde{G}_G(\omega) = \frac{1}{\omega_G^2 - \omega^2 + 2i\omega\beta_G} \tag{2.11}$$

Table 2.1 lists some instrument constants for several seismometer systems. All of these systems are from the early 1900's, except for the WWSSN LP, which was operated as a standard world-wide network from the 1960's to the 1980's. The ground displacement amplification response as a function of period $T = \frac{2\pi}{\omega}$ is shown in Figure 2.3. Figure 2.4 shows a similar plot for a number of important seismographic systems, many of which were operated at the Seismological Laboratory of Caltech.

Table Al. Instrumental Constants

Mechanical Seismograph

Instrument	Pendulum Period, sec	Damping Ratio, ε	Damping Constant, h*	Static Magnification, V
Wiechert 1	5.0	4	0.404	80
Bosch-Omori 1	12.0	4	0.404	40
Bosch-Omori 2	30.0	5.3	0.469	10
Mainka	10.0	2	0.215	120
Wiechert 2	12.6	3.4	0.363	180
Milne-Shaw	12.0	20	0.690	250

*h and ε are related by $\varepsilon = \exp\left[\pi \frac{h}{\sqrt{1-h^2}}\right]$

Electro-Magnetic Seismograph*

Instrument	Pendulum Period, sec	Galvanometer Period, sec	Max. Gain, V
Galitzin 1	12.0	12.0	580
Galitzin 2	25.0	25.0	310
WWSSN LP	15.0	100.0	1500

*Damping constants of the pendulum and the galvanometer are assumed to be 1.0, and the coupling constant is assumed to be 0.02.

Table 2-1

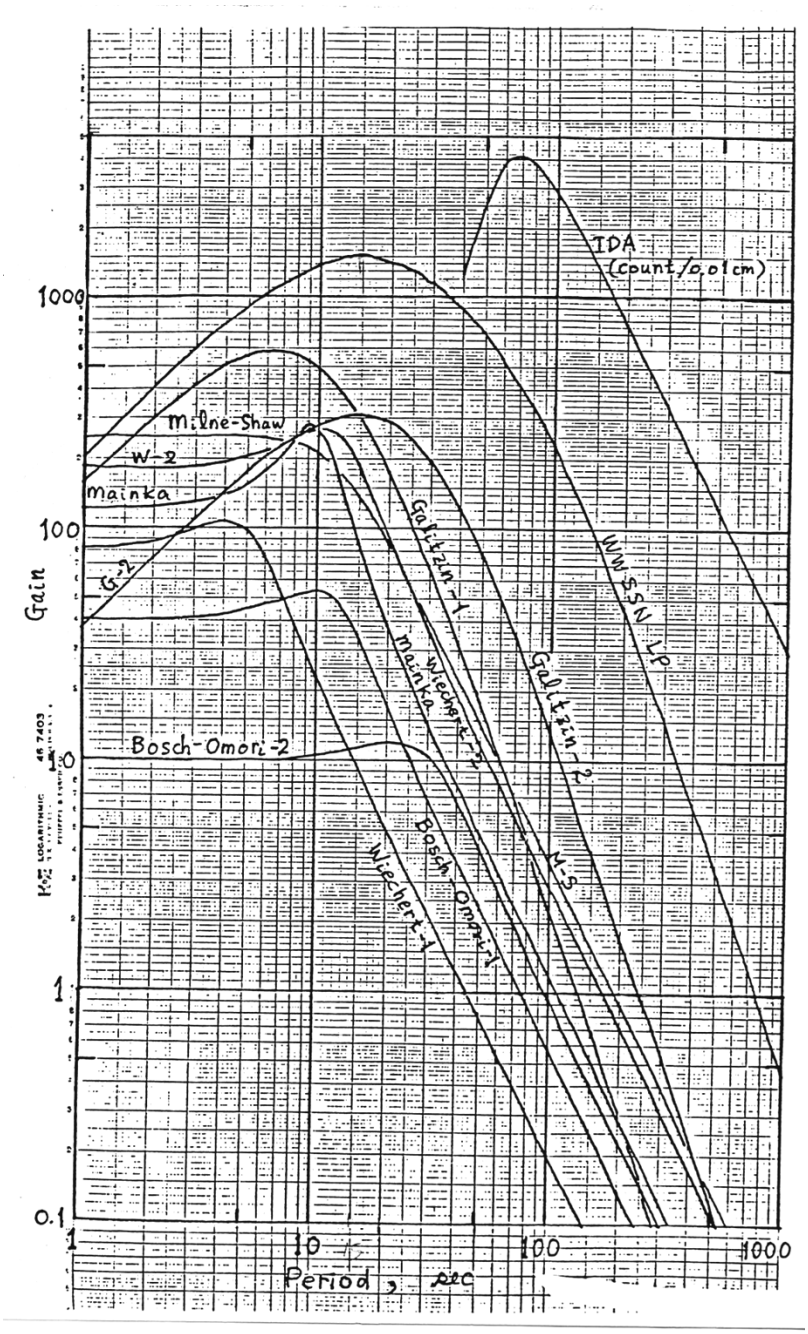


Figure 2.3 Seismometer response for some important early seismometers

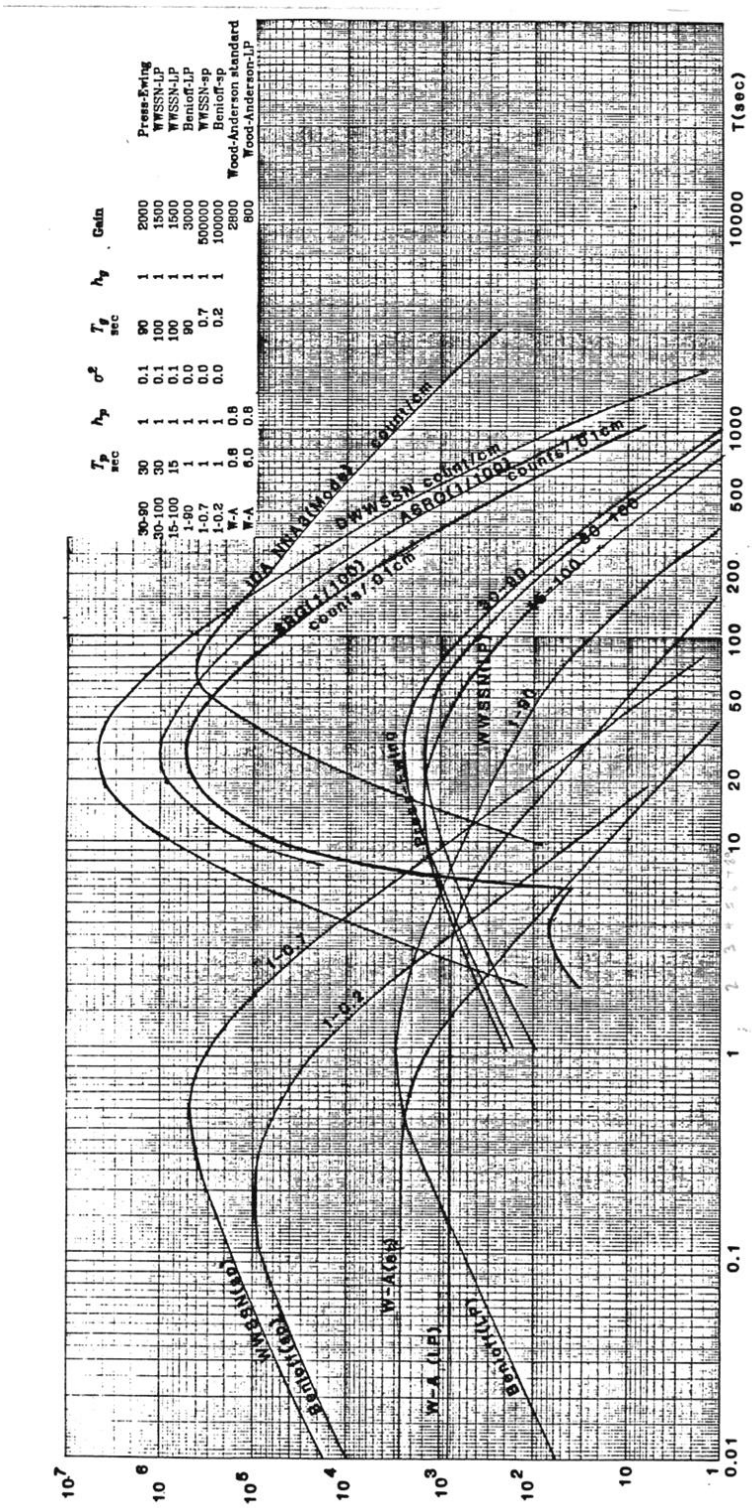


Figure 2.4. Response of seismometer systems used ib the 1950's and 1960/s.

Poles and Zeros

We can rewrite the frequency domain response of our seismometer/galvanometer system (2.9) in the following way

$$y(t) = R(t) * u(t) \tag{2.12}$$

where R(t) is the displacement response of the system given by

$$R(t) = \frac{d}{dt^3}\left[CDG(t) * G_G(t)\right] \tag{2.13}$$

This can be written in the frequency domain as

$$\tilde{y}(\omega) = \tilde{R}(\omega)\tilde{u}(\omega) \tag{2.14}$$

where

$$\begin{aligned}
\tilde{R}(\omega) &= \frac{CD\left(-i\omega^3\right)}{\left(\omega_0^2 - \omega^2 + 2i\omega\beta\right)\left(\omega_G^2 - \omega^2 + 2i\omega\beta_G\right)} \\
&= \frac{CD\left(-i\omega^3\right)}{\left(\omega^2 - \omega_0^2 - 2i\omega\beta\right)\left(\omega^2 - \omega_G^2 - 2i\omega\beta_G\right)} \\
&= CD\frac{(i\omega)^3}{\left[\left(\omega - i\beta\right)^2 + \beta^2 - \omega_0^2\right]\left[\left(\omega - i\beta_G\right)^2 + \beta_G^2 - \omega_G^2\right]} \\
&= iCD\frac{(\omega - z_1)(\omega - z_2)(\omega - z_3)}{(\omega - p_1)(\omega - p_2)(\omega - p_3)(\omega - p_4)}
\end{aligned} \tag{2.15}$$

where

$$z_1 = z_2 = z_3 = 0 \tag{2.16}$$

and

$$p_1 = i\beta - \sqrt{\omega_0^2 - \beta^2} \tag{2.17}$$

$$p_2 = i\beta + \sqrt{\omega_0^2 - \beta^2} \tag{2.18}$$

$$p_3 = i\beta_G - \sqrt{\omega_G^2 - \beta_G^2} \tag{2.19}$$

$$p_4 = i\beta_G + \sqrt{\omega_G^2 - \beta_G^2} \tag{2.20}$$

The p_j and z_i are called the poles and zeros of this system, and together with the station gain, *-iCD*, they define the response of the system.

Unfortunately, there is always confusion about conventions used in transforms. The common convention for poles and zeros is defined by the Standard for Exchange of Earthquake Data (SEED). This standard is described in the SEED user's manual that can be found at http://www.iris.washington.edu/DOCS/manuals.htm. The standard is based on Laplace transforms as opposed to Fourier transforms. These two transforms are very similar except that the poles and zeros are defined in terms of the Laplace transform variable $s = i\omega$. Denoting the poles and zeros by P_i and Z_i, we can rewrite (2.15) as

$$\tilde{R}(\omega) = CD \frac{(s - Z_1)(s - Z_2)(s - Z_3)}{(s - P_1)(s - P_2)(s - P_3)(s - P_4)} \tag{2.21}$$

where

$$Z_1 = Z_2 = Z_3 = 0 \tag{2.22}$$

and

$$P_1 = -ip_1 = \beta + \sqrt{\beta^2 - \omega_0^2} \tag{2.23}$$

$$P_2 = -ip_2 = \beta - \sqrt{\beta^2 - \omega_0^2} \tag{2.24}$$

$$P_3 = -ip_3 = \beta_G + \sqrt{\beta_G^2 - \omega_G^2} \tag{2.25}$$

$$P_4 = -ip_4 = \beta_G - \sqrt{\beta_G{}^2 - \omega_G^2} \tag{2.26}$$

In fact, any complex transfer function that can be written in the form

$$T(\omega) = \frac{a_n \omega^n + a_{n-1}\omega^{n-1} + \ldots + a_0}{b_l \omega^l + b_{l-1}\omega^{l-1} + \ldots + b_0} \tag{2.27}$$

can also be written in the form

$$T(\omega) = \frac{a \prod_{i=1}^{n} (\omega - z_i)}{b \prod_{j=1}^{l} (\omega - p_j)} \tag{2.28}$$

The convention of using poles and zeros is especially useful in systems that can be described as a series of convolutions. Remember that the convolution operator is commutative. That is, these convolutions can be written as a series of multiplications in the frequency domain. Therefore, a linear system can be described by compiling the set of all of the poles and zeros that correspond to each of the functions that are convolved to form the transfer function. If an additional filter or device is added to the system (and if its

effect is that of a convolution), then the poles and zeros of that device are simply added to the set.

Broad-Band Seismometers

It is worth inspecting Figure 2.4 to see that most seismographic systems were designed to have high magnification in either a short-period band (about 1 second) or a long-period band (about 20 seconds). This was accomplished by using short-period galvanometers together with short-period seismometers to make a short period seismograph, or by combining a long-period galvanometer with a long-period seismometer to make a long-period seismograph. However, it was possible to use a short-period seismometer with a long-period galvanometer to make a system which records over a broad range of frequencies. One such instrument was the Benioff 1-90, which had a 1-second velocity transducer seismometer driving a 90-second galvanometer. The response of this instrument (see Figure 2.4) is approximately flat to velocity between 1 and 90 seconds; hence it records velocity over a broad frequency range.

Notice that the amplification of the 1-90 is much less than that of either the short- or long-period systems. This is because there are **microseisms**, which are relatively large amplitude waves continuously, excited by water waves in the ocean at periods between 6 and 12 seconds. There was not much point in making a high-magnification broad-band system since it would fill the seismogram with quasi-harmonic microseisms. The presence of microseismic noise at virtually all stations meant that seismograph designers who wished to detect and locate frequent small-magnitude earthquakes were forced to design either long- or short-period instruments.

Simple optical seismometers (Wood-Anderson, SMA-1 accelerograph) also respond over a broader frequency band, but they have a relatively small overall amplification of signals. Furthermore, their response is flat to acceleration at periods longer than their natural frequency. This means that they are quite insensitive to long-period ground displacements when compared to a seismograph that whose response is flat to velocity.

Figure 2.5 shows the amplitude for many different wave types as a function of frequency. The vertical axis is the log of the max amplitude of a seismogram after filtering with a 1-octave-wide bandpass filter. The curves

labeled maximum and minimum correspond to the background noise level recorded at worldwide seismographic stations. The minimum curve was recorded at a site in Lajitas, Texas. There is really no maximum curve, since one can always find sites with high background noise. It actually represents the noise encountered on ocean island stations where ocean wave generated noise is high.

The various lines shown for different earthquake situations show approximate median amplitudes for earthquakes recorded at approximately 10 km, 100 km, and 3000 km from earthquakes of different magnitudes. You can see that there are more than 200 dB in amplitude difference (10,000,000,000 to 1) between the ambient ground noise and the maximum ground accelerations at seismically quiet sites. The stippled regions show the on-scale range of both an SMA-1 strong motion accelerograph and a typical short-period seismographic channel from a regional seismographic network with analog telemetry (frequency modulated, FM). Almost 1,600 of these short-period seismographs were operating in the United States in the 1980's. These stations were designed to operate at maximum magnification to detect the smallest earthquakes that created motions just larger than the ambient ground motions. Although these stations were well suited for detecting ground motions, they were not well suited for recording them. That is, many earthquake ground motions were too large for the range of the system and they caused clipping.

Some seismological observatories operated a wide variety of seismographs that operated in different amplitude and frequency bands. The Pasadena station routinely recorded several dozen seismograms each day in order to obtain a more or less complete record of ground motion over this vast range of amplitude and frequency.

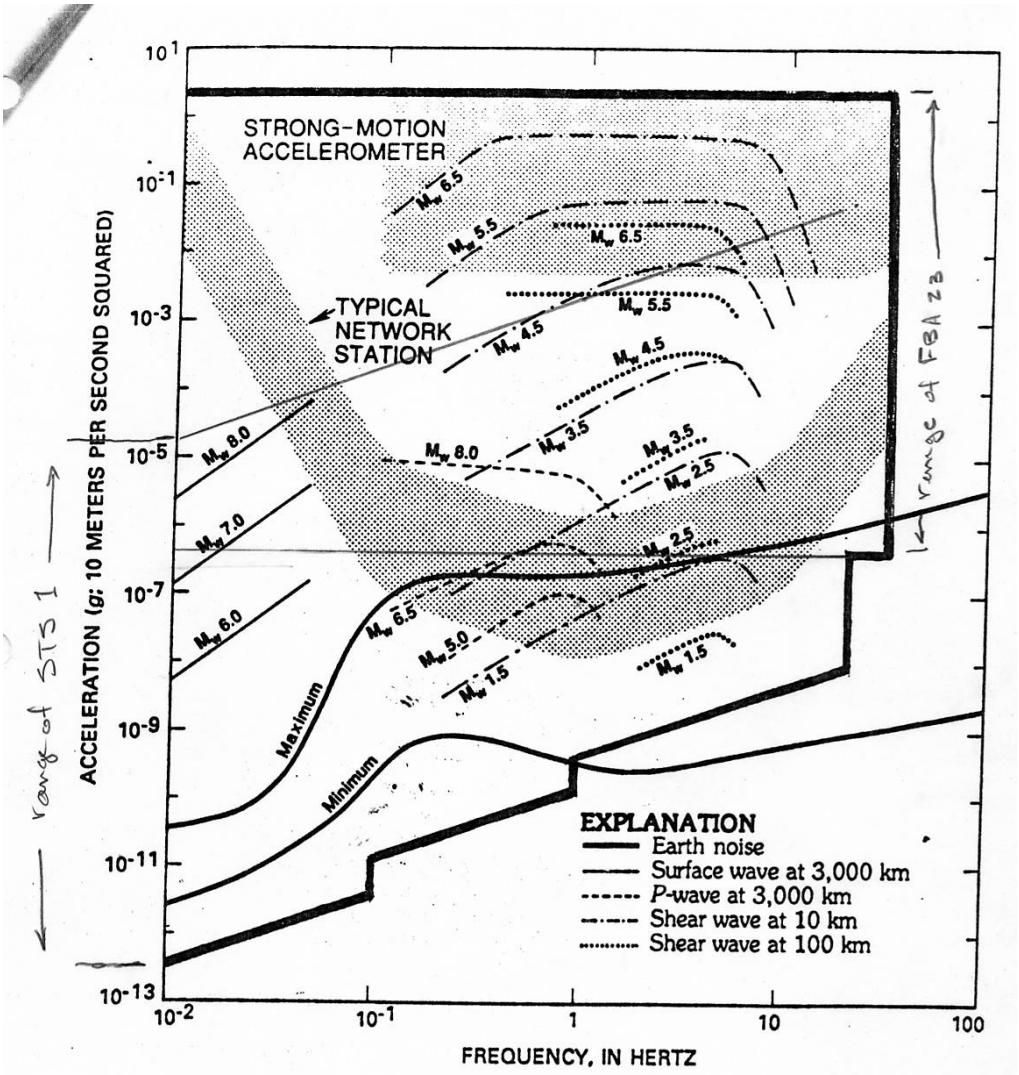


Figure 2. Shaded regions show the dynamic range plotted against frequency for typical analog telemetry station and typical strong-motion station. The heavy lines encompass the region recorded by TERRAscope stations. Also shown are expected levels of ground motion for different seismic arrivals from earthquakes of different sizes and recorded at different distances.

Figure 2.5

Notice that the spectra of ground accelerations from strong motion records of large earthquakes are relatively flat in the frequency band from 0.3 Hz to 5 Hz. Since the recording system of early strong motion accelerographs was less than 60 dB, it was a good choice to record ground acceleration since that was the best way to recover motion in the frequency band from 0.1 Hz to 10 Hz.

In the 1980's seismic instrumentation was revolutionized by the development of **force-feedback** seismometers. These systems are similar to standard seismometers, but they usually have a displacement transducer to measure the motion of the seismometer mass. In addition, they add an electromagnetic forcing system that has the role of minimizing the motion of the mass with respect to the seismometer case. The force necessary to keep the mass stationary is simply the ground acceleration. The essential feature of these systems is that the dynamic range of the instrument is dictated by the dynamic range of the electronic feedback system, and not by the dynamic range of the mechanical seismometer. This essential addition allows modern feedback seismometers to often achieve 140 dB dynamic ranges (a factor of 100 times greater than the dynamic range of mechanical systems).

The electronic feedback system can also be designed to provide the desired instrument response. STS-1 seismometers manufactured by Streckeisen A.G. in Switzerland are considered a standard of excellence for feedback seismometers. They have a mechanical natural period of about 1 Hz that is extended to 0.003 Hz (360 seconds) by the feedback system. Their electronic feedback system is designed to provide an instrument response that is flat to velocity from 360 seconds to 8 Hz. In essence they have a response that is identical to an SDOF with a 360 second natural period and a velocity transducer. The range of amplitudes and frequencies that can be recorded by an STS-1 are shown in Figure 2.5.

Notice that microseismic noise in the .2 Hz to .1 Hz band is several orders of magnitude larger than the minimum motion resolved from an STS-1. Thus, it is necessary to filter in this frequency band if one wants to see small motions in either shorter or longer periods. Such filtering was not feasible when STS-1's were recorded with older systems with limited dynamic range. The development of 24-bit recording systems with dynamic ranges of 140 dB that matched that of the seismometer was the other important development that revolutionized seismographic systems in the 1990's.

A number of other important feedback seismometers have been developed. In particular, the Caltech/USGS network has many stations that use STS-2 seismometers that are flat to velocity from 120 seconds to 30 Hz. These systems are better suited to record small earthquakes and they are also about 1/3 of the cost of STS-1's. STS-1 seismometers are no longer produced.

Currently, strong-motion accelerographs are typically force-feedback systems with stiff (high-frequency) mechanical suspensions. Their output is usually flat to acceleration from static acceleration (sometimes called DC, as in DC current) to 100 Hz. Their dynamic range is also in the range of 140 dB. However, most strong motion accelerographs are still designed to record only during strong shaking (usually a trigger threshold of 0.01 g) and hence it has not been seen as necessary to record with 24-bit resolution. The Kinemetrics K-2 accelerograph has a 20-bit digitizer and it has been a standard at the turn of the millennium.

Stations of the Caltech/USGS seismographic system (Southern California Seismic Network, SCSN; previously,TriNet) have six 24-bit digitizers to record 3 components each of broad-band velocity and strong-motion acceleration. The 24-bit range of the strong-motion accelerometers is also shown in Figure 2.5. The total range of the combined systems is encompassed by the heavy lines in Figure 2.5.
John Clinton prepared the figure on the following page and it shows an updated version of the amplitudes of different signals recorded by the TriNet system in southern California (see www.trinet.org).

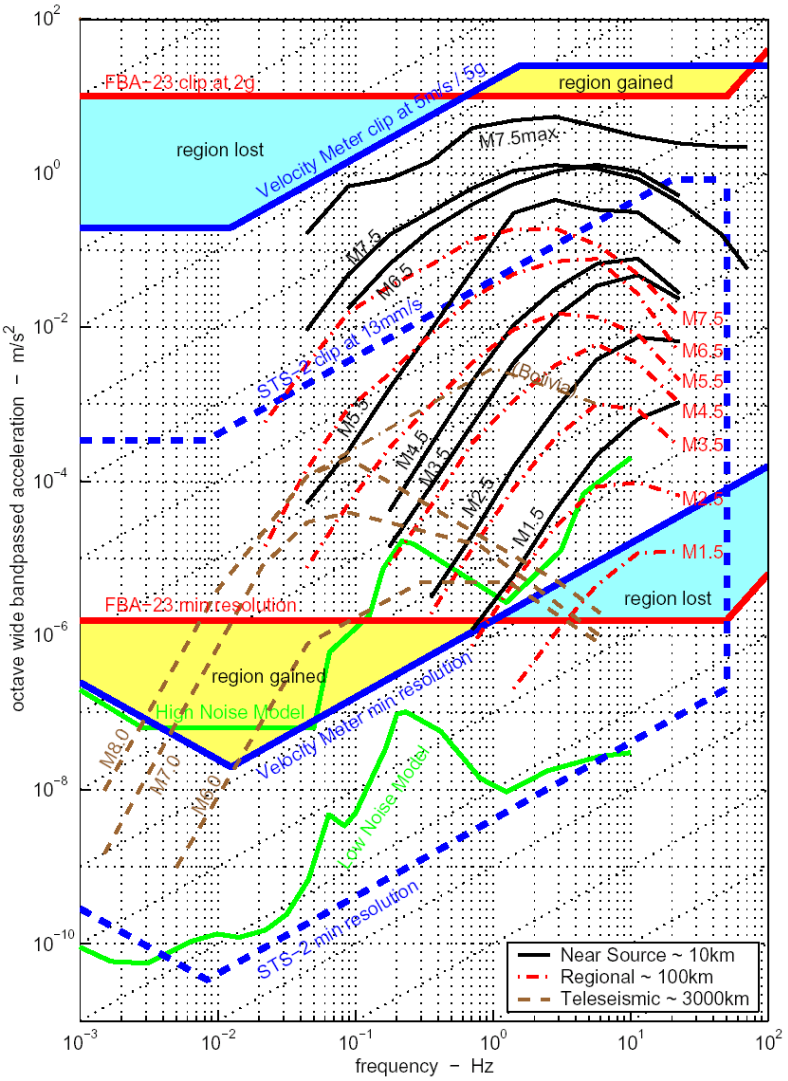


Figure 5. Frequency - Amplitude Plot for Octave Wide Band-passes of Grond Motion Acceleration. Note that instrument limits are scaled down to account for the band-passing of the event data. Ground motions recorded on-scale by the FBA-23 lie between the solid red lines. On-scale motions recorded by the hypothetical low-gain broad-band seismometer lie between the solid blue lines. The dotted blue lines give the dynamic range of the STS-2. The areas shaded light blue are regions of frequency-amplitude space that are recorded by the FBA-23, but not recorded by the hypothetical low-gain broad-band seismometer. Areas shaded yellow are regions of frequency-amplitude space that are recorded by the hypothetical low-gain broad-band seismometer, but not recorded by the FBA-23. Noise levels are the USGS High and Low Noise Models (Peterson, 1993). (*see text for further explanation*)

In this case, the dotted lines outline the range of an STS-2 seismometer and the solid blue lines show the range of a hypothetical strong-motion broad band seismograph that could potentially serve the dual purpose of recording both strong ground motions and large distant earthquakes (teleseisms).

Microelectromechanical systems MEMS Accelerometers

MEMS accelerometers are very small accelerometers that are fabricated with silicon chips. They were first developed to provide inexpensive triggers for air bag deployment in automobiles. They often consist of a small cantilevered conductive plate that is placed in the void space between an upper and lower surface. All of these plates are connected to a circuit that measures the capacitance of the MEMS device. This capacitance is determined by the distance between the plates, which is determined by the flexure of the cantilevered plate. Typically, the undamped frequency of the MEMS is very high (> 100 Hz). Since the output of the MEMS is proportional to the displacement of the mass (displacement transducer), the output of the sensor is proportional to accelerations for frequencies less than the natural frequency.

Of course, the main advantage of a MEMS sensor is that it can be mass produced and can be extremely inexpensive. These sensors are finding more applications as time goes on. Today, all smart phones have MEMS accelerometers; these are used to determine the orientation of the phone (acceleration due to gravity is down) and they are also necessary for some game controllers. The original MEMS accelerometers were designed for automobile collisions and hence their clip range was high (>10 g). Furthermore, there was not much need for precision and the resolution of the devices was often only 40 dB.

Figure 2.7 ONAVI MEMS accelerometer used by QuakeCatcher network. This is a 3-component 80 dB device with a 2 g clip level.

Lately MEMS manufacturers are improving the dynamic range of their devices, which makes them more suitable for use as strong motion seismometers. Towards that end, two separate development projects are attempting to deploy large numbers of MEMS accelerometers to be operated by volunteers. The MEMS accelerometer in Figure 2.7 shows an ONAVI accelerometer that has been deployed by the QuakeCatcher Network that was developed by Jesse Lawrence (Stanford Univ.) and Elizabeth Cochran (USGS, Pasadena). This device was developed for use in navigational equipment and hence it has better fidelity than most other MEMS accelerometers.

Figure 2.8 shows the response of several MEMS accelerometers. Notice that the noise level of most current MEMS devices are significantly higher than that for a current state-of-the-art accelerometer (a Kinemetrics Episensor with a dynamic range of 140 dB). The Caltech Community Seismic Network project is another development project to deploy MEMS accelerometers to volunteers. In this case the accelerometer is a Phidget MEMS accelerometer with a 2 g clip level. The total dynamic range of the current device is in the range of 80 dB. Currently, the cost of the Android phone accelerometer is less than $1, the Phidget is less than $100, and the Episensor is in the range of $4,000.

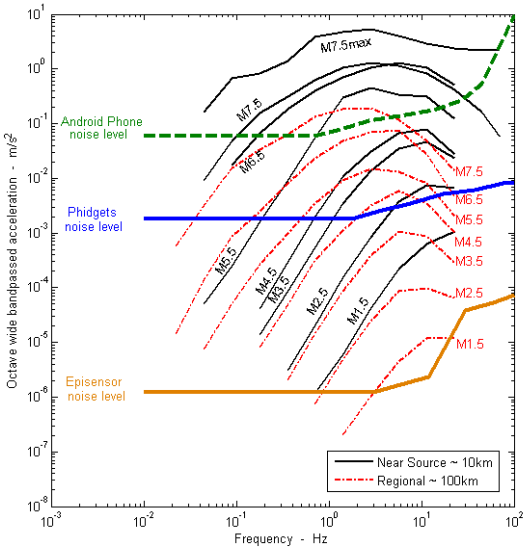


Figure 2.8 Shows the operating range of several MEMS accelerometers as compared with a Kinemetrics Episensor (force balance accelerometer).

Deriving Ground Motion from Seismograms

Chapter 1 provides the basic theory of an SDOF oscillator and deconvolution. While this is straightforward in principal, it is anything but simple in practice. Most seismologists use an excellent signal processing package for UNIX machines that is available from Lawrence Livermore National Lab called SAC (http://www.llnl.gov/sac/). This package has many routines to remove instrument responses, filter, differentiate, integrate, and baseline correct. There are some issues to keep in mind when processing data. First, consider that we typically have three seismometers to record three linear components of motion plus three components of rotation of the ground. In general, seismometers are not directly sensitive to rotation. However, because they sit on the surface of the Earth in the presence of gravity, rotation does have an effect as follows. If $\theta(t)$ is maximum tilt of the site, then

$$\begin{aligned}\theta(t) &= \arctan\sqrt{\left(\frac{\partial u_z(t)}{\partial x}\right)^2+\left(\frac{\partial u_z(t)}{\partial y}\right)^2} \\ &\approx \sqrt{\left(\frac{\partial u_z(t)}{\partial x}\right)^2+\left(\frac{\partial u_z(t)}{\partial y}\right)^2} \quad \text{for small } \theta\end{aligned} \tag{2.29}$$

where *x* and *y* are horizontal Cartesian coordinates of the seismometer and *z* is the vertical component. In most of this text, I simply equate the acceleration of the base of a seismometer with the particle acceleration $\ddot{u}(t)$ of the ground on which it sits. However, if the ground tilts there is an additional acceleration on the instrument caused by changes in the resolved gravitational force on the instrument and we need to be more precise in our definition of the acceleration experienced by the seismometer. We can write the acceleration ***A(t)*** that the seismometer experiences as

$$\begin{aligned}A_z(t) &= \ddot{u}_z(t) - g\cos\theta(t) \\ &\approx \ddot{u}_z(t) - g \quad \text{for small}\theta.\end{aligned} \tag{2.30}$$

and

$$\begin{aligned}A_x(t) &= \ddot{u}_x(t) + g\sin\theta(t) \\ &\approx \ddot{u}_x(t) + g\theta(t) \quad \text{for small } \theta\end{aligned} \tag{2.31}$$

Therefore, vertical-component seismic records are insensitive to small tilts. This is not generally true for the horizontal component seismographs. Fortunately, in most cases the effect of the tilt is small. However, there are

cases when the tilt is important. Tilt can be considered to be the sum of rotations about a horizontal axis due to both elastic strain and rigid body rotations. Tilts can be caused by both traveling elastic waves and also by static (or quasi-static) tilts of the ground surface. Tilts can occur when strong shaking causes compaction of soils.

As we will see in Chapter 3, the strains associated with traveling elastic waves are proportional to the ratio of particle velocity divided by wave velocity. Therefore, we can generally state that

$$\theta(t) \sim B\frac{\dot{u}(t)}{c} \tag{2.32}$$

where B is a constant that depends on the many details of an individual problem and c is a wave speed. Therefore, for traveling waves, we can rewrite (2.31) using (2.32) as

$$A_x(t) \sim \ddot{u}_x(t) + B\frac{g}{c}\dot{u}_x(t) \tag{2.33}$$

The fact that the effect of tilt is proportional to particle velocity as opposed to acceleration means that tilts generally become more important for lower frequency waves. That is (2.33) can be written in the frequency domain as

$$\tilde{A}_x(\omega) \sim \omega^2\tilde{u}_x(\omega)\left(1 - B\frac{g}{c}\frac{i}{\omega}\right) \tag{2.34}$$

The constant B can, in many cases, depend strongly on the local geometry of the seismometer installation. That is, there can be concentrations in strain (i.e., tilt) in corners of rooms. In fact, the relationship between local tilt at a station and the waves passing through are extremely complex and, in most instances, they can only be determined from empirical measurements. In this case, the relationship between Earth strain and local tilt is not a single constant, but is itself a tensor quantity. This unfortunate fact means that it is extremely difficult to determine true horizontal particle motion for long-period seismic waves (see King, ??, for more discussion of this problem). This ambiguity could be resolved if rotations could be independently measured at a site. Unfortunately, the measurement resolution of instruments to measure rotation has not been sufficient to be useful for removing the effects of tilt from long-period seismograms. Fortunately, the vertical particle motions are not affected by this problem.

As a particularly simple example, consider the case of a harmonic Rayleigh wave (we will discuss them in more detail in Chapter 5) with wavenumber k

and frequency ω traveling at velocity c in the x direction. The motion of this wave can be described as

$$u_x(x,t) = a_x \cos(kx - \omega t) = a_x \cos\left[\omega\left(t - \frac{x}{c}\right)\right] \tag{2.35}$$

$$u_z(x,t) = a_z \sin(kx - \omega t) = -a_z \sin\left[\omega\left(t - \frac{x}{c}\right)\right] \tag{2.36}$$

where

$$c = \frac{\omega}{k} \tag{2.37}$$

$$\begin{aligned} A_x(t) &\approx \frac{\partial^2 u_x(kx-\omega t)}{\partial t^2} + g\frac{\partial u_z(kx-\omega t)}{\partial x} \\ &= -\omega^2 \cos(kx-\omega t) - gk\cos(kx-\omega t) \\ &= -\omega^2\left[1 - \frac{g}{c\omega}\right]\cos(kx-\omega t) \end{aligned} \tag{2.38}$$

Thus, for a given wave velocity, the tilt term becomes large with respect to the linear acceleration term when the frequency becomes small. If we assume that $c \approx 3.3\ {}^{\text{km}}\!/\!_{\text{s}}$ and $g \approx 10\text{m}/\text{s}^2$, then (2.38) becomes

$$\begin{aligned} A_x &\approx -\left(1 - \frac{3\cdot 10^{-3}}{\omega}\right)\omega^2 \cos(kx-\omega t) \\ &\approx -\left(1 - 5\cdot 10^{-4} T\right)\omega^2 \cos(kx-\omega t) \end{aligned} \tag{2.39}$$

where T is the period of the wave. That is, the size of the tilt effect is about 10% for a 200-second Rayleigh wave. In actuality, the tilts on a seismometer are very complex since they are really a measure of the local strain at the base of the seismometer. These strains can be strongly affected by the geometry of the seismic recording station. That is, the corners of rooms may cause concentrated strains that are several times larger than the average strain in the earth for the traveling wave that is being considered.

Permanent static tilts can also be caused by other factors, such as differential soil compaction, land sliding, or being next to a fault scarp. In these cases, it is impossible to independently determine both the ground displacement and the ground rotation from just the traces of a seismometer. However, it there is a sudden static change in tilt, it will show up as a step function in the baseline of a horizontal accelerometer. If one assumes a particular function of time over which the static tilt occurs, then it can be removed from the record. However, this usually involves many ad hoc assumptions in practice.

Even if we knew the tilting of the ground and the response of the instrument, there are still difficulties in recovering the true ground displacement. Consider the case of an SDOF in which the seismogram *x(t)* is known (to the resolution of the instrument and the digitizer). We could recover the ground motion U(t) by either deconvolution (discussed in Chapter 1), or by direct integration of the equation of motion (1.2) as follows.

$$\begin{aligned} u(t) &= u_0 + \dot{u}_0 t - \int_0^t \int_0^t \left(\ddot{x} + 2\beta\dot{x} + \omega_0^2 x \right) dt dt \\ &= (u_0 - x_0) + (\dot{u}_0 - \dot{x}_0)t - x - 2\beta\left(\dot{x}_0 + \int_0^t x dt \right) - \omega_0^2 \int_0^t \int_0^t x dt^2 \end{aligned} \tag{2.40}$$

While the implementation of the integrals in (2.40) may seem simple, there are some difficult issues. In particular, what time should we consider zero time to be, and what is the initial velocity, $\dot{u}_0$? Unfortunately, many important strong motions were recorded on analog triggered instruments; there is no recording for the time period prior to the triggering of the instrument. Therefore, the initial velocity is unknown, which may have an important effect on the record.

Another important problem is simply that of obtaining a good record of *x(t)*. In particular, there is often some constant baseline that is superimposed on the initial part of *x(t)*; this is usually called the bias which I will call E_0. Let us further suppose that we have some digital record from our seismograph *y(t)* which is actually composed of the true motion of the seismograph *x(t)* plus some polynomial function of time *E(t)* that represents the bias and other sources of long-period error. That is, assume that

$$y(t) = x(t) + E(t) \tag{2.41}$$

where

$$E(t) = E_0 + E_1 t + E_2 t^2 + \ldots \tag{2.42}$$

Now if we mistakenly substitute y(t) for x(t) in (2.40) (what other choice do we have?), then we derive a flawed ground displacement $u_F(t)$ whose difference from *u(t)* is given by (after some algebra)

$$\begin{aligned} u - u_F &= 2(E_0 + \beta E_1) + (2\beta E_0 + E_1)t + \left(E_1 + \beta E_1 + E_2 + \frac{\omega_0^2}{2} E_0 \right) t^2 \\ &+ \left(E_2 + \frac{2}{3} E_2 + \frac{\omega_0^2}{3} E_1 \right) t^3 + \frac{E_2}{4} t^4 \end{aligned} \tag{2.43}$$

One can see that having an error in the baseline value E_0 causes an error in the displacement that grows as the square of time. If there are further problems in the digital data, such as linear trends, then we can end up with errors that grow as the cube of time. These problems were especially serious for digitization of paper or film records. In these cases, the baseline of the record was often assumed to be the average of the record through time. Unfortunately, the average of the record depends on the time interval that is being averaged. Furthermore, there is no satisfactory way to ensure that there are no linear or quadratic trends in the records. These trends can occur if the film or paper in the recording device is allowed to skew slightly.

SMA-1 film records contain additional null traces, called fixed traces, that are the record described by a rigidly mounted mirror. Trends in these fixed traces are assumed to be the result of wander of the photographic film on the sprockets that transport the film. The trends observed in the fixed traces is subtracted from the live traces in order to minimize baseline shifts.

Because of these problems with trends in the baseline that can cause large errors in displacement that grow large with time, it has been common to subtract best fitting polynomials from records at various stages of processing. Unfortunately, this also introduces new problems. In particular, the process of removing best-fitting polynomial baselines (of any order) is a nonlinear operator. That is, summing two baseline-corrected ground motions does not give the same result as baseline correcting the sum of the motions. Furthermore, when a baseline correction is applied to certain types of true ground motions, it may result in very misleading conclusions about ground

motion. For example, ground displacements near fault scarps often have a static displacement.

Consider the ground motion shown in Figure 2.6. This motion consists of a monotonically increasing displacement up to a new value. The corresponding velocity and accelerations are shown. Acceleration consists of a period of constant positive acceleration followed by an identical negative acceleration. However, if a best-fitting linear baseline is removed from the acceleration record, then we obtain a "corrected" acceleration that is quite different from the true acceleration. Integration of this baseline corrected acceleration will result in a displacement that is quite different from the true displacement.

A real example of this problem is shown in Figure 2.6 from a study by Iwan, Moser, and Peng (BSSA, 1985, 1225-1246). They digitally recorded a Kinemetrics FBA-13 feed-back accelerographs placed on a moving platform. The platform was moved 25 cm and the resulting accelerogram is shown. If the accelerogram is simply integrated into velocity and displacement, then the resulting motions are close to the known input. However, if baselines are removed, then the resulting motion does not look much like the input.

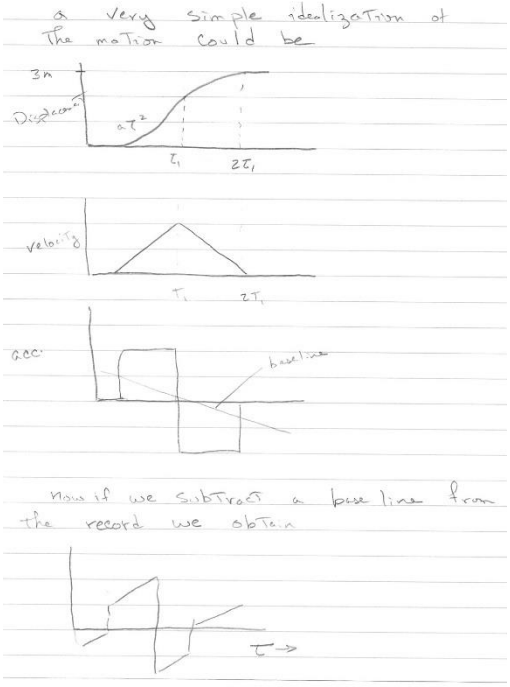


Figure 2.5. Idealized example of how baseline corrections can distort acceleration records for records with net displacements.

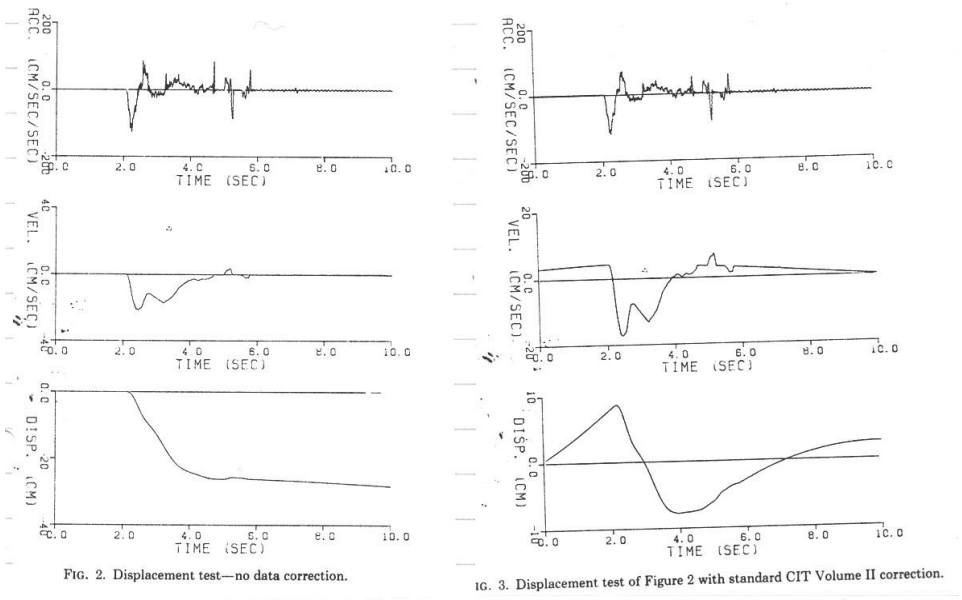


FIG. 2. Displacement test—no data correction.

IG. 3. Displacement test of Figure 2 with standard CIT Volume II correction.

Figure 2.6. From Iwan, Moser, and Chen. Integration of raw displacement records are shown on the left and the effect of baseline correction is shown on the right. The instrument was actually moved to a static displacement of 25 cm.

Records that have no processing, or which only have a bias and the trend of a fixed trace removed are sometimes referred to as Volume I records. This alludes to an important project at Caltech in the 1970's to provide standard processing of most of the known strong-motion records. Records that were corrected (e.g., baselines, initial velocities, and bandpass Ormsby filters) were referred to as Volume II records (everything was published in CIT reports). This processing is described by Trifunac and Lee (Routine Computer Processing of Strong-motion Accelerographs, Earthquake Engineering Research Laboratory Report 73-03, 1973, Pasadena, CA).

An example of how static ground displacement can be recovered is from a study of the 1985 M 8.2 Michoacan, Mexico, earthquake (Anderson and others, 1986, Science, v. 233, 1043-1049). Figure 2.7 shows the locations of strong-motion stations on the Mexican west coast. It also shows the surface projection of the rupture surfaces for several important earthquakes including the Michoacan earthquake that is labeled 19 Sept. 1985. The accelerograms from the four closest digital fba stations (three of which are directly above the rupture) are also shown. The surface projection of the place where rupture originated (called the epicenter) is located near the station Caleta de Campos. It was in this vicinity that the ground first began to shake. It took

tens of seconds for the rupture to propagate throughout the fault surface and stations to the south began to shake at later times; Zihuantanejo did not shake hard until 40 seconds after strong shaking at Caleta de Campos.

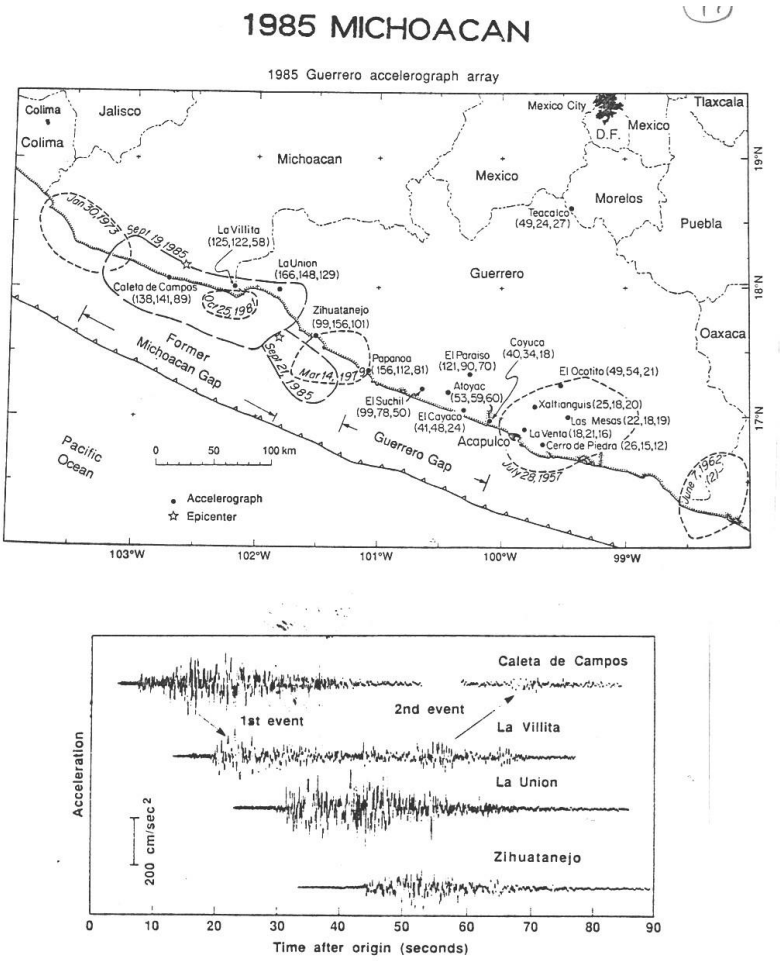


Figure 2.7 North-south component of the ground acceleration for stations above the aftershock zone. The vertical separation of the records is proportional to the NW-SE distance of the station location (along trench distance). Time is measured with respect to the origin time of the earthquake (from Anderson and others, 1986).

Caleta de Campos records that were integrated into velocity and displacement are shown in Figure 2.8.

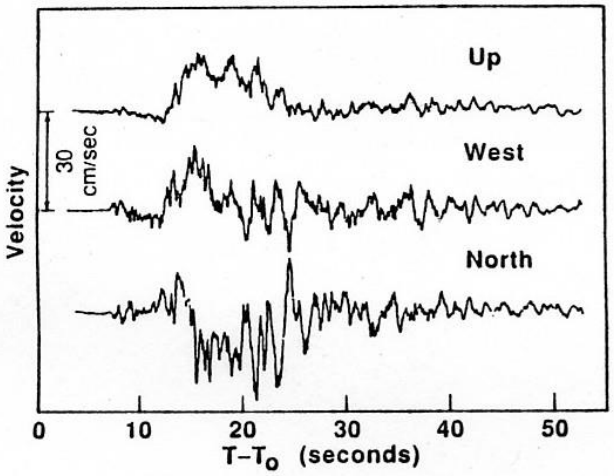


Fig. 5. Ground velocity at Caleta de Campos during the 19 September earthquake derived from accelerograms.

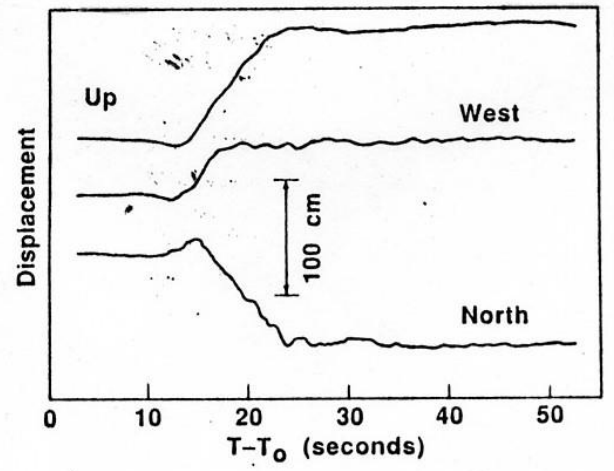


Fig. 6. Ground displacement at Caleta de Campos during the 19 September earthquake derived from accelerograms.

Figure 2.8 (from Anderson and others, 1986)

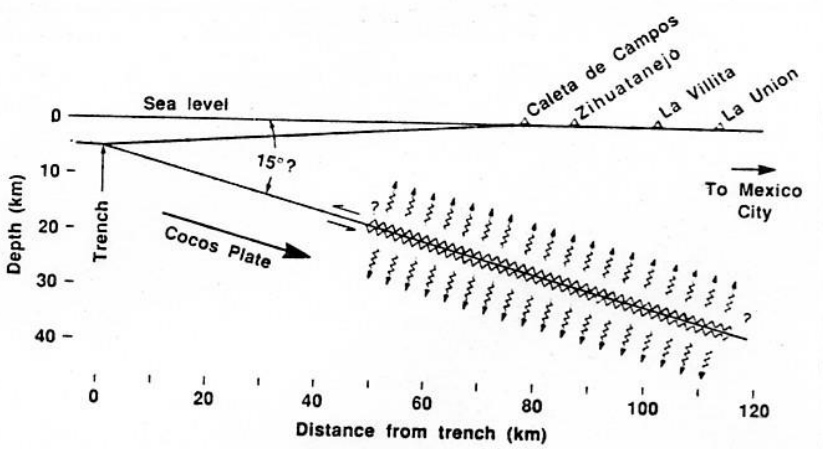


Fig. 1. Cross section of subduction of the Cocos Plate beneath the Mexican mainland. The illustrated extent of faulting corresponds to the aftershock zone in Fig. 3.

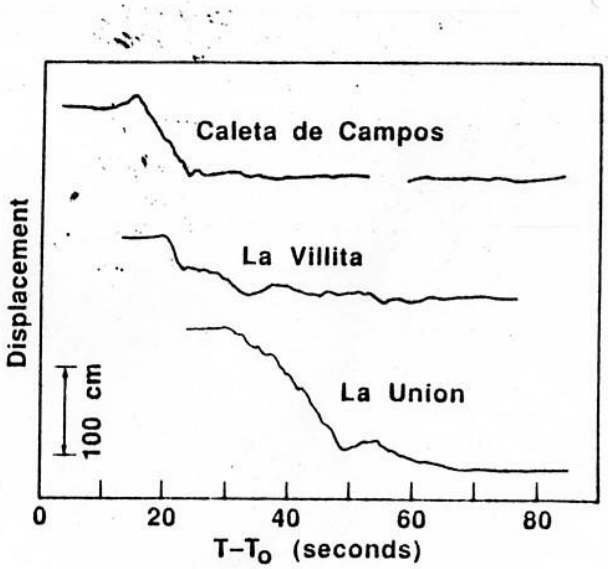


Fig. 7. North component of displacement at three stations above the aftershock zone.

Figure 2.9 (from Anderson and others, 1986)

Permanent displacements of about 1 meter can be seen in the displacement records from Caleta de Campos. Fortuitously, Caleta de Campos is next to the sea shore. The shoreline at this location was permanently uplifted about 1 meter (as derived from killed sea animals such as barnacles), which is in good agreement with the integrated vertical record.

Figure 2.9 shows a cross section view (perpendicular to the oceanic trench) that shows the approximate location of the thrust fault beneath the coast and the relative locations of the strong motion stations. The motion on the thrust fault caused the stations to move upward and towards the ocean. The north components of the derived displacement from the three stations above the faulting are also shown.

As another interesting example of problems with recovering ground displacement, consider the Lucerne station records (station LUC) from the 1992 M 7.2 Landers earthquake. These were recorded on a digital tape system (Kinemetrics SMA-2) which is not widely deployed. The ground velocities recorded for this earthquake are shown in Figure 2.10 (from Wald, Heaton, and Hudnut, 1994, BSSA). The solid line is the surface trace of the faulting and the star is the epicentral location. The recording occurred about 1 km from the fault trace which experienced about 5 meters of strike-slip surface rupture. This means that the east side of the fault moved about 2.5 meters to the south and the west side moved 2.5 meters to the north.

Standard processing was applied to the Lucerne records, and the acceleration, velocity and displacement are shown in Fig. 2.11 and 2.12 for the two horizontal components. Notice that the maximum velocity and displacement that is indicated from these records are only 49 cm/sec and 9 cm, respectively. The displacement is unreasonably small compared to the size of the nearby fault offset. The maximum acceleration, 0.85 g (830 cm/sec) is actually quite large, however.

Iwan and Chen carefully reanalyzed these records; they actually tested the instrument in the lab to see what motions would best reproduce the recordings of the instrument. These motions are shown in Figure 2.13. Notice that the maximum velocity and displacement have increased to 143 cm/sec and 255 cm, respectively. The 255 cm displacement is similar to numbers derived from resurveys of Global Positioning Satellite geodesy network stations in this region.

Figure 2.14 shows the motions of Iwan and Chen after they have been convolved with a 14-second high-pass Butterworth filter. Since high-pass filters do have no response at very long periods, they always remove static offsets from Displacement records. Most strong motion data has been

processed (i.e. filtered) in some way. It is important to understand the processing in order to interpret ground displacement.

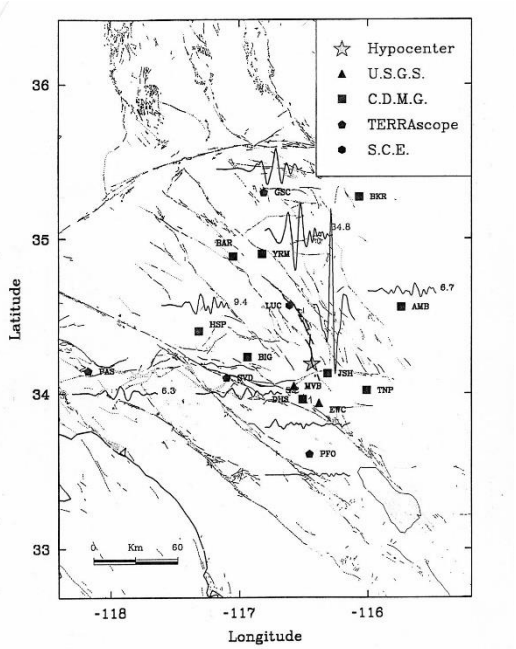


Figure 2.10. Ground velocity records from the 1992 M 7.2 Landers earthquake (from Wald, Heaton, and Hudnut, 1994).

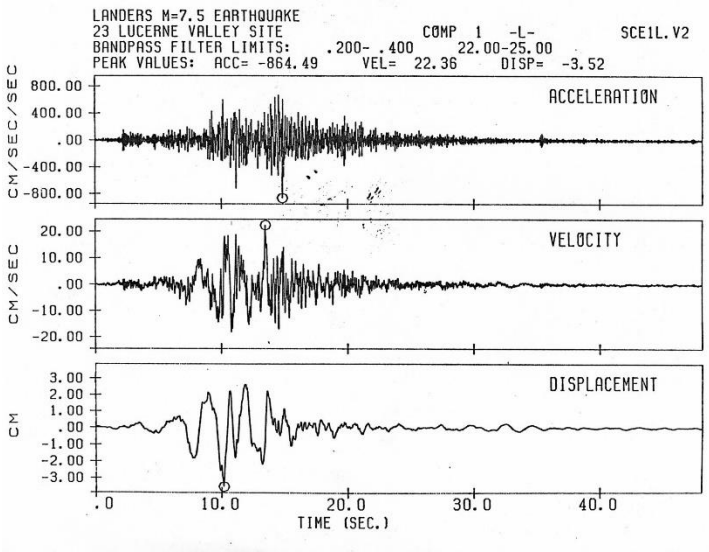


Figure 2.11. Longitudinal ground motions at LUC where the records have a strong band-pass filter to only allow frequencies between 0.4 Hz and 22 Hz.

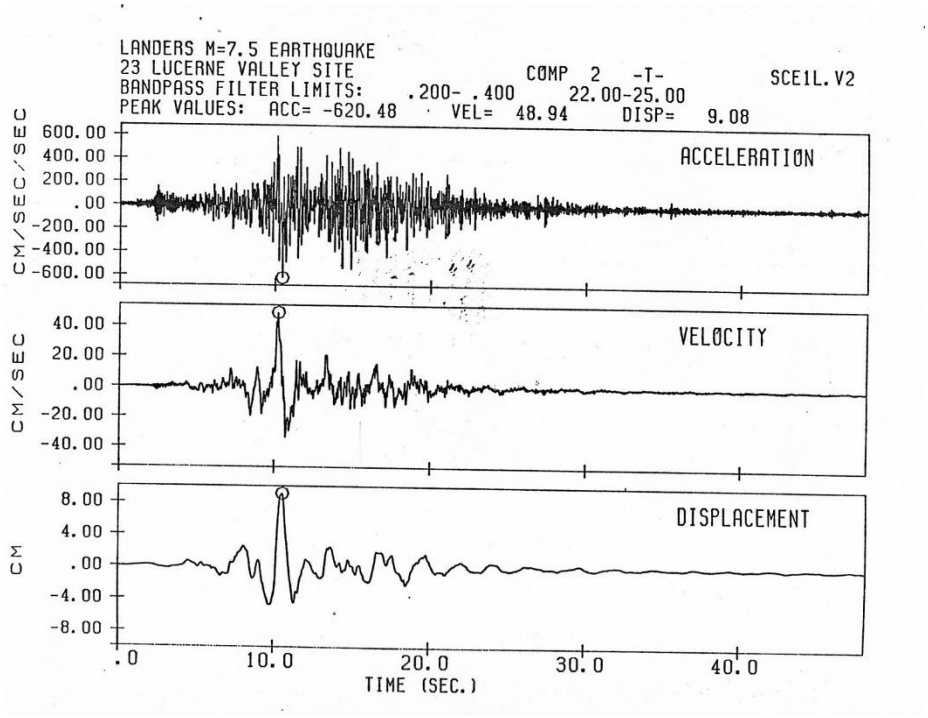


Fig. 2.12 Same as 2.11 except for the transverse component.

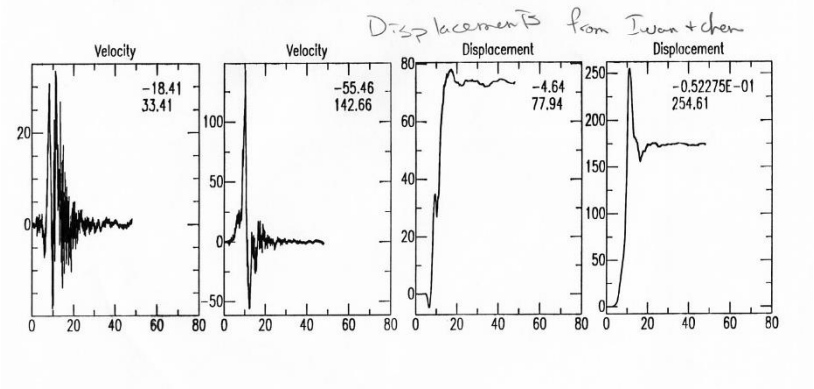


Figure 2.13. Ground velocities and displacements for horizontal components of LUC derived by Iwan and Chen. The top numbers are the peak values in inches or inches/sec, and the bottom numbers are in cm or cm/sec.

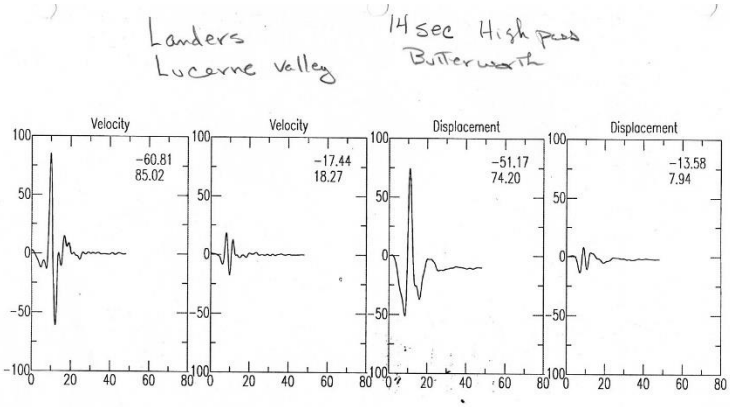


Figure 2.14. Same as 2.13, except that a 14-sec high-pass Butterworth filter has been applied.

As a final example of some of the issues involved with recovering ground displacement from acceleration records, consider the case of the 1999 M 7.6 Chi Chi, Taiwan earthquake. The locations of stations relative to the fault scarp of this east dipping thrust fault are shown in Figure 2.15 (from Boore, D., 2001, Effect of baseline corrections on displacement and response spectra from several recordings of the 1999 Chi-Chi, Taiwan, earthquake, Bull. Seism. Soc. Am., 91, 1199-1210).

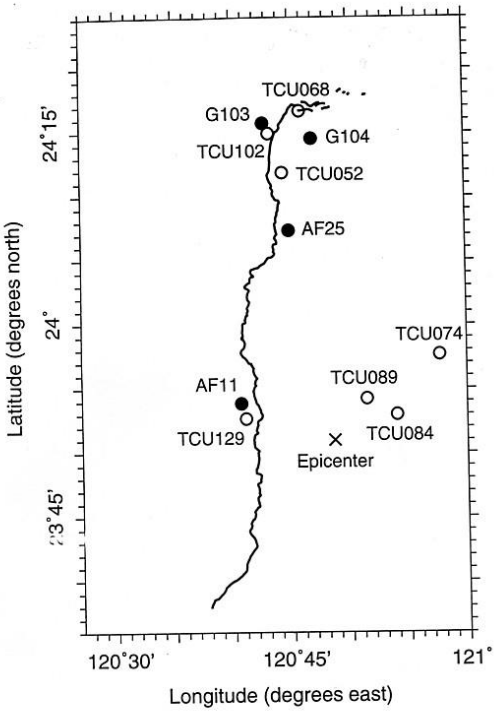


Figure 1. Map showing surface expression of fault (jagged line), strong-motion stations (open circles), and GPS stations (filled circles) that provided data used in this article. The coordinates of the GPS stations are approximate; they were scaled from a figure from the Central Geological Survey (1999).

Figure 2.15 from Boore (2001)

Ground motions were digitally recorded by force-balance accelerometers and the net change in ground displacement was also geodetically recorded by the GPS sites shown in Figure 2.15. In some cases, it was possible to simply doubly integrate the acceleration (after removing a bias) to obtain displacements that were compatible with nearby GPS observations as shown in Figure 2.16.

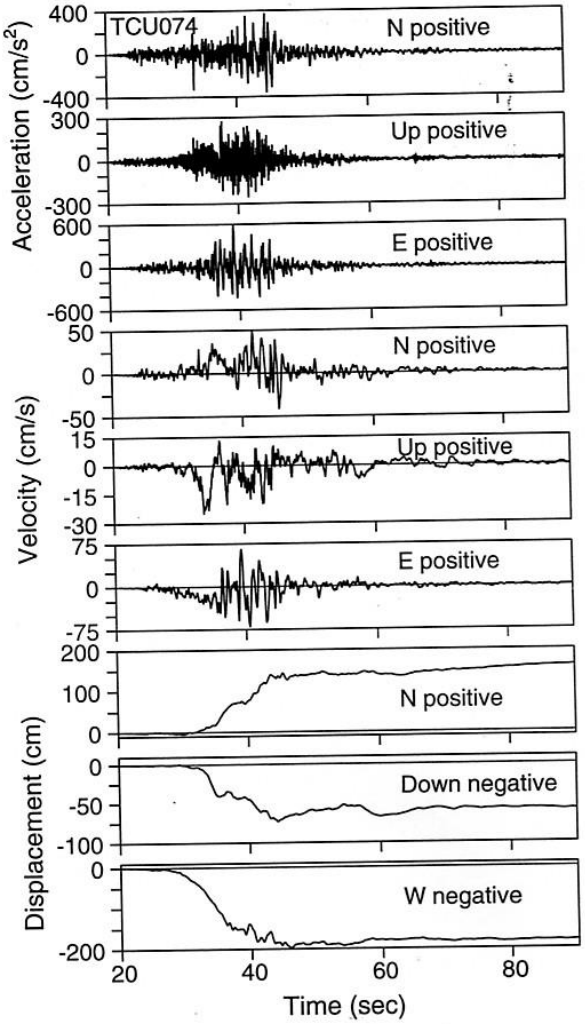


Figure 2. Three components of motion recorded at station TCU074. The top three traces are accelerations for which the pre-event mean was removed from the whole record. These are followed by the velocities and displacements obtained by integrating the accelerations. Note that there is no evidence of a baseline offset.

Figure 2.16 . From Boore (2001).

In other cases, such as that shown in Figure 2.17, removal of the bias was not adequate to obtain a stable ground displacement. Perhaps the site tilted, or perhaps there was some problem with the instrument. In any case, additional

assumptions were necessary in order to derive a reasonable displacement history.

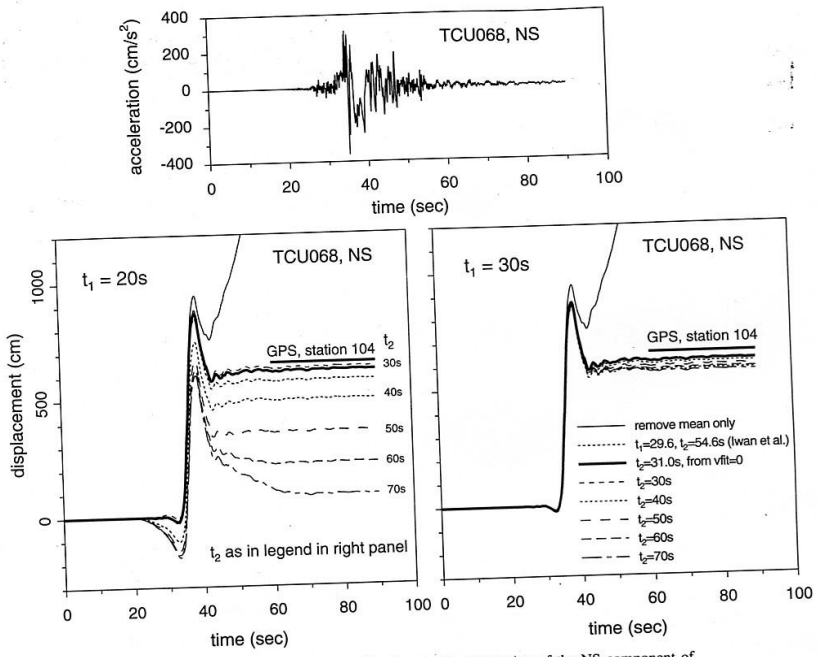


Figure 7. Displacements obtained by double integration of the NS component of acceleration recorded at TCU068 (shown in top panel) and modified using a variety of baseline corrections. The left figure used $t_1 = 20$ sec, while the right figure used $t_1 = 30$ sec. The GPS level was obtained at a station 4.3 km from TCU068, on the hanging-wall side of the fault.

Figure 2.17 from Boore (2001).

Fortunately, the problem of integrating records is mitigated by modern digital instruments that have pre-event memories, force-balance seismometers, and high dynamic range instruments. Nevertheless, it is often a good idea to obtain copies of raw digital records and to then integrate them yourself. Try to understand the source of long period signals so you can decide what to remove from the records.

High-rate GPS

Global Positioning Satellite geodesy has been steadily improving over the past several decades to the point that it can now be used to describe long-period ground motions in large earthquakes. The current accuracy of displacement recording is several mm and numerous stations in the western U.S. and Japan record continuously at 1 sps (or sometimes as high as 20 sps).

Figure 2.18 (top) is from Jing Yang's PhD thesis and shows the amplitude spectra of both a 1 sps GPS station and a 24-bit fba. Notice that these two spectra cross each other near 0.1 Hz; the GPS is quieter than the accelerometer for lower frequencies, while the accelerometer is better for high frequencies. In the bottom panel, the spectra of two nearby recordings of the 2003 M 8.2 Tokachi-Oki earthquake are shown. Notice that in the frequency band from 0.25 Hz to 0.03 Hz, the accelerometer and GPS records are virtually identical. The two signals can be combined in the frequency domain such that the high frequencies are from the accelerometer and the low frequencies are from the GPS. Figure 2.19 shows an example of a very broad band record that was produced by combining the accelerometer and GPS records. While this type of analysis is rare today, it eliminates the problems caused by tilting accelerometers and it's sure to become more common in the future.

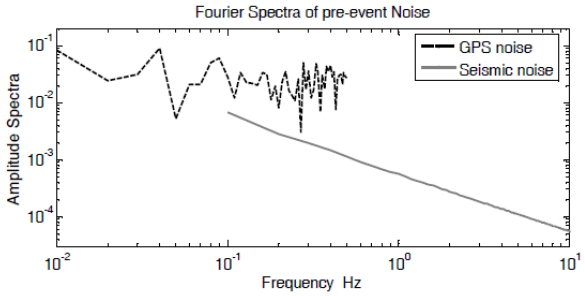


Figure 7.2. Amplitude spectra of east-west preevent noise at GPS station 0124 and K-Net station HKD084 recordings. 100 and 10 seconds preevent signals are chosen for 0124 and HKD084 respectively.

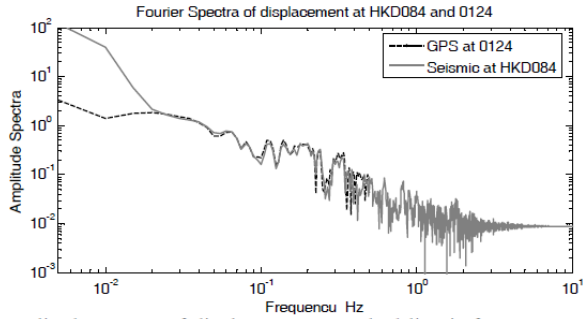


Figure 7.3. Amplitude spectra of displacements. Dashed line is for 200 seconds east-west data recorded at GPS receiver 0124. Solid grey line is for 200 seconds double integration of east-west accelerograms at K-Net HKD084. All the displacements time series are tapered by Chebyshev window at 80 sec to 200 sec to make sure that the signals are periodic function for Fourier transformation.

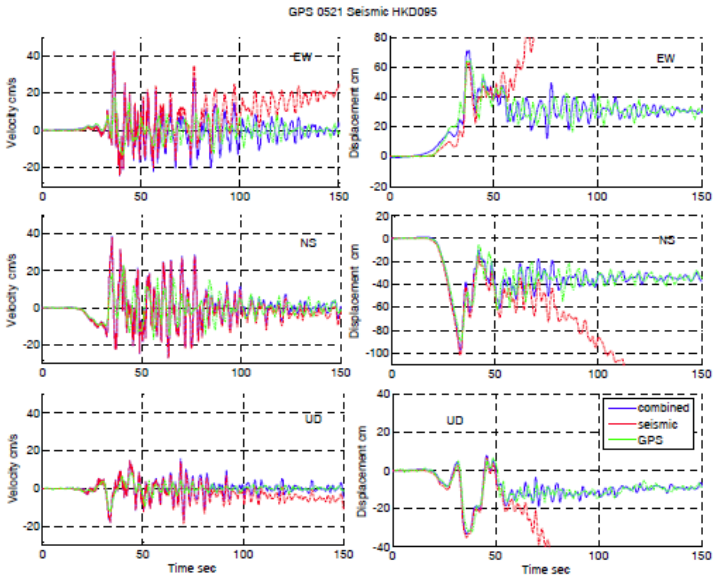


Figure 7.11. Original and corrected velocities (left panel) and displacements (right panel) from GPS station 521 and K-Net station HKD095. The red dashed lines represent seismic time series, the green dash dotted lines represent GPS time series and the solid blue lines are the results processed by our correction scheme.

Another example showing the exceptional ability of GPS stations to record strong ground displacement is shown in Figure 2.20. The GPS data is from the 2010 M8.7 Maule (Chile) earthquake published by Vigny and others. The figure is from Minson, Simons and Heaton (2011 AGU) and it shows the East component displacement. These displacements occurred over tens of seconds and it would be almost impossible to recover these types of records using an accelerometer.

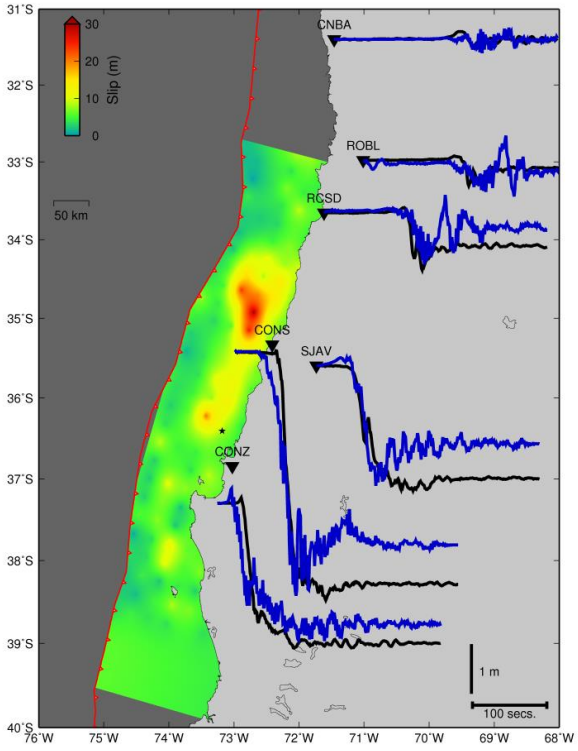


Figure 2.20 East component of 1 sps GPS data for the 2010 M 8.8 Maule earthquake (data from Vigny and others)

Filters

Several different filter types are commonly used in engineering seismology to remove either low-frequency noise (high-pass filter), high-frequency noise (high-pass filter), or both (band-pass filter). In general, filters are designed to have a Fourier amplitude response that is approximately unity in the frequency band that is "passed" by the filter, and some small value in other frequency bands. Before I describe the Fourier amplitude spectra of these filters, I will discuss the phase spectra of the filters. Recall that the phase spectrum is related to the relative amplitude of the cosine and sine series. If $\tilde{F}(\omega)$ is the Fourier transform of some filter function $F(t)$, then $\tilde{F}(\omega)$ is a complex function that can be described by either its real and imaginary parts, or by its amplitude $\left|\tilde{F}(\omega)\right|$ and phase, $\Phi(\omega)$, where

$$\Phi(\omega) \equiv \tan^{-1}\left[\frac{\Im\left(\tilde{F}(\omega)\right)}{\Re\left(\tilde{F}(\omega)\right)}\right] \tag{2.44}$$

The phase function determines how the filter shifts energy with respect to time. Recall the *Shift Theorem* of Fourier transforms.

If $f(t)$ *has the Fourier transform* $\tilde{f}(\omega)$*, then* $f(t-t_0)$ *has the Fourier transform* $e^{-it_0\omega}\tilde{f}(\omega)$.

Since $\left|e^{-it_0\omega}\tilde{f}(\omega)\right| = \left|\tilde{f}(\omega)\right|$, we see that information about shifts in timing are carried by the phase spectrum, and not the amplitude spectrum. Further notice that the shift rule can be written as

$$\Phi\left(e^{-it_0\omega}\right) = \tan^{-1}\left[\frac{\Im\left(e^{-it_0\omega}\right)}{\Re\left(e^{-it_0\omega}\right)}\right] = \tan^{-1}\left[\frac{\sin(-t_0\omega)}{\cos(-t_0\omega)}\right] = t_0\omega \tag{2.45}$$

That is, the phase spectrum associated with a time shift (convolution with $\delta(t-t_0)$) is simply a linear relation between phase and frequency; a positive slope in frequency corresponds to a positive delay in time, and a negative slope corresponds to a negative time delay. This example provides the motivation for a more general understanding of the delay caused by convolution with a filtering function. In general, the phase of a filter is a nonlinear function of the frequency. However, we can define the **group delay** $T_{group}(\omega)$ as

$$T_{group}(\omega) \equiv \frac{d\Phi(\omega)}{d\omega} \tag{2.46}$$

In the case of filtering with a delayed impulse function, the group delay is a constant. That is, all frequencies are delayed by an equal time shift. In the more general case, the main energy is delayed by different amounts at different frequencies. This time of the energy shift is called the group delay.

The principle of **causality** is a statement that effects never precede their causes. That is, if $I(t)$ is the response of a physical system to an impulse applied at $t=0$, then

$$I(t)=0 \quad \text{for all} \quad t<0 \tag{2.47}$$

This is equivalent to saying that **the Group delay must be positive at all frequencies if a filter is causal**. Or alternatively, filters with either zero, or negative, group delays cannot be causal; they create a response before the signal begins. Non-causal filters are a mathematical construct. While it is beyond the scope of this class, it is not too difficult to show that all causal filters $I(t)$ can be written in the Fourier-frequency domain as

$$\tilde{I}(\omega)=G(\omega)+iHT\left[G(\omega)\right] \tag{2.48}$$

Where *HT* signifies **Hilbert transform**, which is defined as

$$\begin{aligned} HT\left[G(\omega)\right] &\equiv \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{G(\omega'-\omega)}{\omega'-\omega}d\omega' \\ &= \frac{-1}{\pi\omega} * G(\omega) \end{aligned} \tag{2.49}$$

Filters generally remove signals in a specified frequency range (usually called a frequency band). For instance, the simplest filters you could think of would consist of a simple cut in frequencies, either above or below a specified frequency. In particular, we could devise a very simple low-pass filter that consists of a rectangle function in the frequency domain. That is, consider convolving with low-pass filter $f_{hp}(t)$ that has a Fourier transform of

$$\tilde{f}_{lp}(\omega)=\Pi(\omega)=\begin{pmatrix} 1 & -\frac{1}{2}<\omega<\frac{1}{2} \\ 0 & \text{elsewhere} \end{pmatrix} \tag{2.50}$$

Since convolution in the time domain is identical to multiplication in the frequency domain, this filter simply eliminates frequencies higher than ½. Since $\tilde{f}_{lp}(\omega)$ is a real number, its phase is zero at all frequencies. That is, it is a non-causal zero-phase filter, which is equivalent to the operation $*\text{sinc}(t)$. Convolution with a sinc function will result in a signal that rings at a frequency of 1. That is, our simple low-pass filter will usually produce a signal that looks as if it is dominated by harmonic waves with a frequency defined by our limiting frequency.

While a filter of the type defined by (2.50) is generally a very poor choice for a filter, it is actually commonly encountered. In particular, there are several important techniques to solve the wave equation only up to a specified frequency. That is, the solution does not contain its high-frequency terms. These solutions are often characterized by a waveform that looks harmonic at the cut-off frequency. If a similar filter is used on data for comparison, then both the data and the synthetic look similar because they both look like harmonic waves at the chosen cut-off frequency.

In a likewise manner, we could construct a simple rectangular high-pass filter $f_{hp}(t)$, where

$$\tilde{f}_{hp}(\omega) = \begin{pmatrix} 0 & -\frac{1}{2} < \omega < \frac{1}{2} \\ 1 & \text{elsewhere} \end{pmatrix} = 1 - \Pi(\omega) \tag{2.51}$$

This filter would be equivalent to the time domain operation $*\left(\delta(t) - \text{sinc}(t)\right)$. That is, the filter is the same as subtracting a low-pass filtered version of the signal from itself. It's easy to see that this also will ring at the cutoff frequency.

Finally, we could construct a band-pass filter $\tilde{f}_{bp}(\omega)$ that is unity between between specified frequencies or

$$\tilde{f}_{bp}(\omega) = \begin{pmatrix} 0 & -\omega_1 < \omega < \omega_1 \\ 1 & \omega_1 < \omega < \omega_2 \\ 0 & \omega > \omega_2 \end{pmatrix} = \Pi\left(\frac{\omega}{\omega_1}\right) - \Pi\left(\frac{\omega}{\omega_2}\right) \tag{2.52}$$

This filter is the same operation as $*\left(\text{sinc}\,\omega_1 t - \text{sinc}\,\omega_2 t\right)$.

Ormsby filters are sometimes encountered in engineering seismology. The Ormsby filter is best described in the Fourier domain as having a trapezoidal shape for the amplitude spectrum and a phase spectrum equal to zero at all frequencies. It is in the class of filters known as **zero-phase filters**. We can write the Ormsby filter $\tilde{O}(\omega)$ as

$$\tilde{O}(\omega) \equiv \begin{Bmatrix} 0 & \omega < \omega_1 \\ \dfrac{\omega - \omega_1}{\omega_2 - \omega_1} & \omega_1 \le \omega \le \omega_2 \\ 1 & \omega_2 \le \omega \le \omega_3 \\ \dfrac{\omega_3 - \omega}{\omega_4 - \omega_3} & \omega_3 \le \omega \le \omega_4 \\ 0 & \omega > \omega_4 \end{Bmatrix} \tag{2.53}$$

This filter is an extension of our simple bandpass filter and it has the same undesirable features. Fortunately, it is rarely used anymore. However, it was heavily used in the 1970's and you should be cautious interpreting records that were filtered with it.

Butterworth filters are far and away the most common type of filter encountered in engineering seismology. They are a type of causal filter that has a frequency amplitude response that is optimally flat. The Fourier amplitude response of an n^{th} order Butterworth is given by

$$\left|\tilde{B}_{lp}(\omega)\right| = \frac{1}{\sqrt{1 + \left(\dfrac{\omega}{\omega_c}\right)^{2n}}} \tag{2.54}$$

This is a minimum-phase filter, which means that the group delay is the minimum possible for the shape of the amplitude spectrum. It is beyond the scope of the class to derive the phase spectrum, but the entire filter is best described by its poles and zeros. Two-pole Butterworth filters are most commonly used in seismology, and ***their response looks very similar to a 71% damped SDOF***. A high-pass Butterworth is formed by taking the reciprocal of frequency or

$$\left|\tilde{B}_{hp}(\omega)\right| = \frac{1}{\sqrt{1 + \left(\dfrac{\omega_c}{\omega}\right)^{2n}}} \tag{2.55}$$

While Butterworth filters are causal filters, they are often applied as zero-phase filters. This is accomplished by filtering twice, once in the forward direction, and once again in the negative direction. Time reversal is the same as taking the complex conjugate in the frequency domain, which is the same

as taking the negative phase. Causal filters are useful if one wants to identify the first arrival of a signal. However, if one is interested in the timing of a main group of energy, then causal filters introduce a group delay. Zero-phase filters eliminate this group delay, but they result in non-causal signals.

Additional Resources

A worldwide database of strong ground motions can be found at

http://db.cosmos-eq.org/scripts/default.plx

A worldwide seismic database of seismic data is available from IRIS (Incorporated Research Institutes for Seismology). Iris also maintains a software download site for analysis of seismic data. Many researchers use SAC to analyze seismic data.

http://www.iris.edu/

Data from the California Integrated Seismic Network (Caltech, UC Berkeley, USGS, Calif. Geol. Survey) can be found at

http://www.cisn.org/

Many researchers find it helpful to plot map data using GMT

http://gmt.soest.hawaii.edu/

GPS data and resources can be found at UNAVCO

http://www.unavco.org/

NGA-PEER processing

In 2008. The Pacific Earthquake Engineering Research Center (PEER) published a set of **ground motion prediction equations (gmpe's)** to describe the statistical properties of ground shaking as a function of earthquake magnitude, distance to the rupture, and characteristics of sites (shear-wave velocity near the site). These gmpe's are referred to as the Next Generation Attenuation (NGA) models. The equations are formulated in log space and they have numerous terms to account for how ground motions vary with magnitude, distance, and site characteristics. There are literally

hundreds of constants that are used cover the plethora of amplitudes that are used to describe the amplitude of shaking. For example, peak ground acceleration (pga), peak ground velocity (pgv). and response spectral acceleration at 5 % damping and at least 10 different natural frequencies (see Chapter 6). The PEER NGA project compiled a large set of strong motion records that can be accessed at Next Generation Attenuation Relationships for Western US (NGA-West) - Databases (berkeley.edu).

This project has been updated to include a large set of strong motion records. Raw, unfiltered acceleration records are available. However, most practitioners use records that are filtered with a zero-phase high-pass Butterworth filter. This simplifies any analysis since it removes long-period trends in the data. Unfortunately, it also means that these filtered records cannot be used to recover true displacements. Most earthquake engineers argue that the bandpass of the filter is wide enough that dynamic analysis of buildings is unaffected by the filtering. Unfortunately, this is not true since simulations of collapse of tall buildings is a highly nonlinear operation. You can find more about this problem in a paper byBuyco, Roh, and Heaton (2020) who concluded that it's generally best to use uncorrected acceleration records for structural analysis.

Homework Chapter 2

Problem 2.1 Explain why an L4-C seismometer that has its case filled with highly viscous oil has an output voltage that is approximately equal to a constant times ground acceleration.

Problem 2.2 Find the poles and zeros of seismograph system that has a 20 sec displacement transducer seismometer (70.7% damped) that is driving a 100 sec galvanometer (also 70.7% damped). Sketch the response.

Chapter 3 Waves in an Elastic Whole Space

8/28/2022

Equation of Motion of a Solid

Hopefully, many of the topics in this chapter are review. However, I find it useful to discuss some of the key characteristics of elastic continuous media. These concepts are critical for understanding both seismic waves in the Earth and also the response of engineered structures (e.g., buildings). I will assume that you already know what stress and strain are and I will begin with the equation of motion. I will use **Einstein's summation convention** that any repeated index signifies summation over three spatial coordinates.

In the first two chapters we considered dynamics problems in which time was the only dependent variable. However, in a continuum, the motion is a function of both time and space. Consider an infinitesimally small cube of elastic solid shown in Figure 3.1. Although this cube is surrounded by a continuous solid, we can ask about the net forces on the cube. This is called **free-body analysis** in engineering mechanics.

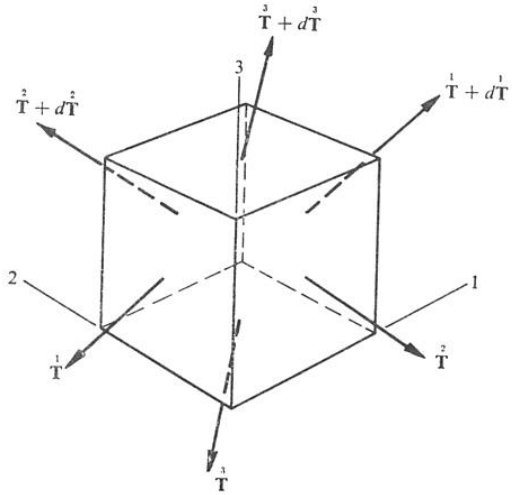


Figure 3.1. Distribution of tractions on the faces of an infinitesimal cube of matter. $\overset{i}{\mathbf{T}}$ is the vector traction (force per unit area) on the i^{th} face of the cube.

We inquire about the net force $\mathbf{F}$ on the cube. We begin by noting that the traction vector on the i^{th} face is given by

$$\overset{i}{\mathbf{T}} = \sigma_{ij}\mathbf{n}_j \tag{3.1}$$

where σ_{ij} is stress in cartesian corrdinates and $\mathbf{n}_j$ is a unit vector normal to the j^{th} face of the cube. For example,

$$\overset{1}{\mathbf{T}} = \sigma_{11}\mathbf{e}_1 + \sigma_{12}\mathbf{e}_2 + \sigma_{13}\mathbf{e}_3 \tag{3.2}$$

We begin by assuming that there is no net torque on the cube, otherwise it would start to spin. This condition is satisfied if and only if the stress tensor is symmetric; that is

$$\sigma_{ij} = \sigma_{ji} \tag{3.3}$$

We next employ Newton's 2nd law to derive the rectilinear acceleration of the mass,

$$\mathbf{F} = \dot{\mathbf{P}} \approx m\ddot{\mathbf{u}} \tag{3.4}$$

The cube is assumed to have a density of ρ and dimensions of dx_1, dx_2, dx_3. The $\approx$ becomes a true $=$ if we take $\mathbf{u}$ to be the position of the center of mass of our infinitesimal cube. The i^{th} component of net force on the cube is

$$F_i = d\overset{1}{T}_i\, dx_2 dx_3 + d\overset{2}{T}_i\, dx_1 dx_3 + d\overset{3}{T}_i\, dx_1 dx_2 \tag{3.5}$$

Recognizing that

$$d\overset{j}{T}_i = \frac{\partial \sigma_{ij}}{\partial x_j} dx_j \quad \text{(no summation)} \tag{3.6}$$

we can rewrite Newton's law (3.4) for the i^{th} component of net force and acceleration as

$$\frac{\partial \sigma_{ij}}{\partial x_j} dx_1 dx_2 dx_3 = \rho \ddot{u}_i dx_1 dx_2 dx_3 \text{ (summation on } j\text{)} \tag{3.7}$$

or using the notation where $,i$ signifies differentiation with respect to the i^{th} coordinate, this can be written

$$\sigma_{ij,j} = \rho \ddot{u}_i \tag{3.8}$$

We can obtain a slightly more general expression by allowing there to be some external "body" force $\boldsymbol{f}$ that is acting on the cube (e.g., gravity) and we then obtain

$$\sigma_{ij,j} + f_i = \rho \ddot{u}_i \tag{3.9}$$

Equation (3.9) is the basic equation of motion of a solid continuum. Although we derived it from Newton's law, it is fundamentally different in that it contains a spatial derivative of forces in addition to the usual time derivative of linear momentum. As we will see, this fundamentally changes the nature of the forces in the problem. It says that acceleration at a point is not related to stress at that point (force per unit area), but to the spatial derivative of stress. We can generalize (3.9) by noting that it can be written as

$$\underset{\rightarrow}{\nabla} \bullet \sigma + \mathbf{f} = \rho \ddot{\mathbf{u}} \tag{3.10}$$

Where $\nabla\bullet$ is the divergence operator (operating on the stress tensor). This operation is a 3-vector whose components are the divergence of the three columns of the stress tensor.

Strain and Constitutive Laws

In order to actually solve elasticity problems, we must have some relationship between the deformation of the body and the internal stresses. If we consider our infinitesimal cube as shown in Figure 3.2, then we can describe the motion of the cube as a combination of a rigid-body motion and rotation and, internal strain. We will keep track of the motions of our cube by characterizing the position $\mathbf{u}$ and the diagonal vector $\mathbf{R}$. We will call the diagonal of the unstrained element $\mathbf{R}$ and the diagonal of the element after straining $\mathbf{R'}$. We define the change in the diagonal element as

$$\delta \mathbf{R} = \mathbf{R'} - \mathbf{R} \tag{3.11}$$

If the motion of the infinitesimal cube is small, then in component form

$$\delta R_i = u_{i,j} dx_j \tag{3.12}$$

which can be rewritten in the form of

$$\delta R_i = \omega_{ij} dx_j + \varepsilon_{ij} dx_j \tag{3.13}$$

where

$$\omega_{ij} = \frac{1}{2}\left(u_{i,j} - u_{j,i}\right) \tag{3.14}$$

and

$$\varepsilon_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) \tag{3.15}$$

ω_{ij} represents rigid body rotation and it is anti-symmetric. ε_{ij} is the infinitesimal strain tensor and it is symmetric.

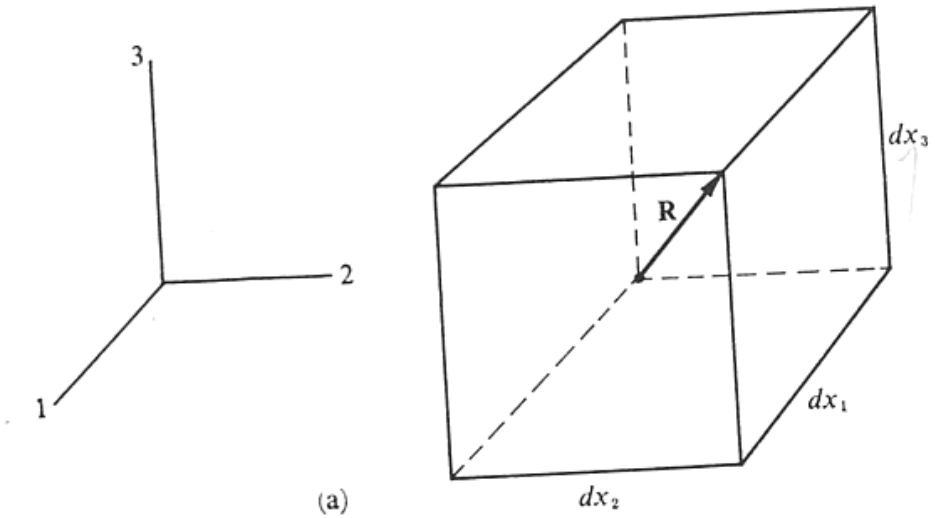


Figure 3.2. Deformation of an infinitesimal element.

The relationship between stress and strain is called the constitutive relation. For small strains, most materials exhibit a linear relationship between stress and strain that can be generally written as

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \tag{3.16}$$

where there are 81 elastic coefficients C_{ijkl}. However, due the symmetry of the stress and strain tensor, and due to the requirement for a unique strain energy, there are at most 21 independent elastic coefficients. If the material is isotropic (no intrinsic directionality to the properties), then there are only 2 independent elastic coefficients. Table 3.1 provides a handy conversion between several different elastic coefficients for an isotropic solid. For our discussion, we will use the 1st and 2nd Lame constants λ and μ. In this case (3.16) simplifies to

$$\sigma_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + 2\mu\varepsilon_{ij} \tag{3.17}$$

where

$$\delta_{ij} = \begin{Bmatrix} 0 & i \neq j \\ 1 & i = j \end{Bmatrix} \equiv \text{Kronecker delta} \tag{3.18}$$

	LAMÉ'S MODULUS λ	SHEAR MODULUS μ	YOUNG'S MODULUS η	POISSON'S RATIO ν	BULK MODULUS κ
λ, μ			$\frac{\mu(3\lambda+2\mu)}{\lambda+\mu}$	$\frac{\lambda}{2(\lambda+\mu)}$	$\frac{3\lambda+2\mu}{3}$
λ, η		irrational		irrational	irrational
λ, ν		$\frac{\lambda(1-2\nu)}{2\nu}$	$\frac{\lambda(1+\nu)(1-2\nu)}{\nu}$		$\frac{\lambda(1+\nu)}{3\nu}$
λ, κ		$\frac{3(\kappa-\lambda)}{2}$	$\frac{9\kappa(\kappa-\lambda)}{3\kappa-\lambda}$	$\frac{\lambda}{3\kappa-\lambda}$	
μ, η	$\frac{(2\mu-\eta)\mu}{\eta-3\mu}$			$\frac{\eta-2\mu}{2\mu}$	$\frac{\mu\eta}{3(3\mu-\eta)}$
μ, ν	$\frac{2\mu\nu}{1-2\nu}$		$2\mu(1+\nu)$		$\frac{2\mu(1+\nu)}{3(1-2\nu)}$
μ, κ	$\frac{3\kappa-2\mu}{3}$		$\frac{9\kappa\mu}{3\kappa+\mu}$	$\frac{3\kappa-2\mu}{2(3\kappa+\mu)}$	
η, ν	$\frac{\nu\eta}{(1+\nu)(1-2\nu)}$	$\frac{\eta}{2(1+\nu)}$			$\frac{\eta}{3(1-2\nu)}$
η, κ	$\frac{3\kappa(3\kappa-\eta)}{9\kappa-\eta}$	$\frac{3\eta\kappa}{9\kappa-\eta}$		$\frac{3\kappa-\eta}{6\kappa}$	
ν, κ	$\frac{3\kappa\nu}{1+\nu}$	$\frac{3\kappa(1-2\nu)}{2(1+\nu)}$	$3\kappa(1-2\nu)$		

Table 3.1. Relationship between elastic constants for an isotropic elastic medium

Navier's Equation

We are now able to write the equation of motion entirely in terms of displacement of the medium. Combining equations (3.9), (3.15), and (3.17), we obtain

$$\rho\ddot{u}_i = f_i + \mu u_{i,jj} + (\lambda+\mu)u_{j,ji} \tag{3.19}$$

This is Navier's equation, and it is such an important equation that it is worth writing it out to see the terms more explicitly.

$$\rho\frac{\partial^2 u_i}{\partial t^2} = f_i + \sum_{j=1}^{3}\left[\mu\frac{\partial^2 u_i}{\partial x_j^2} + (\lambda+\mu)\frac{\partial^2 u_j}{\partial x_j \partial x_i}\right] \tag{3.20}$$

In Navier's equation 2nd derivatives of displacements with respect to time are linearly related to 2nd derivatives of displacement with respect to space.

Everything that happens in an isotropic linearly elastic solid is a solution to this equation.

We can also write Navier's equation in vector form as

$$\mu\nabla^2\mathbf{u}+(\lambda+\mu)\underset{\rightarrow}{\nabla}(\underset{\rightarrow}{\nabla}\bullet\mathbf{u})+\mathbf{f}=\rho\ddot{\mathbf{u}} \tag{3.21}$$

Where Laplacian operator $\nabla^2\mathbf{u}\equiv\underset{\rightarrow}{\nabla}\bullet(\underset{\rightarrow}{\nabla}\mathbf{u})$ is the divergence (a 3-vector) of the gradient of the displacement vector (a 3-tensor). The term $\underset{\rightarrow}{\nabla}\bullet\mathbf{u}$ is seen to be the *dilatation*, or the net volume change of our infinitesimal element. This vector form of the equation has the advantage that we can rewrite it in any type of coordinate frame for which we know the Laplacian operator, the gradient operator, and the acceleration vector. In particular, we can write these operators for

Cartesian coordinates

$$\ddot{\mathbf{u}}=\ddot{u}_i\mathbf{e}_i \tag{3.22}$$

$$\underset{\rightarrow}{\nabla}\cdot\mathbf{u}=u_{i,i} \tag{3.23}$$

$$\underset{\rightarrow}{\nabla}=\mathbf{e}_i\frac{\partial}{\partial x_i} \tag{3.24}$$

$$\nabla^2\mathbf{u}=\frac{\partial^2 u_i}{\partial x_j\partial x_j}\mathbf{e}_i \quad \text{(note the double sum on } i \text{ and } j\text{)} \tag{3.25}$$

$$\underset{\rightarrow}{\nabla}\otimes\mathbf{u}=\left(\frac{\partial u_3}{\partial x_2}-\frac{\partial u_2}{\partial x_3}\right)\mathbf{e}_1+\left(\frac{\partial u_1}{\partial x_3}-\frac{\partial u_3}{\partial x_1}\right)\mathbf{e}_2+\left(\frac{\partial u_2}{\partial x_1}-\frac{\partial u_1}{\partial x_2}\right)\mathbf{e}_3 \tag{3.26}$$

Cylindrical coordinates

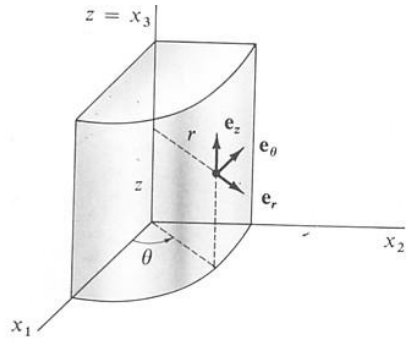
$$\ddot{\mathbf{u}}=\ddot{u}_r\mathbf{e}_r+\ddot{u}_\theta\mathbf{e}_\theta+\ddot{u}_z\mathbf{e}_\mathbf{z} \tag{3.27}$$

$$\underset{\rightarrow}{\nabla}\cdot\mathbf{u}=\frac{1}{r}\frac{\partial}{\partial r}(ru_r)+\frac{1}{r}\frac{\partial u_\theta}{\partial\theta}+\frac{\partial u_z}{\partial z} \tag{3.28}$$

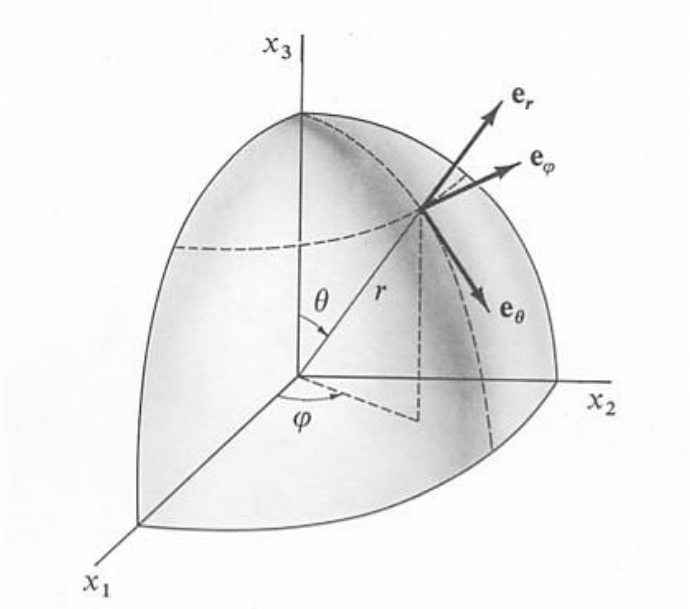
$$\underset{\rightarrow}{\nabla}=\mathbf{e}_r\frac{\partial}{\partial r}+\mathbf{e}_\theta\frac{1}{r}\frac{\partial}{\partial\theta}+\mathbf{e}_z\frac{\partial}{\partial z} \tag{3.29}$$

$$\nabla^2=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2} \tag{3.30}$$

$$\underset{\rightarrow}{\nabla}\otimes\mathbf{u}=\left(\frac{1}{r}\frac{\partial u_z}{\partial\theta}-\frac{\partial u_\theta}{\partial z}\right)\mathbf{e}_r+\left(\frac{\partial u_r}{\partial z}-\frac{\partial u_z}{\partial r}\right)\mathbf{e}_\theta+\left[\frac{1}{r}\frac{\partial}{\partial r}(ru_\theta)-\frac{1}{r}\frac{\partial u_r}{\partial\theta}\right]\mathbf{e}_z \tag{3.31}$$



Spherical coordinates



$$\ddot{\mathbf{u}} = \ddot{u}_r \mathbf{e}_r + \ddot{u}_\theta \mathbf{e}_\theta + \ddot{u}_\varphi \mathbf{e}_\varphi \tag{3.32}$$

$$\underset{\rightarrow}{\nabla} \cdot \mathbf{u} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 u_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}\left(u_\theta \sin\theta\right) + \frac{1}{r\sin\theta}\frac{\partial u_\varphi}{\partial \varphi} \tag{3.33}$$

$$\underset{\rightarrow}{\nabla} = \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_\theta \frac{1}{r}\frac{\partial}{\partial \theta} + \mathbf{e}_z \frac{1}{r\sin\theta}\frac{\partial}{\partial z} \tag{3.34}$$

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial \varphi^2} \tag{3.35}$$

$$\underset{\rightarrow}{\nabla}\otimes\mathbf{u}=\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}\left(u_\varphi\sin\theta\right)-\frac{\partial u_\theta}{\partial\varphi}\right]\mathbf{e}_r+\frac{1}{r\sin\theta}\left[\frac{\partial u_r}{\partial\varphi}-\sin\theta\frac{\partial}{\partial r}\left(ru_\varphi\right)\right]\mathbf{e}_\theta$$
$$+\frac{1}{r}\left[\frac{\partial}{\partial r}\left(ru_\theta\right)-\frac{\partial u_r}{\partial\theta}\right]\mathbf{e}_\varphi \tag{3.36}$$

There are infinitely many solutions to Navier's equation and the solution to any individual problem is the one that has the correct initial conditions and boundary conditions for that particular problem. In general, it is not possible for humans to analytically solve 3.18 for all classes of three-dimensional solutions to (3.20). However, there are a number of analytic solutions to (3.20) if the problem is assumed to be uniform in one direction (two-dimensional). This is ultimately due to the fact that division is defined for two dimensional vectors (the same as division by complex numbers). In contrast, division cannot be defined for higher dimension vectors. Therefore, there are analytic (well mostly analytic) solutions to problems in which the elastic media is described by a stack of horizontal plane layer, but entirely numerical procedures (finite-element or finite-difference) must be used to solve problems in which the structure is truly three dimensional. The techniques for solving general layer problems often rely on expressing the displacement vector field as the sum of potentials (Helmholtz decomposition). That is, we can decompose the displacement as

$$\mathbf{u}=\underset{\rightarrow}{\nabla}\phi+\underset{\rightarrow}{\nabla}\otimes\underset{\rightarrow}{\psi} \tag{3.37}$$

where φ and $\underset{\rightarrow}{\psi}$ are scalar and vector functions of time and space. If we make this change of variables, then Navier's equation separates into several wave equations as follows.

$$\nabla^2\phi=\frac{1}{\alpha^2}\ddot{\phi} \tag{3.38}$$

$$\nabla^2\psi_i=\frac{1}{\beta^2}\ddot{\psi}_i \tag{3.39}$$

Of course, the boundary conditions must also be transformed into potential form. These potential forms can be used in any coordinate system if you know how to compute the Laplacian, the gradient and the curl.

It is beyond the scope of this class to demonstrate general solution techniques for Navier's equation (see Achenbach for a nice treatment), but we can demonstrate several simple solutions which have attributes similar to those of solutions encountered in the real world. **Since Navier's equation is linear, any solution that is added to any other solution is also a solution.** Therefore, we can often build the appropriate solution by adding together known simple solutions in such a way that they produce the desired stresses or displacements on the boundary of

a domain; that is, they match boundary conditions. When a domain contains layers, the solutions apply inside the individual layer, and they are constructed to produce continuous displacement at the boundaries and balanced tractions on the boundaries.

Plane P-waves

Suppose that we consider a motion defined by

$$u_1\left(x_1, x_2, x_3, t\right) = f\left(t - \frac{x_1}{\alpha}\right) \tag{3.40}$$

and

$$u_2 = u_3 = 0 \tag{3.41}$$

then it is a simple matter of substituting (3.40) and (3.41) into (3.20) to show that this is a valid solution for any single-variable function f that is twice differentiable, and provided that

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}} \tag{3.42}$$

We could have alternatively chosen the potentials,

$$\phi = -\alpha \int f\left(\xi\right) d\xi; \qquad \xi \equiv t - \frac{x_1}{c} \tag{3.43}$$

$$\underset{\rightarrow}{\psi} = \mathbf{0} \tag{3.44}$$

It is a trivial matter to show that its gradient is the displacement field given by (3.40) and (3.41), and that it satisfies the wave equations (3.38) and (3.39).

This is the equation of a planar P-wave traveling at velocity α in the positive x_1 direction. Since the material is isotropic, this direction is arbitrary, and it could just as well be traveling in the negative x_1 direction. Note that the shape of the waveform is unchanged as it propagates through the medium. This property is called **nondispersive,** and it contrasts with some other solutions that we will explore later where the wave velocity depends on the frequency of the oscillation.

Since the equation is linear, we could write a more general solution that has different P-waves traveling in both positive and negative directions as

$$u_1 = f\left(t - \frac{x_1}{\alpha}\right) + g\left(t + \frac{x_1}{\alpha}\right) \tag{3.45}$$

where *g* is some other twice differentiable function. **P-waves** are also called **longitudinal waves** since their particle motion is in the same direction as the wave propagates. They are also called **compressional waves,** although they have both compressional and shear stresses as shown by the computing the strain and stress tensor for (3.40) as follows.

$$\begin{aligned}\varepsilon_{11} &= \frac{\partial u_1}{\partial x_1} \\ &= -\frac{f'\left(t-\frac{x_1}{\alpha}\right)}{\alpha} = -\frac{\dot{f}\left(t-\frac{x_1}{\alpha}\right)}{\alpha} \\ &= -\frac{\dot{u}\left(t-\frac{x_1}{\alpha}\right)}{\alpha}\end{aligned} \tag{3.46}$$

and all other strain components are zero. Don't be confused by the f' , it simply means differentiation with respect to the argument, $\left(t-\frac{x_1}{\alpha}\right)$. We see that the **strain in this wave is proportional to the particle velocity divided by the wave speed**. This will be a recurring theme for other solutions of Navier's equation.

We can substitute (3.46) into (3.17) to obtain the stress, which gives

$$\begin{aligned}\sigma_{11} &= \lambda\left(\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33}\right)+2\mu\varepsilon_{11} \\ &= \left(\lambda+2\mu\right)\varepsilon_{11}\end{aligned} \tag{3.47}$$

and

$$\sigma_{22} = \sigma_{33} = \lambda\varepsilon_{11} \tag{3.48}$$

$$\sigma_{12} = \sigma_{23} = \sigma_{13} = 0 \tag{3.49}$$

Substituting (3.42) and(3.46) into (3.47) and (3.48) we find that

$$\sigma_{11} = -\rho\alpha\dot{u} \tag{3.50}$$

and

$$\sigma_{22} = \sigma_{33} = \frac{\lambda}{\lambda+2\mu}\sigma_{11} \tag{3.51}$$

Equation (3.50) tells us that the **stress in this wave is related to the particle velocity times the product of the density and the wave speed.** The ratio of

the stress to the particle velocity $\sigma/\dot{u} = \rho\alpha$ is called the **mechanical impedance**; it measures the stress that is needed to make a particular ground motion. In our particular example,

$$\text{mecahnical impedance} = \frac{\sigma_{11}}{\dot{u}_1} = \rho\alpha = \sqrt{\rho(\lambda + 2\mu)} \tag{3.52}$$

Notice that although there are no explicit shear stresses in this coordinate frame (which is the principal coordinate frame for this problem), there are shear stresses in other coordinate frames. The maximum shear stress is in the frame rotated 45 degrees from the principal frame and in this frame the maximum shear stress is

$$\sigma_{1'2'} = \frac{1}{2}(\sigma_{11} - \sigma_{22}) = \frac{2\mu}{\lambda + 2\mu}\sigma_{11} \tag{3.53}$$

Therefore, there are shear stresses associated with these P-waves.

We can also calculate the **power** $P(x_1, t)$ associated with this wave as the energy flux in the x_1 direction. This energy flux is the rate of work per unit area done by the traction vector on a plane perpendicular to the velocity of propagation. This rate of work (power P) per infinitesimal unit area dS is the stress times the particle velocity, or

$$\frac{P(x_1, t)}{dS} = -\sigma_{11}\dot{u}_1 = \rho\alpha\dot{u}_1^2 \tag{3.54}$$

The energy per unit volume $E(x_1, t)$ associated with the wave is just the energy flux divided by the wave velocity, or

$$\frac{E(x, t)}{dV} = \rho\dot{u}^2 \tag{3.55}$$

As is the case for all linear dynamic systems, this energy is evenly divided between kinetic energy and potential (strain) energy if averaged throughout the system.

Finally, we can inquire about the maximum accelerations that can occur in an elastic continuum. We can differentiate equation (3.50) to obtain

$$\ddot{u}_1(x_1, t) = -\frac{\dot{\sigma}_{11}\left(t - \frac{x_1}{\alpha}\right)}{\rho\alpha} \tag{3.56}$$

That is, the acceleration of a point scales like the time derivative of the compressive stress. If a finite compressive stress is suddenly applied to a surface then it generates a P-wave whose acceleration is described by a Dirac-delta function, which has infinite acceleration. That is, if

$$\sigma_{11} = \sigma_0 H\left(t - \frac{x_1}{\alpha}\right) \tag{3.57}$$

where *H(t)* is a Heaviside step function, then

$$\ddot{u}_i = \frac{\sigma_0}{\rho\alpha}\delta\left(t - \frac{x_1}{\alpha}\right) \tag{3.58}$$

Notice that the acceleration is infinite, whereas the stress is finite.

Plane Shear Waves

Another important solution to Navier's equation can be expressed as

$$u_2 = f\left(t - \frac{x_1}{\beta}\right) \tag{3.59}$$

$$u_1 = u_3 = 0 \tag{3.60}$$

It is a simple matter to substitute (3.59) and (3.60) into Navier's equation (3.20) to find that this is a solution so long as

$$\beta = \sqrt{\frac{\mu}{\rho}} \tag{3.61}$$

As before, we could have used the displacement potentials

$$\phi = 0 \tag{3.62}$$

$$\psi_1 = \psi_2 = 0 \tag{3.63}$$

$$\psi_3 = \beta f\left(t - \frac{x_1}{\beta}\right) \tag{3.64}$$

where the curl of $\vec{\psi}$ is the displacement and (3.64) solves the scalar wave equation
(3.39).

This is the description of a planar shear wave (S-wave) traveling in the positive x_1 direction with velocity β. The particle motion is in the x_2 direction and it is parallel to the wave front and perpendicular to the direction of motion. As was the case with P-waves, *f(t)* is any function with a finite 2nd derivative. Like the planar P-wave, planar S-waves are also nondispersive.

Notice that the S-wave is slower than the P-wave and that the ration of the velocities is

$$\frac{\alpha}{\beta} = \sqrt{\frac{\lambda + 2\mu}{\mu}} \tag{3.65}$$

This can be expressed in terms of Poisson's ratio ν by using Table 3.1. In this case,

$$\frac{\alpha}{\beta} = \sqrt{\frac{2 - 2\nu}{1 - 2\nu}} \tag{3.66}$$

So the ratio of P- to S-wave velocities depends only on Poisson's ratio. For many solids, $\lambda \approx \mu$, or $\nu \approx 1/4$, in which case we call the solid Poissonian and $\alpha/\beta \approx \sqrt{3} = 1.717$. The typical S- and P-wave speeds in the Earth's crust are 4 km/s and 6.5 km/s, respectively. A handy trick is to estimate the distance Δ between an earthquake and a seismic station using the following simple formula

$$\Delta \approx (t_S - t_P) \cdot 7\left(\mathrm{km/s}\right) \tag{3.67}$$

There are important cases where the P-wave speed is much higher than the S-wave speed. In particular, the types of water saturated muds found in coastal areas can have P-wave speeds that are more than 10 times the S-wave speed. In this case Poisson's ratio approaches its upper limit of $\frac{1}{2}$.

We can also compute strain, stress, and energy flux for the S-wave wave as we did for the planar P-wave. In this case,

$$\varepsilon_{12} = -\frac{1}{2}\frac{\dot{u}_2}{\beta} \tag{3.68}$$

$$\varepsilon_{11} = \varepsilon_{22} = \varepsilon_{33} = \varepsilon_{13} = \varepsilon_{23} = 0 \tag{3.69}$$

$$\sigma_{12} = \rho\beta\dot{u}_2 \tag{3.70}$$

$$\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma_{13} = \sigma_{23} = 0 \tag{3.71}$$

$$\frac{P(x_1, t)}{dS} = -\sigma_{12}\dot{u}_2 = \rho\beta\dot{u}_2^2 \tag{3.72}$$

Diagrams of the motion of Planar P- and S-waves are shown in Figure 3.3.

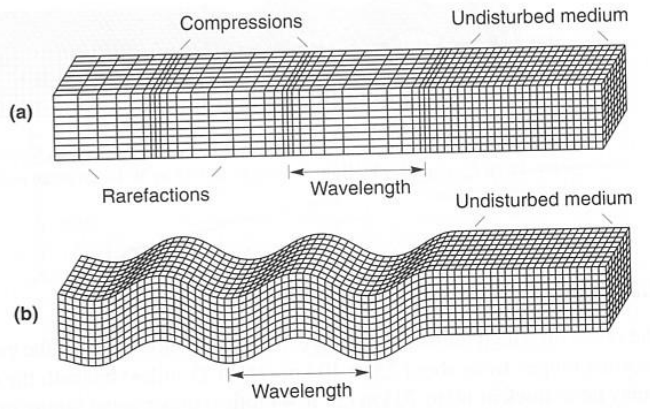


Figure 3.3. a) longitudinal P-wave, b) Transverse S-wave

Harmonic Plane Waves

While planar P- and S-waves can be expressed for any function of the variable, $\left(t-\frac{x}{c}\right)$, where c is the wave velocity, it is instructive to investigate the solution if the function is harmonic, a sinusoid or cosinusoid. That is, there are many instances in which the superposition of harmonic solutions can be used to construct solutions to more general problems. To demonstrate, let's consider the planar S-wave in the previous section, but we will assume that our function is a cosine. That is,

$$\begin{aligned} u_2 &= \cos\left[\omega\left(t-\frac{x_1}{\beta}\right)\right] \\ &= \cos\left(kx_1-\omega t\right) \end{aligned} \tag{3.73}$$

where k is spatial wavenumber given by

$$k=\frac{\omega}{\beta}=\frac{2\pi}{\Lambda} \tag{3.74}$$

and Λ is the wavelength. We can now consider what happens when two harmonic plane waves of identical strength and frequency, but traveling in opposite directions are added together. We can use standard trigonometric identities to easily show that.

$$\begin{aligned} u_2 &= \cos\left(kx_1-\omega t\right)+\cos\left(kx_1+\omega t\right) \\ &= 2\cos\left(kx_1\right)\cos\left(\omega t\right) \end{aligned} \tag{3.75}$$

Equation (3.75) is therefore a **standing wave** with the same frequency and wavenumber as the two traveling waves. Since Navier's equation is linear, and since the waves traveling in each direction are solutions, then their sum (the standing wave) is also a solution of Navier's equation. Obviously, standing wave solutions are natural when identical waves are traveling in opposite directions. This is a common occurrence when harmonic waves are reflected off of an interface. It also happens in our spherical Earth when waves that travel around the Earth in opposite directions meet. In this case the interference makes the free oscillations of the Earth.

In a similar fashion, it is possible to add two harmonic standing waves together to produce a single harmonic traveling wave. Again we can use standard trig identities to show that

$$\begin{aligned} u_2 &= \cos\left(kx_1\right)\cos\left(\omega t\right) + \sin\left(kx_1\right)\sin\left(\omega t\right) \\ &= \cos\left(kx_1 - \omega t\right) \end{aligned} \tag{3.76}$$

We have shown that we can represent **any** harmonic plane wave as **either** the sum of traveling waves **or** the sum of standing waves. Obviously, it works for P-waves too since we use the same trig identities. As it turns out, this duality of representations is far more general and can be applied to a variety of more complex problems. These two solutions are sometimes referred to as **characteristic** solutions and **mode** solutions. Figure 3.4 shows a schematic of how sinusoids traveling in opposite directions sum to make a standing wave.

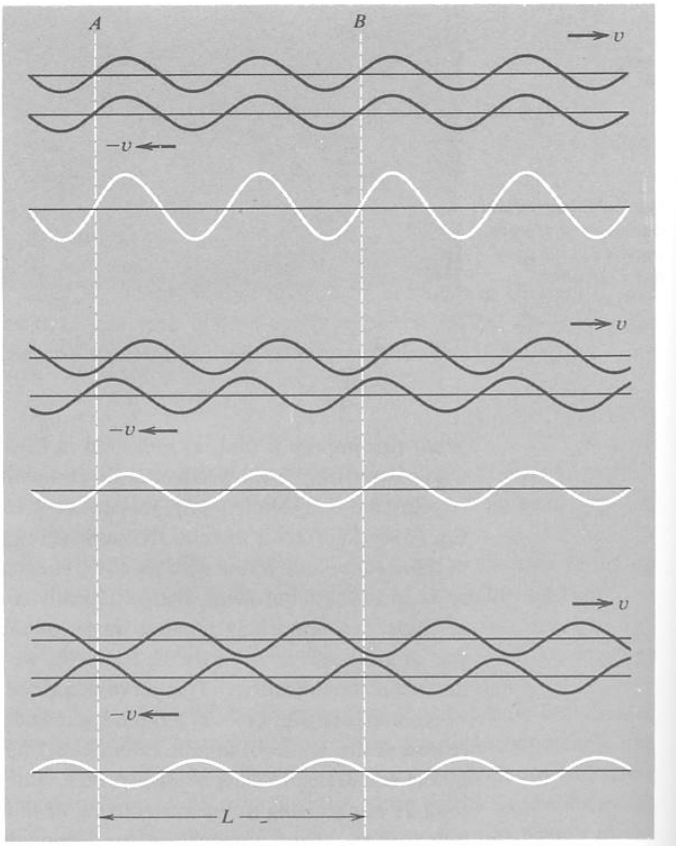


Fig. 7–3 Two exactly similar sinusoidal waves traveling in opposite directions and the resultant standing waves.

Figure 3.4. From "Vibration and Waves" by A. P. French. W.W. Norton and Co., 1971.

Spherical Waves

Many problems that we encounter concern the radiation of waves from a point in the medium. These waves spread spherically through the medium and their representation with Cartesian coordinates is awkward. In a homogeneous whole space, it is usually most natural to solve these problems in spherical coordinates. However, if there are layers in the medium, then it usually is more convenient to solve these problems in cylindrical coordinates. General solutions for these problems are quite complex and beyond the scope of this class. However, we can consider the following potential in spherical coordinates. This potential has radial symmetry.

$$\varphi(r,t)=\frac{1}{r}f\left(t-\frac{r}{\alpha}\right)+\frac{1}{r}g\left(t+\frac{r}{\alpha}\right) \tag{3.77}$$

This solves the transformed form of Navier's equation given by (3.35) and (3.38). When the problem is radially symmetric, this can be written as

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\varphi}{\partial r}\right)=\frac{1}{\alpha^2}\ddot{\varphi} \tag{3.78}$$

The displacement that results from this is

$$\begin{aligned} u_r=\frac{\partial\varphi}{\partial r}&=\frac{-1}{r^2}\left[f\left(t-\frac{r}{\alpha}\right)+g\left(t+\frac{r}{\alpha}\right)\right]-\frac{1}{\alpha r}\left[f'\left(t-\frac{r}{\alpha}\right)-g'\left(t+\frac{r}{\alpha}\right)\right] \\ &=\frac{-1}{r^2}\left[f\left(t-\frac{r}{\alpha}\right)+g\left(t+\frac{r}{\alpha}\right)\right]-\frac{1}{\alpha r}\left[\dot{f}\left(t-\frac{r}{\alpha}\right)-\dot{g}\left(t+\frac{r}{\alpha}\right)\right] \end{aligned} \tag{3.79}$$

I have chosen a solution with waves that travels both radially outward (the *f* terms) and inwards (the *g* terms). Each of these has terms that decay with distance as both r^{-1} and r^{-2}; these are called far-field and near-field terms, respectively. They are both required to solve Navier's equation for this radial wave problem. Notice that the far-field term has a time dependence that looks like the time derivative of the near-field term. Also notice that the far-field term is scaled by the factor α^{-1}.

We can enquire about the energy in the spherically symmetric P-wave by integrating the power that is exerted on a shell of radius, r . Recall that power per unit area is given by equation (3.54), or

$$P(t)=\left(4\pi r^2\right)\rho\alpha\dot{u}^2 \tag{3.80}$$

Inserting the far-field term from (3.79) into (3.80), we obtain

$$P(t)=4\pi\frac{\rho}{\alpha}\ddot{f}^2(t) \tag{3.81}$$

That is, the energy in the radiated far-field P-wave is the same as it passes through any spherical shell at any distance; the wave energy of far-field waves is conserved.

Pressure Step in a Spherical Cavity

We can explore this difference between near-field and far-field terms by investigating the exact solution to the problem of a step change in pressure p_0 inside a spherical cavity of radius *a*. The derivation is somewhat lengthy and is given by Achenbach. The answer for a Poisson solid is

$$u_r=p_0H(\hat{t})\frac{a^3}{4\mu r^2}\left\{1+\left[\left(\frac{r}{a}-\frac{1}{2}\right)\sqrt{2}\sin\omega_1\hat{t}-\cos\omega_1\hat{t}\right]e^{-b\hat{t}}\right\} \tag{3.82}$$

where

$$\hat{t}\equiv t-\frac{r}{\alpha} \tag{3.83}$$

$$\omega_1 = \frac{\alpha 2\sqrt{2}}{3a} \tag{3.84}$$

$$b = \frac{2\alpha}{3a} \tag{3.85}$$

At the surface of the cavity the displacement is

$$u_r\big|_{r=a} = p_0 \frac{a}{4\mu}\left(1 + e^{-bt}\frac{b}{\omega_1}\sin\omega_1 t - e^{-bt}\cos\omega_1 t\right) \tag{3.86}$$

This looks like the pressure rate convolved with the solution of damped harmonic oscillator problem subjected to a step in force (see equation (1.39)). The period of the undamped oscillator is given by (1.37), which when combined with (3.84) and (3.85) gives

$$\omega_0^2 = \omega_1^2 + b^2 = 3b^2 \tag{3.87}$$

The fraction of critical damping of this system is given by (1.5) and is equal to

$$\zeta = \frac{b}{\omega_0} = \frac{1}{\sqrt{3}} = 0.58 \tag{3.88}$$

So, the surface of the cavity is a 58% damped oscillator that settles about its new static equilibrium position. With each harmonic swing, it radiates wave energy to the far-field term, which at large r become.

$$u_r\big|_{r>>a} \approx p_0 \frac{a^2}{4\mu r} e^{-b\hat{t}}\sqrt{2}\sin\omega_1\hat{t} \tag{3.89}$$

The damping of the oscillating cavity is sometimes referred to as **radiation damping** and since it is linear and depends on the velocity at the source, it is very analogous to viscous damping discussed in the SDOF problem of chapter 1. The concept of radiation damping can become useful when investigating the damping of an oscillating building that excites seismic waves as it oscillates.

Of course, a spherical cavity has many other modes besides the radially symmetric mode just described. Each mode has its own natural frequency, mode shape, and radiation damping. The mode shapes are best described with spherical harmonics. Since the pressure problem is radially symmetric, we only need the fundamental mode solution that is given by (3.82).

Point Force

The displacement in the *i* direction from a point force in the *k* direction with time history *f(t)* was given by Love (The mathematical theory of elasticity, Dover Pubs., 1944) and is

$$u_i = \frac{1}{4\pi}\left\{\frac{\partial^2}{\partial x_i \partial x_k}\frac{1}{r}\int_{r/\alpha}^{r/\beta}\tau f(t-\tau)d\tau + \frac{1}{2}\left(\frac{\partial r}{\partial x_i}\right)\left(\frac{\partial r}{\partial x_k}\right)\left[\frac{1}{r\alpha^2}f\left(t-\frac{r}{\alpha}\right)-\frac{1}{r\beta^2}f\left(t-\frac{r}{\beta}\right)\right]+\frac{\delta_{ik}}{r\beta^2}f\left(t-\right.\right.$$

(3.90)

where

$$r^2 = x_i x_i \tag{3.91}$$

This is an important building point in seismology, since it allows us to calculate the wave field that results from distributions of forces. Although this solution is relatively compact, it is written in terms of both Cartesian coordinates and radial distance. It is easier to write the full solution in spherical coordinates in which case all of the spatial derivatives turn into a relatively complex set of sines and cosines of the angular geometric parameters. This is called "radiation pattern" and an example will be given in Chapter 7 (Sources).

Anelastic Attenuation of a Traveling Wave

The solutions discussed above are for an elastic medium. However, it is useful to introduce the concept that their energy slowly decays as they travel due to some inelastic response of the medium. In addition, there are basic physical considerations that require that waves eventually attenuate. One convenient approach to this problem is to break a waveform into its harmonic constituent parts and to then introduce the following definition of Q which is entirely analogous to the one that we used in Chapter 1 for the SDOF problem. Recall that for a lightly damped oscillator (equation 1.30),

$$Q \approx -2\pi\frac{E}{\Delta E} \tag{3.92}$$

where E and ΔE are the total energy and energy lost per cycle. We can also define the logarithmic decrement of the amplitude lost per cycle as

$$\delta \equiv \ln\left(\frac{A_1}{A_2}\right) \tag{3.93}$$

since energy is proportional to the square of amplitude,

$$\ln A = \frac{1}{2}\ln E \tag{3.94}$$

from which it follows that

$$Q \approx \frac{\pi}{\delta} \tag{3.95}$$

We can now write the expression for the amplitude A of a harmonic wave as a function of distance traveled r as

$$A(r) = A_0 e^{-\left(\omega/2Qc\right)r} \tag{3.96}$$

where c is the velocity of the wave. Sometimes the attenuation is described by the parameter t^* which is defined to be

$$t^* = \frac{r}{cQ} = \frac{\text{travel time}}{\text{quality factor}} \tag{3.97}$$

Homework for Chapter 3

1. Show that (3.40) and (3.59) are solutions to Navier's equation.

2. Show that (3.77) is a solution to Navier's equation.

3. If a plane harmonic wave with a frequency of 1 Hz and a propagation velocity of 3 km/sec is ½ the amplitude after traveling 100 km through an attenuating medium, then what is the Q and t^*?

Planar Waves in Layered Media

The behavior of planar P- and S-waves in a layered elastic medium is one of the most useful (and mercifully one of the easiest) problems in seismology. There are many instances where the Earth can be approximated by plane layers and wave fronts are also approximately planar. As an added bonus, these problems can also provide useful insights into the dynamic behavior of buildings.

Unlike the last chapter in which there was no inherent coordinate frame for a homogeneous isotropic whole space, it is natural for the layered space problem to choose one of the coordinate axes to be perpendicular to the layering. In most cases it is customary to choose this to be the x_3 axis. If the waves are approximately planar, we can use Cartesian coordinates which simplifies the problem considerably. It is important to realize, though, that there is no physical situation in which truly planar waves exist. If a full solution is desired to the problem of a point source in a layered medium, then cylindrical coordinates are the best choice. However, the problem becomes much more complex.

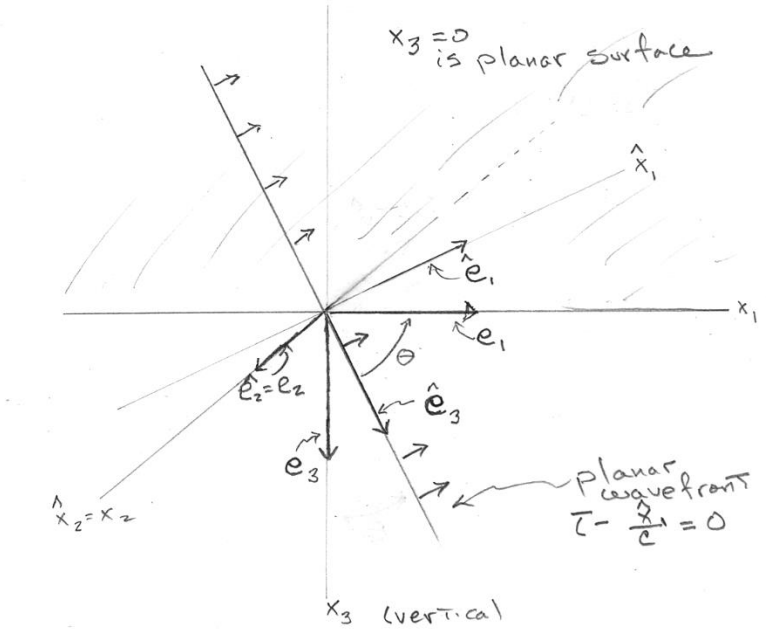


Figure 4.1. Natural Cartesian coordinate frames for planar waves in a layered medium

For the moment, let us consider Cartesian coordinate frames. There are two natural coordinate frames in the problem of planar waves in layered problems (see

Figure 4.1). The first coordinate frame $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ is that used in Chapter 3 and it is defined by the planar wavefronts, with $\hat{\mathbf{e}}_1$ being in the direction of wave propagation. The second coordinate frame is natural to the layers with (x_1, x_2, x_3) perpendicular to the layering. In the Earth, we can choose $\mathbf{e}_3$ as either up or down (both are used … entropy of the universe increases). Because it is useful to have positive numbers for this coordinate and since the surface of the Earth at the top, it is most common to choose it as down. Now we choose $\hat{\mathbf{e}}_2 = \mathbf{e}_2$ to be the intersection of the plane of the layering with the plane of the wavefront. $\mathbf{e}_1$ is now defined to be perpendicular to $\mathbf{e}_2$ and $\mathbf{e}_3$, which is the projection of the propagation direction on the layering. Finally, $\hat{\mathbf{e}}_3$ is chosen to be perpendicular to both $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$. θ is the angle between $\mathbf{e}_1$ and $\hat{\mathbf{e}}_3$ and it is usually referred to as the incidence angle; it is 0 degrees for a wavefront traveling perpendicular to the layering (a wave coming straight up). Using this coordinate frame, we see that planar P-waves can generally be written as

$$\mathbf{u}^P = f\left(t - \frac{\hat{x}_1}{\alpha}\right)\hat{\mathbf{e}}_1 \tag{4.1}$$

and shear waves can be generally written as

$$\mathbf{u}^S = \mathbf{u}^{SH} + \mathbf{u}^{SV} \tag{4.2}$$

where

$$\mathbf{u}^{SH} = f\left(t - \frac{\hat{x}_1}{\beta}\right)\hat{\mathbf{e}}_2 \tag{4.3}$$

and

$$\mathbf{u}^{SV} = g\left(t - \frac{\hat{x}_1}{\beta}\right)\hat{\mathbf{e}}_3 \tag{4.4}$$

Therefore, it is natural to decompose S-waves, whose particle motion is a 2-dimensional vector in the plane of the wavefront, into 1) the component that is horizontal (called the **SH wave**) and, 2) the component that has some vertical motion (called the **SV wave**). As before, *f* and *g* are independent functions with finite second derivatives.

When the source of the wave is a point, then the approximately planar waves are traveling radially away from the source, and we identify $\mathbf{e}_1, \mathbf{e}_2,$ and $\mathbf{e}_3$ as the **radial, transverse, and vertical** directions, respectively.

Now the solutions (4.1), (4.3) and (4.4) are written in the $\hat{\mathbf{x}}$-coordinate frame. However, it is most convenient to write these solutions in the $\mathbf{x}$-coordinate frame since the boundary conditions are more naturally described there. Now for the

plane wave shown in Fig 4.1, which is traveling in the plus x_1 direction and the minus x_3 direction,

$$\mathbf{u}^P = f\left(t - p_\alpha x_1 + \eta_\alpha x_3\right)\left(\mathbf{e}_1 \sin\theta - \mathbf{e}_3 \cos\theta\right) \tag{4.5}$$

$$\mathbf{u}^{SV} = f\left(t - p_\beta x_1 + \eta_\beta x_3\right)\left(-\mathbf{e}_1 \cos\theta - \mathbf{e}_3 \sin\theta\right) \tag{4.6}$$

$$\mathbf{u}^{SH} = f\left(t - p_\beta x_1 + \eta_\beta x_3\right)\mathbf{e}_2 \tag{4.7}$$

Where

$$p_\alpha = \frac{\sin\theta}{\alpha} \equiv \text{P-wave horizontal slowness} \equiv \text{ P-wave ray parameter} \tag{4.8}$$

$$p_\beta = \frac{\sin\theta}{\beta} \equiv \text{S-wave horizontal slowness} \equiv \text{ S-wave ray parameter} \tag{4.9}$$

$$\eta_\alpha = \frac{\cos\theta}{\alpha} \equiv \text{P-wave vertical slowness} \tag{4.10}$$

$$\eta_\beta = \frac{\cos\theta}{\beta} \equiv \text{S-wave vertical slowness} \tag{4.11}$$

SH-Waves in Two Welded Half-Spaces

The simplest problem of a plane wave in layered media is that of planar SH-waves in two welded half-spaces. The geometry of the problem is shown in Figure 4.3, a and b. I show the location of a traveling wavefront in part a, and in part b, I show an equivalent **ray diagram,** which is a shorthand from optics that shows the vector directions of the normal to wavefronts.

As in all of these problems, we have the correct solution when it 1) solves Navier's equation, 2) has continuous displacements on either side of the boundary, and 3) has equal and opposite traction vectors operating on either side of the boundary. In this simple problem it turns out that, except for certain incidence angles, it is possible to match these 3 conditions by the appropriate sum of three planar SH-waves, the **incident wave** in medium 1, the **reflected wave** in medium 1, and the **transmitted wave** in medium 2. As we have seen, planar SH-waves already solve Navier's equation. We can write the solution as

$$\mathbf{u}^1 = A_I f\left(t - p x_1 + \eta_1 x_3\right)\mathbf{e}_2 + A_R f\left(t - p x_1 - \eta x_3\right)\mathbf{e}_2 \tag{4.12}$$

$$\mathbf{u}^2 = A_T f\left(t - px_1 + \eta_2 x_3\right)\mathbf{e}_2 \tag{4.13}$$

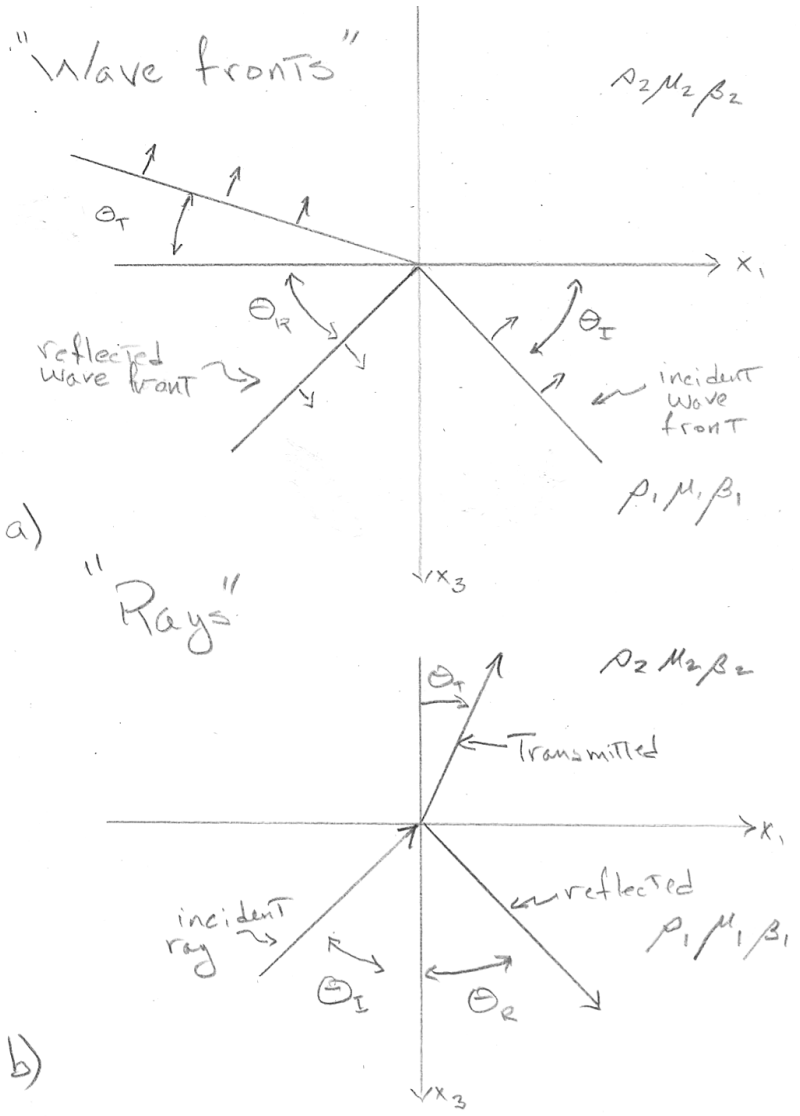


Figure 4.2 SH plane wave from a faster to a slower medium

The requirement of continuous displacements on either side of the boundary can only occur if the planar wave fronts travel along the boundary at the same **apparent horizontal velocity** *c*. This is simply a statement of Snell's law. Since apparent velocities are infinite for vertically incident waves, seismologists sometimes use a parameter called slowness, which is defined to be the inverse of velocity. As it turns out for vertically layered problems, the apparent horizontal velocity of the wavefront (or **horizontal phase velocity)** is constant in the

problem. The slowness of this horizontal phase velocity is called the **ray parameter**, *p*. By the same token, we can define a vertical phase velocity and its corresponding **vertical slowness** η, which is *not* the same in both media. That is

$$c = \frac{\beta_1}{\sin\theta_I} = \frac{\beta_1}{\sin\theta_R} = \frac{\beta_2}{\sin\theta_T} \tag{4.14}$$

$$p \equiv \frac{1}{c}, \text{ the same in both media} \tag{4.15}$$

$$\eta_1 = \frac{\cos\theta_I}{\beta_1} = \frac{\cos\theta_R}{\beta_1} \tag{4.16}$$

$$\eta_2 = \frac{\cos\theta_T}{\beta_2} \tag{4.17}$$

While (4.14) is a necessary condition for continuity of displacement, it is not sufficient. To solve the problem fully, we need to use the boundary condition that tractions at the interface are equal, which can be stated as

$$\sigma_{i3}^1\left(x, y, z = 0^+\right) = \sigma_{i3}^2\left(x, y, z = 0^+\right), \quad i = 1, 2, 3 \tag{4.18}$$

Because we have postulated a planar SH wave traveling in the x_1-direction, the only stress involved is σ_{23}, so

$$\sigma_{23}^1\left(x, y, z = 0^+\right) = \sigma_{23}^2\left(x, y, z = 0^+\right) \tag{4.19}$$

and

$$\sigma_{13}^1\left(x, y, z\right) = \sigma_{13}^2\left(x, y, z\right) = \sigma_{33}^1\left(x, y, z\right)\sigma_{33}^2\left(x, y, z\right) = 0 \tag{4.20}$$

It can be shown that condition (4.18), together with the continuity of displacement is satisfied if

$$\frac{A_R}{A_I} \equiv R_{SS}^{SH} = \frac{\mu_1\eta_1 - \mu_2\eta_2}{\mu_1\eta_1 + \mu_2\eta_2} = \frac{\mu_1\beta_2\cos\theta_1 - \mu_2\beta_1\cos\theta_2}{\mu_1\beta_2\cos\theta_1 + \mu_2\beta_1\cos\theta_2} \tag{4.21}$$

and

$$\frac{A_T}{A_I} \equiv T_{SS}^{SH} = \frac{2\mu_1\eta_1}{\mu_1\eta_1 + \mu_2\eta_2} = \frac{2\mu_1\beta_2}{\mu_1\beta_2\cos\theta_1 + \mu_2\beta_1\cos\theta} \tag{4.22}$$

R_{SS}^{SH} and T_{SS}^{SH} are called the reflection and transmission coefficients, respectively.

We can look at the special case of an SH plane wave that is reflecting off of a free surface. There is no transmitted wave in this case and the reflection coefficient is +1. This means that, at the free surface the incident and transmitted waves sum together in phase, or from (4.12) we conclude that

$$\mathbf{u}^{SH}\left(t; x_3 = 0\right) = 2A_I\, f\left(t - px_1\right)\mathbf{e}_2 \tag{4.23}$$

That is, the waves are twice as large at the free surface as they are inside the medium. Although the displacement is large, the stress is zero because of the free surface boundary condition.

The astute reader will note that this solution cannot make sense if the horizontal phase velocity of the incident wave is less than the shear-wave velocity in the transmitting medium; that is, β_2 is the minimum possible horizontal phase velocity in medium 2, and therefore if

$$c = \frac{\beta_1}{\sin\theta_1} < \beta_2 \tag{4.24}$$

then a purely plane wave solution is not possible. This is called a **post-critical reflection** and it turns out that the wave is totally reflected, but as we will see in a later chapter, the solution is fairly complex. $\theta_c = \sin^{-1}\left(\beta_1 \big/ \beta_2\right)$ is called the **critical angle**.

It is instructive to investigate the amplitude of the transmitted SH wave at vertical incidence, in which case (4.22) becomes

$$T_{SS}^{SH}\Big|_{\theta_1=0} = \frac{2\mu_1\beta_2}{\mu_1\beta_2 + \mu_2\beta_1} = \frac{2}{1+\sqrt{\dfrac{\mu_2\rho_2}{\mu_1\rho_1}}} \tag{4.25}$$

Therefore, as the density times the rigidity of the transmitted medium becomes small, the amplitude of the transmitted wave can become twice as large as the incident wave. This may seem like a violation of conservation of energy, but it is not. Recall from equation (3.68) that energy flux is

$$P = \rho_2\beta_2\dot{u}_2^2 = \sqrt{\rho_2\mu_2}\,\dot{u}_2^2 \tag{4.26}$$

While the maximum amplitude of the transmitted wave is twice that of the incident wave, we can achieve even higher amplifications by transmitting the wave

through a stack of smoothly decreasing velocity. In this case, there is very little reflection from each layer and most of the energy is transmitted through the entire medium. When this happens, the energy flux at the top is approximately that which is fed into the bottom. So for a **smooth velocity gradient**,

$$P_I \approx P_T \tag{4.27}$$

Or

$$u_2^T \approx u_2^I \sqrt{\frac{\rho_I \beta_I}{\rho_T \beta_T}} \tag{4.28}$$

where the indices *T* and *I* represent the material properties on the transmitted and incident sides of the gradient, respectively. Therefore, the transmitted wave can become very large in amplitude with respect to the incident wave.

P- and SV- Plane Waves

The problem gets far more complex when we consider either a P-wave or an SV-wave as the incident wave. The motion now involves two components of motion, which must be continuous on the boundary, and there are now two components of stress to match across the boundary.

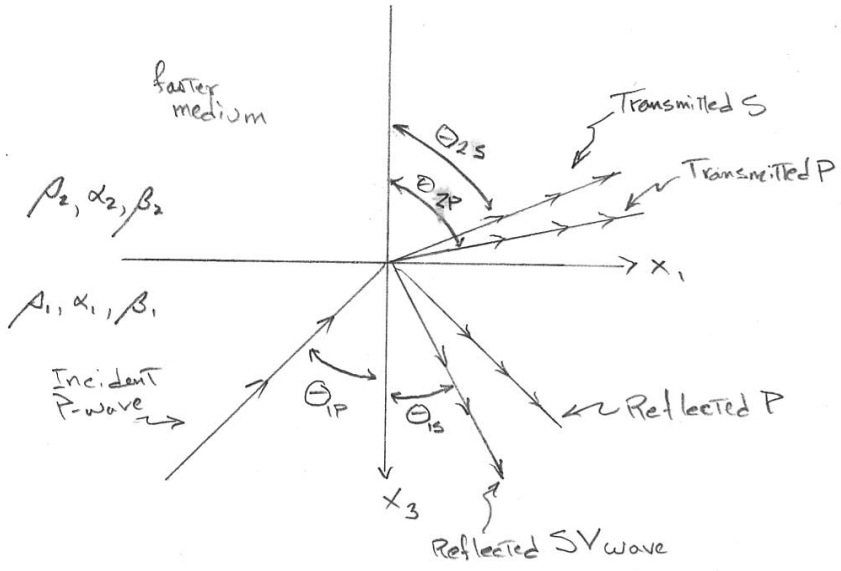


Figure 4.2 Ray diagram for a planar P-wave incident on a faster medium

Let us first consider the case of an incident planar P-wave. It turns out that the boundary conditions can be satisfied with the following superposition of plane waves. The ray diagram for these waves is shown in Figure 4.2.

$$\begin{aligned}\mathbf{u}^1 = {} & A_{IP} f\left(t - px_1 + \eta_{\alpha_1} x_3\right)\left(\mathbf{e}_1 \sin\theta_{P1} + \mathbf{e}_3 \cos\theta_{P1}\right) \\ & + A_{IP} R_{PP} f\left(t - px_1 - \eta_{\alpha_1} x_3\right)\left(\mathbf{e}_1 \sin\theta_{P1} - \mathbf{e}_3 \cos\theta_{P1}\right) \\ & + A_{IP} R_{PS} f\left(t - px_1 - \eta_{\beta_1} x_3\right)\left(\mathbf{e}_1 \cos\theta_{S1} - \mathbf{e}_3 \sin\theta_{S1}\right)\end{aligned} \tag{4.29}$$

The three terms represent incident P, reflected P, and reflected SV waves. The solution in medium 2 has the form

$$\begin{aligned}\mathbf{u}^2 = {} & A_{IP} T_{PP} f\left(t - px_1 + \eta_{\alpha_2} x_3\right)\left(\mathbf{e}_1 \sin\theta_{P2} - \mathbf{e}_3 \cos\theta_{P2}\right) \\ & + A_{IP} T_{PS} f\left(t - px_1 + \eta_{\beta_2} x_3\right)\left(\mathbf{e}_1 \sin\theta_{S2} + \mathbf{e}_3 \cos\theta_{S2}\right)\end{aligned} \tag{4.30}$$

As in the case of the SH wave, all of these wavefronts have identical horizontal phase velocities or

$$c = \frac{1}{p} = \frac{\alpha_1}{\sin\theta_{1P}} = \frac{\alpha_2}{\sin\theta_{2P}} = \frac{\beta_1}{\sin\theta_{1S}} = \frac{\beta_2}{\sin\theta_{2S}} \tag{4.31}$$

Two critical angles are now possible if

$$\theta_{cPP} = \sin^{-1}\left(\frac{\alpha_1}{\alpha_2}\right) \tag{4.32}$$

$$\theta_{cPS} = \sin^{-1}\left(\frac{\alpha_1}{\beta_2}\right) \tag{4.33}$$

The transmission and reflection coefficients are derived in such a way that the following boundary conditions are satisfied. There are four non-zero boundary conditions for this problem

$$u_1^+ = u_1^- \tag{4.34}$$

$$u_3^+ = u_3^- \tag{4.35}$$

$$\sigma_{33}^+ = \sigma_{33}^- \tag{4.36}$$

$$\sigma_{13}^+ = \sigma_{13}^- \tag{4.37}$$

The following Table from Lay and Wallace gives the values of the reflection and transmission coefficients.

TABLE 3.1 Displacement Reflection and Transmission Coefficients

Coefficient	Formula
Solid–free surface (P–SV)	
R_{PP}	$\{-[(1/\beta^2)-2p^2]^2+4p^2\eta_\alpha\eta_\beta\}/A$
R_{PS}	$\{4(\alpha/\beta)p\eta_\alpha[(1/\beta^2)-2p^2]\}/A$
R_{SP}	$\{4(\beta/\alpha)p\eta_\beta[(1/\beta^2)-2p^2]\}/A$
R_{SS}	$\{-[(1/\beta^2)-2p^2]^2+4p^2\eta_\alpha\eta_\beta\}/A$
$R_{SS}(SH)$	1
Solid–solid (P–SV)	
R_{PP}	$[(b\eta_{\alpha_1}-c\eta_{\alpha_2})F-(a+d\eta_{\alpha_1}\eta_{\beta_2})Hp^2]/D$
R_{PS}	$-[2\eta_{\alpha_1}(ab+cd\eta_{\alpha_2}\eta_{\beta_2})p(\alpha_1/\beta_1)]/D$
T_{PP}	$[2\rho_1\eta_{\alpha_1}F(\alpha_1/\alpha_2)]/D$
T_{PS}	$[2\rho_1\eta_{\alpha_1}Hp(\alpha_1/\beta_2)]/D$
R_{SS}	$-[(b\eta_{\beta_1}-c\eta_{\beta_2})E-(a+b\eta_{\alpha_2}\eta_{\beta_1})Gp^2]/D$
R_{SP}	$-[2\eta_{\beta_1}(ab+cd\eta_{\alpha_2}\eta_{\beta_2})p(\beta_1/\alpha_1)]/D$
$R_{SS}(SH)$	$\dfrac{\mu_1\eta_{\beta_1}-\mu_2\eta_{\beta_2}}{\mu_1\eta_{\beta_1}+\mu_2\eta_{\beta_2}}$
$T_{SS}(SH)$	$\dfrac{2\mu_1\eta_{\beta_1}}{\mu_1\eta_{\beta_1}+\mu_2\eta_{\beta_2}}$
$a=\rho_2(1-2\beta_2^2p^2)-\rho_1(1-2\beta_1^2p^2)$	$E=b\eta_{\alpha_1}+c\eta_{\alpha_2}$
$b=\rho_2(1-2\beta_2^2p^2)-2\rho_1\beta_1^2p^2$	$F=b\eta_{\beta_1}+c\eta_{\beta_2}$
$c=\rho_1(1-2\beta_1^2p^2)+2\rho_2\beta_2^2p^2$	$G=a-d\eta_{\alpha_1}\eta_{\beta_2}$
$d=2(\rho_2\beta_2^2-\rho_1\beta_1^2)$	$H=a-d\eta_{\alpha_2}\eta_{\beta_1}$
$D=EF+GHp^2$	
$A=[(1/\beta^2)-2p^2]^2+4p^2\eta_{\alpha_1}\eta_{\beta_1}$	

From Lay and Wallace

These coefficients are algebraically complex, and their behavior as a function of incidence angle is quite complex. Ewing, Jardetsky, and Press (Elastic Waves in Layered media) shows several plots of the behavior of these coefficients.

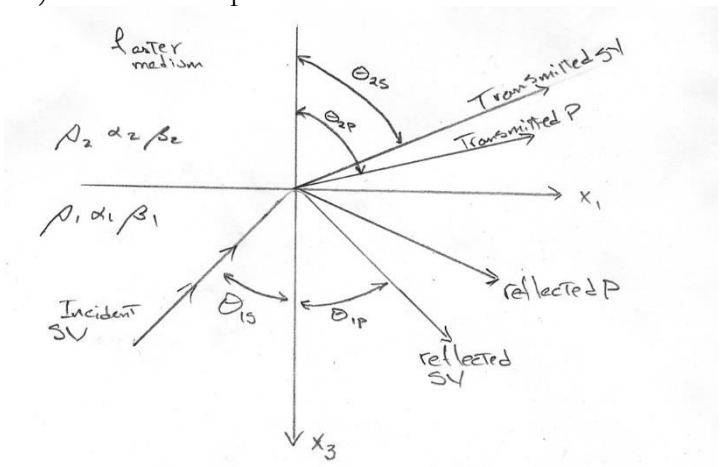


Figure 4.3 Ray diagram for a Planar SV wave incident on a faster medium.

The case for an **incident SV-wave** is very similar to the incident P-wave and the ray diagram for this case is shown in Figure 4.3. However, now there are as many as **three possible critical angles**, depending on the wave speeds. They can occur if

$$\theta_{cSS} = \sin^{-1}\left(\frac{\beta_1}{\beta_2}\right) \tag{4.38}$$

$$\theta_{cSP} = \sin^{-1}\left(\frac{\beta_1}{\alpha_2}\right) \tag{4.39}$$

$$\theta_{cSP} = \sin^{-1}\left(\frac{\beta_1}{\alpha_1}\right) \tag{4.40}$$

SH Waves in a Plate

We now address the problems of planar SH-waves in a plate. We will first solve the problem where both top and bottom surfaces are free, and then we will solve the problem where the bottom surface is fixed.

We begin with the problem where both the top and bottom surfaces are free. In this case there will be an infinite series of planar SH waves that are shown schematically in Figure 4.4. The boundary conditions are satisfied by the free surface reflection coefficient which, for SH waves is simply 1. To keep things simple, we restrict ourselves to looking at a single point on the free surface that we will set to be the origin.

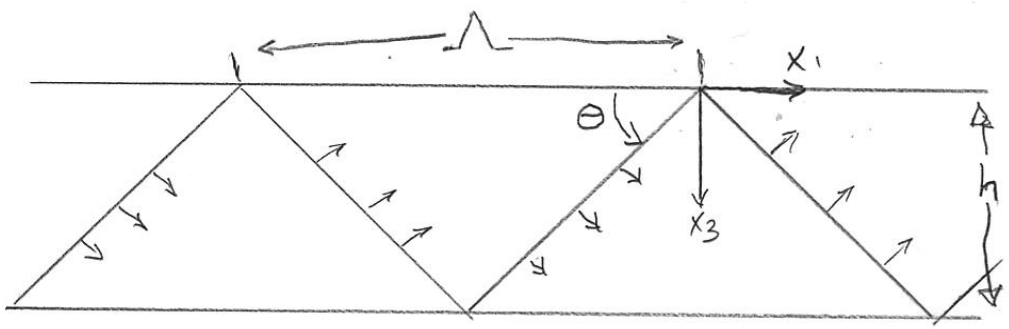


Figure 4.4. Geometry of a multiply reflecting SH-wave in a plate of thickness *h*.

In this case we can write the solution as

$$u_2\left(\mathbf{x}=\mathbf{0};t\right)=f\left(t\right)*\sum_{n=-\infty}^{\infty}\delta\left(t-nT\right)=f\left(t\right)*\mathrm{III}\left(\frac{\mathrm{t}}{\mathrm{T}}\right) \tag{4.41}$$

where III is called the **sampling function** (or sometimes called a comb function) and is defined as

$$\mathrm{III}\left(\frac{\mathrm{t}}{\mathrm{T}}\right)=\sum_{n=-\infty}^{\infty}\delta\left(t-nT\right) \tag{4.42}$$

The sampling function (and several other related functions) is shown in Figure 4.4. The Fourier transform of a sampling function is also a sampling function (see Bracewell) and it is also shown in Figure 4.5. In this case

$$\mathrm{FT}\left[\mathrm{III}\left(\frac{t}{T}\right)\right]=T\mathrm{III}\left(\frac{T\omega}{2\pi}\right) \tag{4.43}$$

Therefore we can write the Fourier transform of the displacement as

$$\tilde{u}\left(\omega\right)=\tilde{f}\left(\omega\right)T\,\mathrm{III}\left(\frac{T\omega}{2\pi}\right)=T\sum_{n=-\infty}^{\infty}\tilde{f}\left(\omega_n\right) \tag{4.44}$$

for our purposes we will only consider non-negative values of n.

$$\omega_n=\frac{2\pi n}{T} \tag{4.45}$$

Therefore we see in (4.44) the solution just picks off discrete values of the Fourier transform of our function *f*. We can write the solution back in the time domain as

$$u\left(t\right)=T\sum_{n=-\infty}^{\infty}\left\{\mathrm{Re}\left[\tilde{f}\left(\omega_n\right)\right]\cos\left(\omega_n t\right)+\mathrm{Im}\left[\tilde{f}\left(\omega_n\right)\right]\sin\left(\omega_n t\right)\right\} \tag{4.46}$$

We call the discrete frequencies the **modes** of the solution. However, notice that these modes represent traveling waves except at vertical incidence. While they are modes, they are not normal modes.

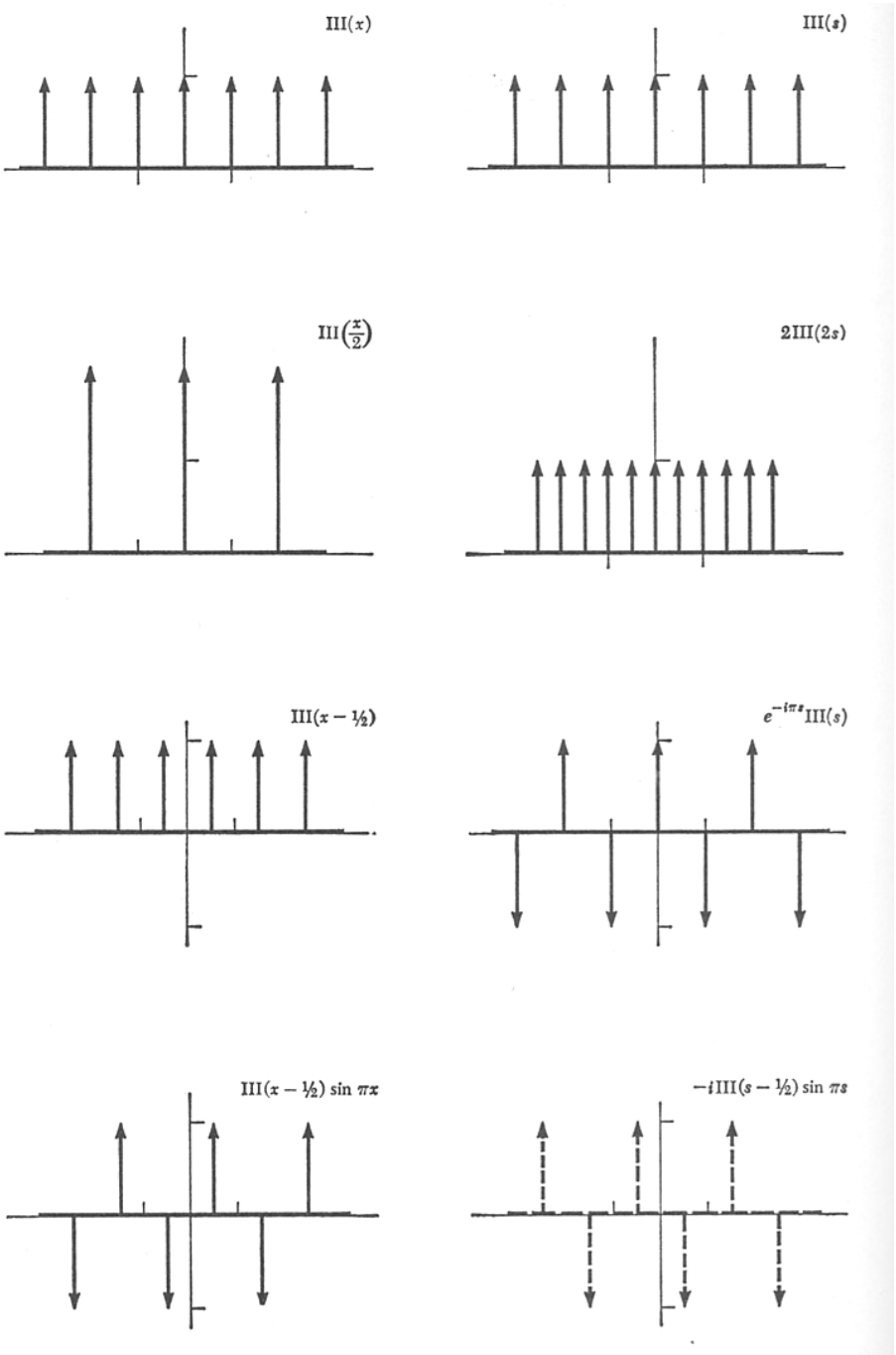


Figure 4.5 (from Bracewell, The Fourier Transform and its Applications). Notice that Bracewell uses the notation $\omega = 2\pi s$.

Recall though, that our original problem only specifies the thickness of the layer and *T* is still undetermined. From Figure 4.4 we see that

$$T = \frac{\Lambda}{c} \tag{4.47}$$

Where

$$c = \frac{\beta}{\sin\theta} \tag{4.48}$$

Now

$$\Lambda = \frac{2h}{\tan\theta} = 2h\frac{\cos\theta}{\sin\theta} = 2h\frac{\sqrt{1-\sin^2\theta}}{\sin\theta} = 2h\sqrt{\frac{1}{\sin^2\theta}-1} \tag{4.49}$$

substituting (4.48) into (4.49) gives

$$\frac{c^2}{\beta^2} = 1+\left(\frac{\Lambda}{2h}\right)^2 \tag{4.50}$$

Therefore the horizontal phase velocity depends on the wavelength of the propagating wave. We can combine (4.45) and (4.47) to write

$$\Lambda = cT = 2\pi n\frac{c}{\omega_n} \tag{4.51}$$

substituting (4.51) into (4.50), and solving for c gives

$$c = \frac{\beta}{\sqrt{1-\left(\frac{n\pi\beta}{\omega_n h}\right)^2}} \tag{4.52}$$

We can interpret (4.52) in the following way. We can choose any mode number n and then we can choose any frequency ω_n and then we can use (4.52) to calculate the phase velocity. The phase velocity as a function of frequency is called the **dispersion curve** for the n^{th} mode of the problem.

Notice that when the wave is traveling vertically $(\theta = 0)$ then both the phase velocity and the wavelength become infinite. However, we can still compute the modal frequencies for this case by recognizing that the periodicity of our sampling function in (4.41) is just twice the travel time through the layer or

$$T\Big|_{\theta=0} = \frac{2h}{\beta} \tag{4.53}$$

and then using (4.45) we see that

$$\omega_n\Big|_{\theta=0} = \frac{n\pi\beta}{h};\ \ n = 1,2,3,... \tag{4.54}$$

where we have dropped the 0^{th} mode, since it yields zero frequency and is not physically meaningful in this case. Notice that when there is a 0^{th} mode, it corresponds to any horizontally traveling SH wave. The modes become true standing waves when the waves are traveling vertically.

We can also inquire what the full spatial solution looks like for each of these modes. That is, we assume that the time history is a cosine wave having a modal frequency and use (4.12) to write

$$\begin{aligned}\mathbf{u^1} &= \left\{\cos\left[\omega_n\left(t - px_1 + \eta x_3\right)\right] + \cos\left[\omega_n\left(t - px_1 - \eta x_3\right)\right]\right\}\mathbf{e_2} \\ &= 2\cos\left[\omega_n\left(t - px_1\right)\right]\cos\left(\omega_n \eta x_3\right)\mathbf{e_2}\end{aligned} \tag{4.55}$$

Therefore we see that this is a cosine traveling with a particular phase velocity and with a depth dependence that is a fixed cosine function. We can now recognize that the vertical mode shapes must be such that the shear stress and shear strain are zero at the boundaries of the plate, or

$$\sigma_{23}\Big|_{x_3=0} = \mu\varepsilon_{23}\Big|_{x_3=0} = 0 \tag{4.56}$$

$$\sigma_{23}\Big|_{x_3=h} = \mu\varepsilon_{23}\Big|_{x_3=h} = 0 \tag{4.57}$$

This must mean that

$$\varepsilon_{23}\Big|_{x_3=0,h} = \frac{\partial u_2}{\partial x_3}\Big|_{x_3=0,h} = -2\omega_n\eta\cos\left[\omega_n\left(t - px_1\right)\right]\sin\left(\omega_n\eta x_3\right)\Big|_{x_3=0,h} = 0 \tag{4.58}$$

This condition is only satisfied if the vertical distribution of the displacement has the forms shown in Figure 4.5.

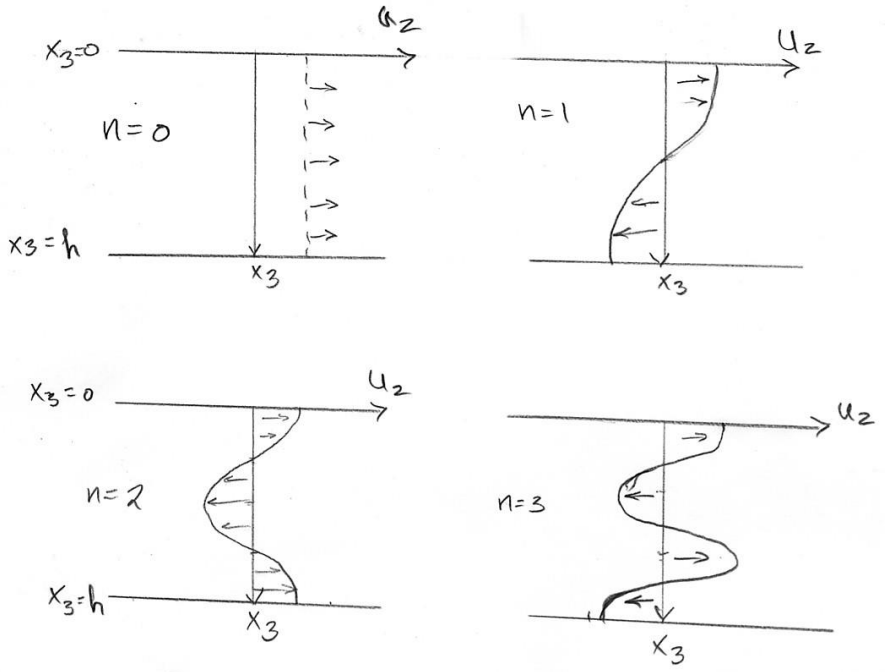


Figure 4.5. Depth dependence of horizontal displacement for the first 4 modes of harmonic SH-waves in a plate with free boundaries.

Planar SH-waves in a Plate with a Rigid Base

While the free plate problem is simple and useful to show modes and dispersion, a more important problem for us is one in which the base of the plate is fixed to a rigid medium.

This problem is identical to the free-boundary problem, except that the reflection coefficient at the bottom of the layer is -1 instead of +1. That is, the up- and down-going waves in Figure 4.3 destructively interfere at the bottom of the layer so that there is no displacement here. If the plane waves are impulse functions, then the ground motion at the origin is now an alternating series of positive and negative impulses as given by

$$u_2(\mathbf{x}=\mathbf{0};t) = f(t) * \sum_{n=-\infty}^{\infty} \left[\delta(t-nt)\right] - \delta\left(t-\frac{T}{2}-nT\right)$$

$$= f(t) * \left[\mathrm{III}\left(\frac{\mathrm{t}}{\mathrm{T}}\right) - \mathrm{III}\left(\frac{\mathrm{t}}{\mathrm{T}}-\frac{1}{2}\right)\right] = f(t) * \mathrm{III}\left(\frac{\mathrm{t}}{\mathrm{T}}-\frac{1}{2}\right)\sin\left(\frac{\pi t}{T}\right) \tag{4.59}$$

Figure 4.4 shows the Fourier transform of this alternating comb function (the bottom function in the figure) is also an alternating comb function, which can be written

$$FT\left[\text{III}\left(\frac{\text{t}}{\text{T}}-\frac{1}{2}\right)\sin\left(\frac{\pi t}{T}\right)\right]=-iT\,\text{III}\left(\frac{\text{T}\,\omega}{2\pi}-\frac{1}{2}\right)\sin\left(\frac{T\omega}{2}\right) \tag{4.60}$$

This solution is very similar to the free plate solution, except the frequencies of the modes have been changed. There is no mode for $n=0$, since that corresponds to uniform horizontal motion with depth, which cannot be possible if the motion at the bottom is zero. The modal frequencies are given by

$$\omega_n=\frac{\pi\left(2n-1\right)}{T};\qquad n=1,2,3,... \tag{4.61}$$

In this case the alternating comb function repeats itself only every other reflection, and so in this case

$$T=\frac{2\Lambda}{c} \tag{4.62}$$

where Λ is the same as shown in Figure 4.4, and as before

$$c=\frac{\beta}{\sin\theta} \tag{4.63}$$

Just as before,

$$\frac{c^2}{\beta^2}=1+\left(\frac{\Lambda}{2h}\right)^2 \tag{4.64}$$

However, in this case,

$$2\Lambda=cT=2\pi n\frac{c}{\omega_n} \tag{4.65}$$

Therefore, the horizontal phase velocity is

$$c=\frac{\beta}{\sqrt{1-\left(\frac{n\pi\beta}{2\omega_n h}\right)^2}} \tag{4.66}$$

As before, if the wavefront is vertically incident, the phase velocity become infinite, and the interfering waves form a true standing wave with

$$T\Big|_{\theta=0}=\frac{4h}{\beta} \tag{4.67}$$

And

$$\omega_n\Big|_{\theta=0} = \frac{(2n-1)\pi\beta}{2h}; \qquad n = 1,2,3,... \tag{4.68}$$

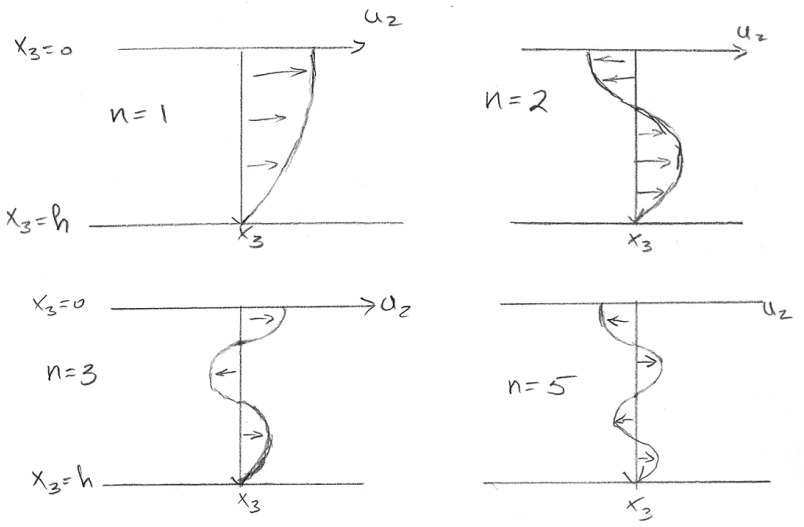


Figure 4.6. Amplitude as a function of depth for the first 4 SH modes of a plate with a fixed bottom and a free surface at the top.

We can now compute the mode shape by noting that we must now add both positive and negative waves that are shifted by a quarter wavelength, so that

$$\begin{aligned}
\mathbf{u^1} &= \cos\left[\omega_n\left(t - px_1 + \eta x_3\right)\right]\mathbf{e_2} + \cos\left[\omega_n\left(t - px_1 - \eta x_3\right)\right]\mathbf{e_2} \\
&\quad -\sin\left[\omega_n\left(t - px_1 + \eta x_3\right)\right]\mathbf{e_2} - \sin\left[\omega_n\left(t - px_1 - \eta x_3\right)\right]\mathbf{e_2} \\
&= 2\cos\left[\omega_n\left(t - px_1\right)\right]\cos\left(\omega_n \eta x_3\right)\mathbf{e_2} - 2\sin\left[\omega_n\left(t - px_1\right)\right]\cos\left(\omega_n \eta x_3\right)\mathbf{e_2} \\
&= 2\cos\left(\omega_n \eta x_3\right)\left\{\cos\left[\omega_n\left(t - px_1\right)\right] - \sin\left[\omega_n\left(t - px_1\right)\right]\right\}\mathbf{e_2} \\
&= 2\sqrt{2}\cos\left(\omega_n \eta x_3\right)\cos\left[\omega_n\left(t - px_1\right) - \frac{\pi}{4}\right]\mathbf{e_2}
\end{aligned} \tag{4.69}$$

So we end up with a sum of a sinusoid and a cosine in the x_1 direction, and our familiar cosine in the vertical direction. Because there are different natural frequencies, the mode shapes are different as is shown in Figure 4.6. Notice that the displacement is zero at the bottom boundary, but the strain

(derivative of displacement with respect to vertical position) is zero at the top.

Comparing Figures 4.5 and 4.6, it is easy to see why free plates have mode frequencies that are integer multiples of the fundamental mode, whereas a plate with one fixed boundary has modal frequencies that are odd integer multiples of the fundamental mode frequency.

Phase Velocity and Group Velocity

We have seen that planar plate waves have horizontal phase velocities $c = 1/p$ that vary with the frequency of the motion, or alternatively with the wavelength of the waves. In particular, we can write them in generalized form as

$$\begin{aligned} u_2 &= g(x_3)\sin\left[\omega\left(t - \frac{x_1}{c}\right)\right] \\ &= g(x_3)\sin(\omega t - kx_1) \end{aligned} \tag{4.70}$$

where $g(x_3)$ is a harmonic function of depth and $c = \Lambda/T = \omega/k$, and k is the horizontal spatial frequency which is commonly referred to as horizontal wavenumber. That is for a harmonic traveling wave, the phase velocity can always be measured by taking the frequency divided by the appropriate wavenumber. Remember, though, this is just the apparent velocity of some particular phase in the harmonic wave. Since harmonic waves have no beginning or end, they do not carry any information. In order to carry information (or energy) the wave must vary in time and space; it must be possible to follow a wave packet. The easiest way to understand a wave packet is to consider the case of two harmonic waves of similar (but different) frequencies. That is, consider

$$\begin{aligned} u_2(x_1,t) &= A_0\cos(\omega_1 t - k_1 x_1) + A_0\cos(\omega_2 t - k_2 x_1) \\ &= 2A_0\cos\left(\frac{\omega_1-\omega_2}{2}t - \frac{k_1-k_2}{2}x_1\right)\cos\left(\frac{\omega_1+\omega_2}{2}t - \frac{k_1+k_2}{2}x_1\right) \\ &= u_{group}u_{carrier} \end{aligned} \tag{4.71}$$

Where $u_{carrier}$ is the sinusoid with the average frequency of the signals and it's frequency is that of the "carrier signal." The two signals beat against each other and the beat frequency is the difference in frequencies of the two signals. Of course, there are similar carrier and beat wavenumbers. As before the *phase velocity of the carrier* is just

$$c_{carrier} = \frac{\omega_{carrier}}{k_{carrier}} = \frac{\omega_1 + \omega_2}{k_1 + k_2} \tag{4.72}$$

And the *group velocity* of the amplitude envelope is just

$$c_{group} = \frac{\omega_{group}}{k_{group}} = \frac{\omega_1 - \omega_2}{k_1 - k_2} = \frac{\Delta\omega}{\Delta k} \approx \frac{d\omega}{dk} \tag{4.73}$$

The wavegroup carries energy and information and its speed must not exceed the intrinsic velocity of the medium. The phase velocity can be any number higher than the intrinsic wave velocity. If you know $\omega(k)$, then you can always obtain the group velocity by differentiating with respect to *k*. For example, if we return the problem of a plane wave in a plate, the phase velocity is given by (4.52), or

$$c_{carrier} = \frac{\omega_n}{k} = \frac{\beta}{\sqrt{1 - \left(\frac{n\pi\beta}{\omega_n h}\right)^2}} \tag{4.74}$$

Or

$$\omega_n = \frac{\beta}{h}\sqrt{k^2h^2 + n^2\pi^2} \tag{4.75}$$

So

$$c_{group} = \frac{d\omega_n}{dk} = \beta h \frac{k}{\sqrt{k^2h^2 + n^2\pi^2}} \tag{4.76}$$

Or we can infer the group velocity as a function of ω by noting that $k = \frac{\omega_n}{c_{carrier}}$, so that (4.76) becomes

$$c_{group} = \frac{\beta h \omega_n}{c_{carrier}\sqrt{\omega_n{}^2 \frac{h^2}{c_{carrier}^2} + n^2\pi^2}} \tag{4.77}$$

Substituting (4.74) into (4.77), we obtain

$$c_{group} = \frac{\beta h \omega_n}{\frac{\beta}{\sqrt{1-\left(\frac{n\pi\beta}{\omega_n h}\right)^2}} \sqrt{\omega_n{}^2 \frac{h^2}{\frac{\beta^2}{1-\left(\frac{n\pi\beta}{\omega_n h}\right)^2}} + n^2\pi^2}}$$

$$= \beta \frac{1}{\sqrt{1+\frac{n^2\pi^2\beta^2}{\left(\omega_n^2 h^2 - n^2\pi^2\beta^2\right)}}} \qquad (4.78)$$

Planar SH-Waves in Layered Media

The problem of horizontally-polarized S-waves in a vertically stratified space is one of the most fundamental problems in Engineering Seismology. This is primarily because of the large changes in seismic velocity near the Earth's surface. That is, seismic velocities generally increase dramatically with depth as overburden pressures increase. There can also be large velocity contrasts at the water table. Furthermore, the seismic velocity of loosely consolidated sediments can be much lower than that of crystalline rock. Since the average S-wave velocity at the typical depth of earthquakes is 3.5 km/sec, most shear-waves are traveling nearly vertically as they propagate through the upper 100 meters of the Earth. For near-vertical incidence, there is little distinction between SV and SH; they are both nearly horizontally polarized. Therefore, it is of great interest to investigate how a horizontally polarized S-wave propagates through a series of horizontal layers.

The following Figures are from "Borehole velocity measurements and geological conditions at thirteen sites in the Los Angeles, Californis, region," by Gibbs, Tinsely, Boore, and Joyner (USGS Open File Report 00-470). They provide a good idea of the nature of near-surface seismic velocities encountered at real-world construction sites.

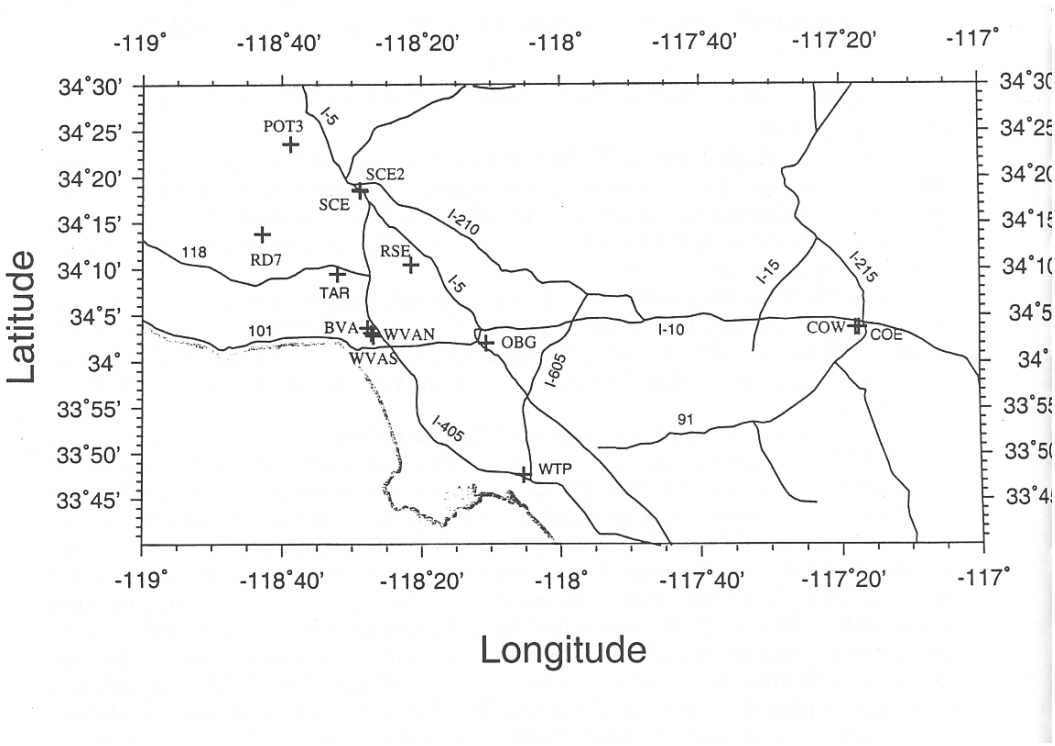


Figure 4.7 Locations of boreholes for velocity profiles shown in following figures

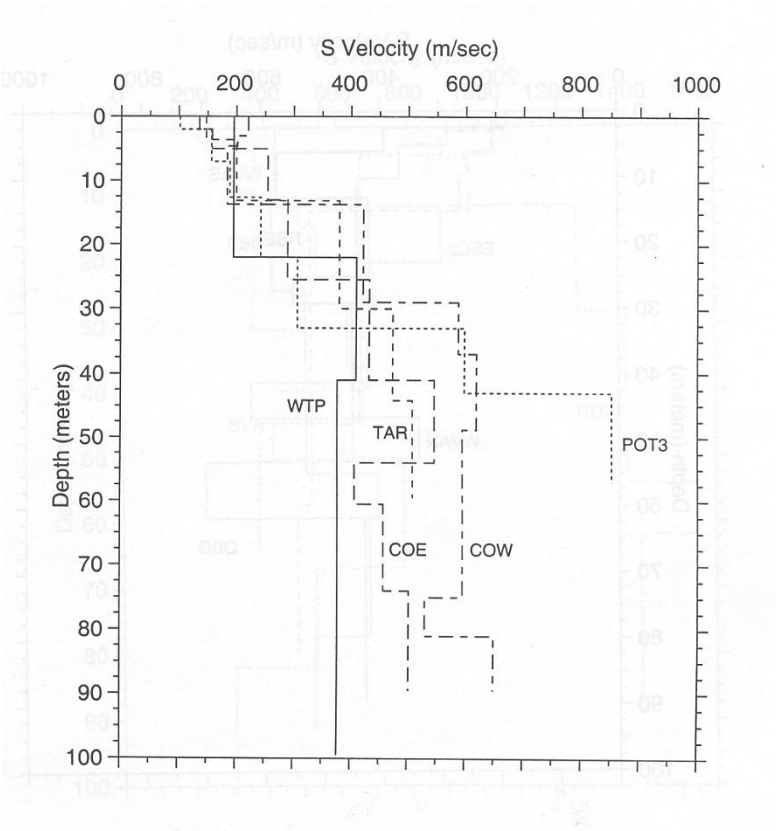
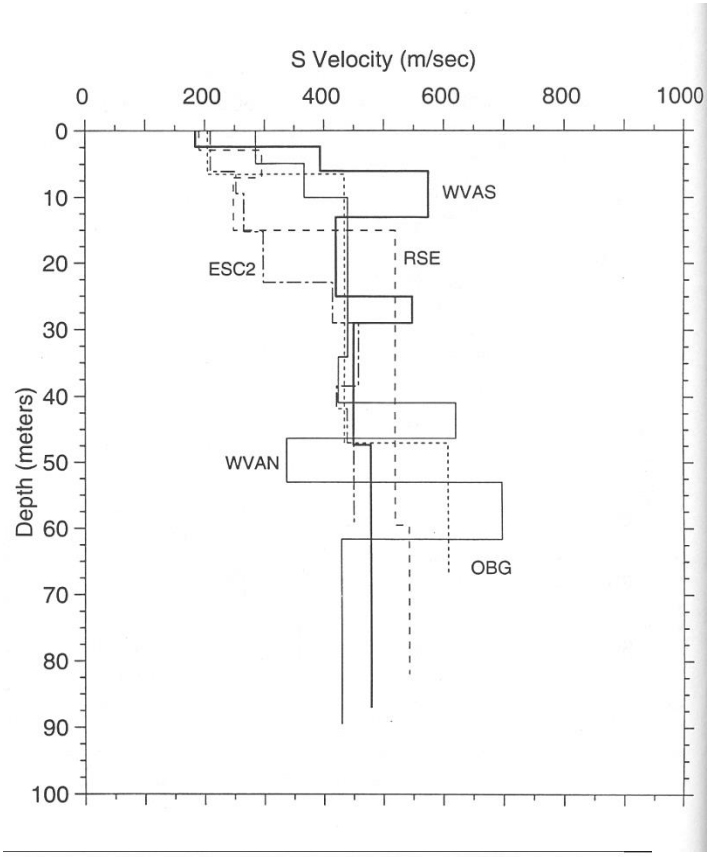


Figure 4.8 Shear-wave velocities for boreholes at stations shown in Figure 4.7



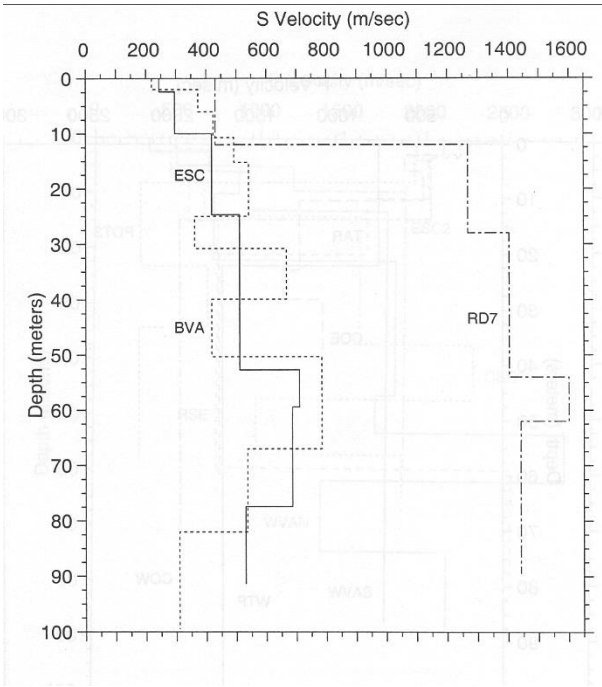


Figure 4.9 and 4.10 Shear-wave velocities for boreholes at stations shown in Figure 4.7

Thompson (Transmission of elastic waves through a stratified solid medium, J. Appl. Phys., 21, 89-93, 1950) and Haskell (The dispersion of surface waves on multi-layered media, Bull. Seism. Soc. Am., 43, 17-34, 1953) describe a clever method for solving the problem of a harmonic planar P-, SV-, or SH-wave through a vertically stratified medium. The algebra of the problem is beyond the scope of these notes, but I will sketch out the solution for a planar harmonic SH-wave. The methodology is known as **Thompson-Haskell Propagator Matrices** and it is the basis of a widely used computer program known as **Shake**. The following is a modification of Haskell's classic paper, "Crustal reflection of plane SH-waves" (JGR, 1960, 4147-4150).

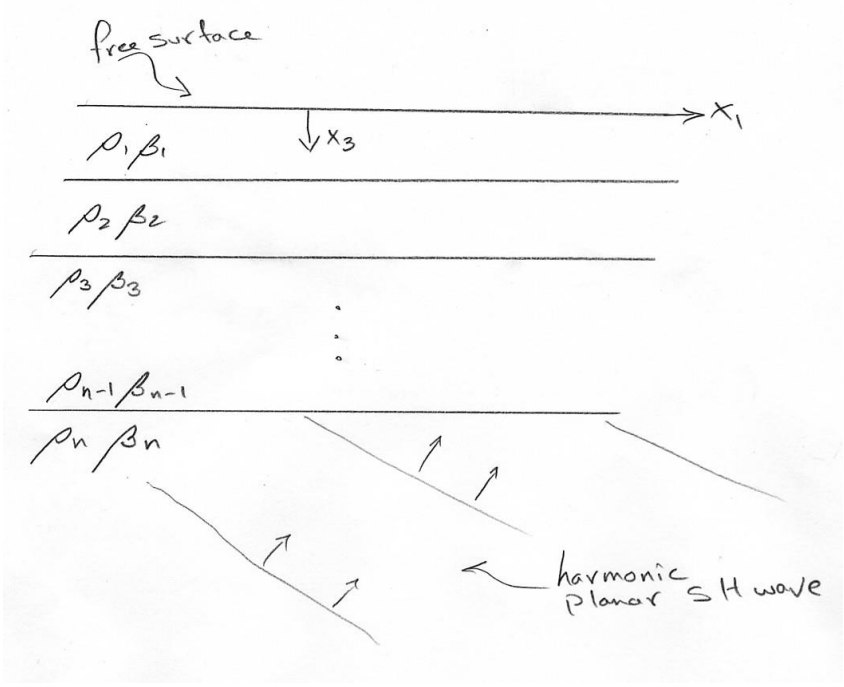


Figure 4.11

Consider the stack of *n-1* plane layers over an elastic half space with a planar SH-wave incident from the half-space as is shown in Figure 4.11. We will consider that the plane wave is harmonic. This requirement can be relaxed later, since we can sum harmonic solutions to form the solution for any time history for the incident wave. Motivated by the solutions for a planar SH-wave in a plate, we recognize that solutions to our problem will have some part that travels horizontally at the same phase velocity throughout the medium and another part of the solution that describes the depth dependence of the motion. That is, the solution will be of the form

$$\mathbf{u} = g\left(x_3\right)\cos\left[\omega\left(t - px_1\right) - \delta\right]\mathbf{e_2} \tag{4.79}$$

where $g(x_3)$ is a function of depth that will allow us to match the appropriate boundary conditions, and δ is a phase lag that may also be a function of depth.

At each boundary, there are two important conditions, continuous displacement and stress. Thompson and Haskell devised a clever solution method that writes these boundary conditions in matrix form. If we denote the amplitude of the up-going SH wave incident at the bottom of the stack as g^I, then the amplitude at the surface of the Earth g^0 can be written as

$$g^0 = g^I \left(\frac{2\mu_n \gamma_n}{\mu_n \gamma_n A_{11} + A_{21}} \right) \tag{4.80}$$

Where

$$\gamma_n \equiv \sqrt{\frac{c^2}{\beta_n^2} - 1} \tag{4.81}$$

And

$$\mathbf{A} = \prod_{m=n-1}^{1} \mathbf{a}_m = \mathbf{a}_{n-1}\mathbf{a}_{n-2}\mathbf{a}_{n-2}..\mathbf{a}_1 \tag{4.82}$$

And

$$\mathbf{a}_m = \begin{bmatrix} \cos Q_m & i(\mu_m \gamma_m)^{-1} \sin Q_m \\ i(\mu_m \gamma_m)^{-1} \sin Q_m & \cos Q_m \end{bmatrix} \tag{4.83}$$

And

$$Q_m = \frac{\omega h_m \gamma_m}{c} = \omega h_m \sqrt{\frac{1}{\beta_m^2} - \frac{1}{c^2}} \tag{4.84}$$

Unfortunately, these equations become singular for vertical incidence. That is, at vertical incidence, the horizontal phase velocity becomes infinite in which case,

$$\gamma_m \Big|_{c\to\infty} = \frac{c}{\beta} \tag{4.85}$$

$$Q_m \Big|_{c\to\infty} = \frac{\omega h_m}{\beta_m} \tag{4.86}$$

and then

$$\mathbf{a}_m \Big|_{c\to\infty} = \begin{bmatrix} \cos\left(\omega h_m \beta_m^{-1}\right) & 0 \\ 0 & \cos\left(\omega h_m \beta_m^{-1}\right) \end{bmatrix} \tag{4.87}$$

Thus, $a_{21} \to 0$ at vertical incidence.

We can gain some insight by considering the simple case of a single layer over a half space. In this case, $n = 2$ and we need only consider one propagator matrix. Therefore,

$$A_{11} \cos Q_1 \tag{4.88}$$

$$A_{21} = i\left(\mu_1 \gamma_1\right)^{-1} \sin Q_1 \tag{4.89}$$

And

$$g^0 = g^I \left(\frac{2}{\cos Q_1 + ib \sin Q_1} \right) \tag{4.90}$$

Where

$$b = \frac{\mu_1 \gamma_1}{\mu_2 \gamma_2} \tag{4.91}$$

we can write the amplitude of the amplification as

$$\left| \frac{g^0}{g^I} \right|^2 = \frac{2}{\cos^2 Q_1 + b \sin^2 Q_1} = \frac{2}{1 + (b-1) \sin^2 Q_1} \tag{4.92}$$

The nature of the amplification depends on whether or not *b* is greater than 1. Let c_1 be the phase velocity for which $b = 1$, then

$$c_1^2 = \frac{\rho_2^2 \beta_2^4 - \rho_1^2 \beta_1^4}{\rho_2^2 \beta_2^2 - \rho_1^2 \beta_1^2} \tag{4.93}$$

If $c > c_1$, then $b < 1$ and the maxima of $\left| \frac{g^0}{g^I} \right|$ occur at frequencies given by $\cos Q_1 = 0$, with maximum values given by $\left| \frac{g^0}{g^I} \right|_{\max} = \frac{2}{b}$. The minima occur at frequencies given by $\sin Q_1 = 0$ and the minimum amplifications are 2.

If $c < c_1$, then $b > 1$ and the maximum amplifications occur when $\sin Q_1 = 0$ and the maximum amplifications are 2. The minima occur when $\cos Q_1 = 0$ and have values of $\left| \frac{g^0}{g^I} \right|_{\min} = \frac{2}{b}$.

Figure 4.12 is from Haskell's paper and shows the amplification of planar SH-waves as a function of incidence angle (0 is vertical incidence). Amplification is

given relative to a half-space which always has a free-surface amplification of 2. The model is layer of average crustal properties over a half-space with average upper mantle properties.

$$\begin{aligned} \rho_1 &= 2.869\, g/cm^3 \\ \beta_1 &= 3.635\, km/\sec \\ h_1 &= 37 km \\ \rho_2 &= 3.337\, g/cm^3 \end{aligned} \tag{4.94}$$

For vertical incidence, you can see that the maximum amplification occurs at periods given by

$$T_n = \frac{4h}{(2n-1)\beta} \qquad n = 1,2,3,... \tag{4.95}$$

which is identical to the modal frequencies obtained from the plate with a rigid bottom in (4.68).

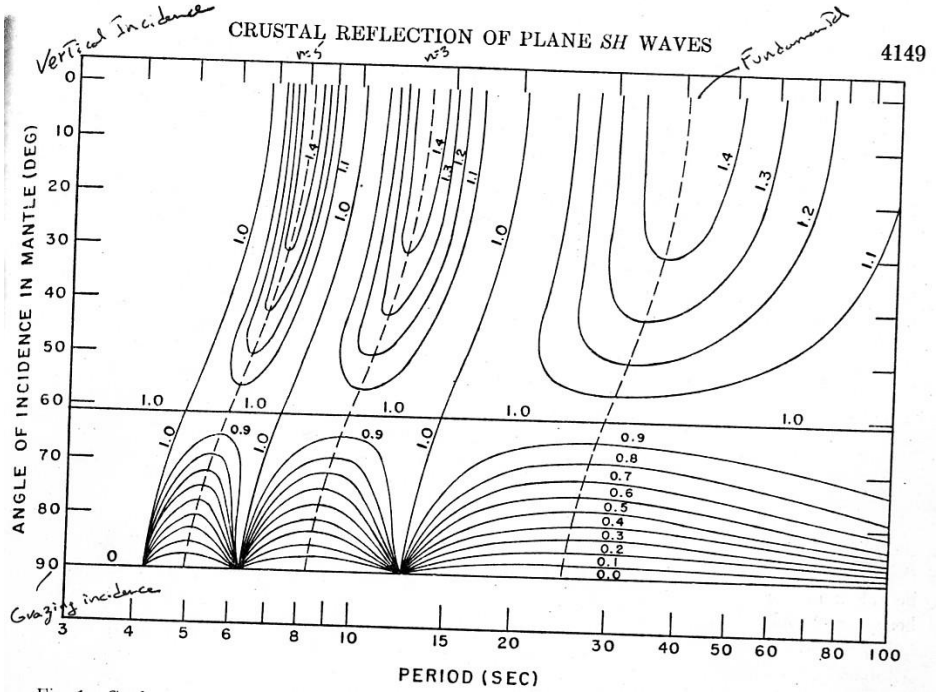


Figure 4.10 Amplification of SH-waves relative to a half-space as a function of incidence angle and period of a harmonic SH-wave incident on the crust. From Haskell (1960).

For vertical incidence the maximum amplification at these frequencies is simply

$$\left|\frac{g^0}{g^I}\right| = \frac{\rho_2 \beta_2}{\rho_1 \beta_1} \tag{4.96}$$

Whereas the maximum amplification of an SH-wave transmitted across an interface is 2 (see equation(4.25)**), the maximum amplification of a harmonic wave can be very large at certain frequencies if there is a very low velocity layer over a half space**.

San Francisco Bay Mud

Propagator matrix solutions are very important for analyzing problems where there is a layer of very low velocity material overlying a much higher velocity material. One of the most dramatic examples of this is on the margins of the San Francisco Bay where poorly consolidated Holocene mud deposits overly crystalline rocks. The following figure are from "Analysis of seismograms from a downhole array in sediments near San Francisco Bay" by Joyner, Warrick and Oliver (1970, BSSA, 60, 937-958). Figure 4.11 shows the location of a 186-m deep borehole adjacent to the Dumbarton Bridge.

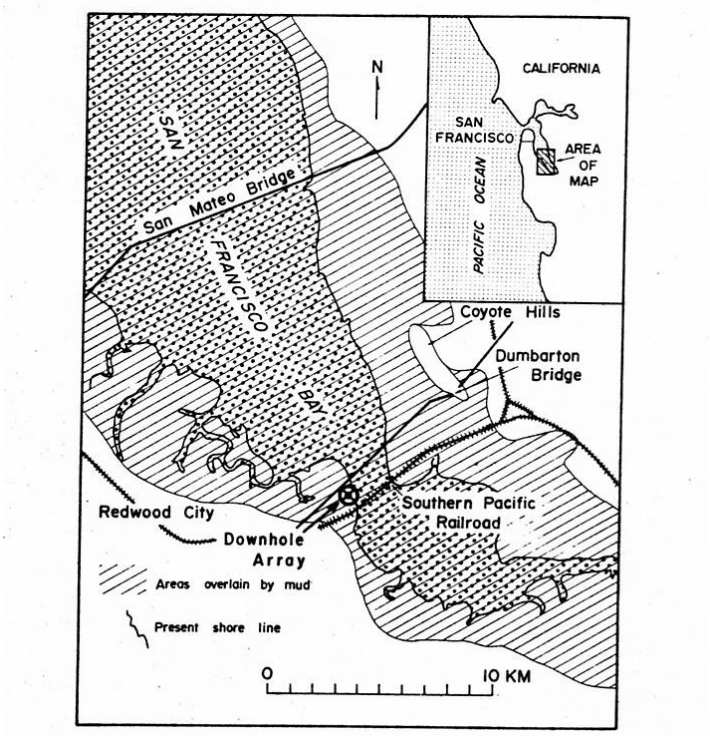


Figure 4.11 Map of southern part of San Francisco Bay showing location of the borehole. From Joyner and others.

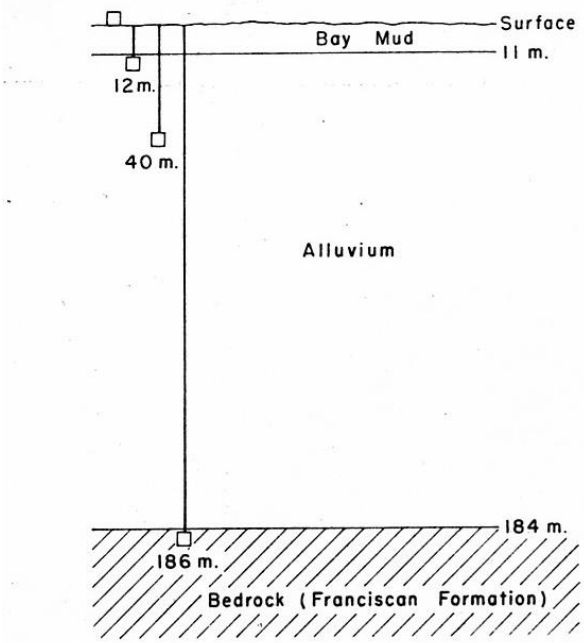


Figure 4.12 Schematic of the locations of seismometers in the borehole. From Joyner and others

Four 3-component sets of velocity-transducer seismometers with a natural frequency of 4.5 Hz were installed at depths of 0, 12, 40, and 186 m below the surface. The Bay mud is a Holocene mud whose mass is 50% water. The Alluvial deposits are thought to be Pleistocene.

Figure 4.13 show the estimated S-wave velocity in the borehole and Figure 4.14 shows the seismograms from different levels for a M 3.1 earthquake. Notice that not all of the channels were working. However, as we would predict from the wave propagation discussion in this chapter, the signals are much larger at the surface than they are at the bottom of the hole.

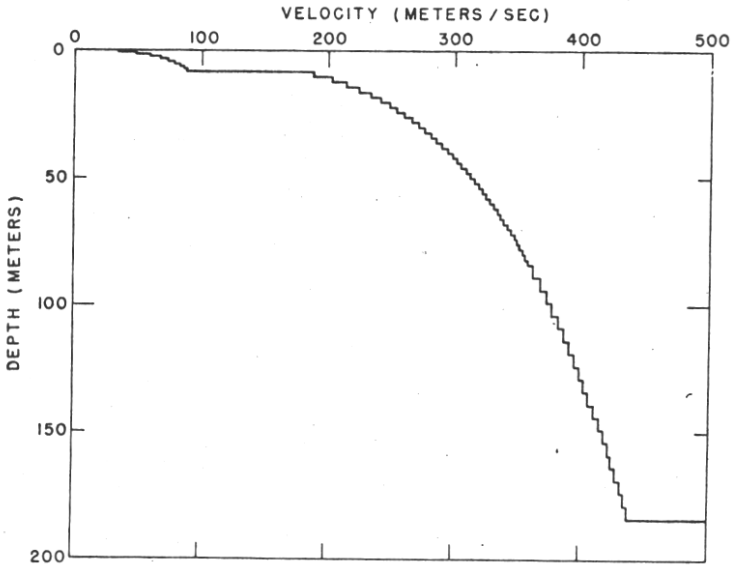


Figure 4.12 Shows an estimate of shear-wave velocity as a function of depth. The velocity model was developed by a combination of seismic travel-time data and from other studies of the properties of bay mud. The velocity in the bedrock is 2200 m/sec. From Joyner and others.

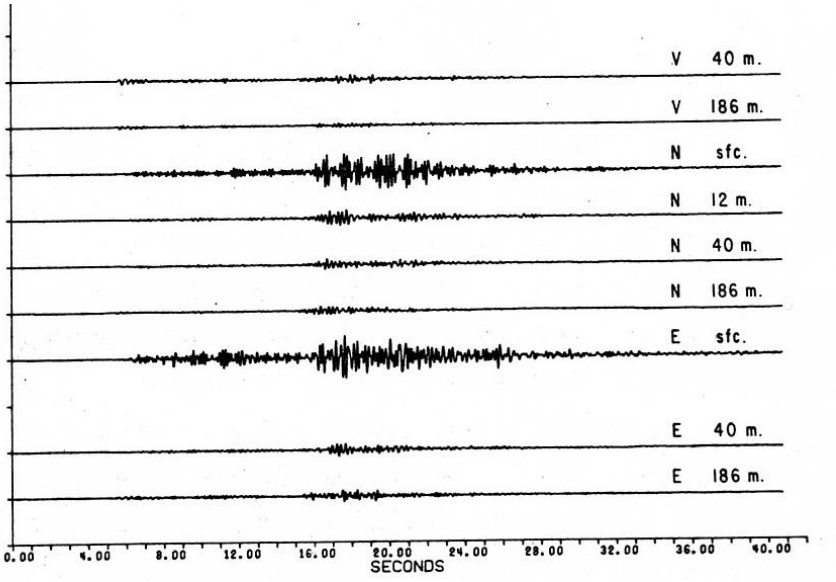


Figure 4.13 Seismograms from a M 3.1 at a distance of 79 km.

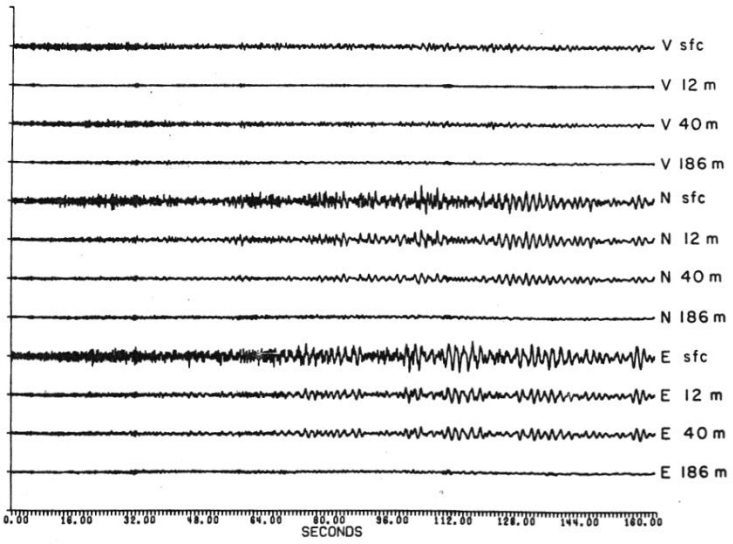


Figure 4.14 Seismograms from the 1971 M 6.7 San Fernando earthquake at a distance of 485 km. From Joyner and others. Notice the harmonic nature of the seismograms from the shallow sites. This is caused by resonance of vertically propagating S-waves.

There are actually several resonance periodicities that we expect since there are two strong velocity contrasts; one at the mud-alluvium interface (11 m) and another at the alluvium-bedrock interface (180 m). The velocity of the bay mud averages about 70 m/sec (150 mph); the velocity of sound in air is 330 m/sec. Equation (4.95) tells us that we expect to see amplification at frequencies of about 1.6 Hz, 4.8 Hz, etc. for the mud layer. The alluvium layer is faster, but much thicker and it has a resonance somewhere near 0.5 Hz.

The mud-layer resonance can be seen in Figure 4.15 where the ratios of the Fourier amplitude spectra of the seismometers at the top and bottom of the mud are shown for a M3.6 earthquake. Also shown are predictions of these amplifications using Thompson-Haskell propagator matrices and the velocity model shown in Figure 4.12. The computations were performed for several different values of attenuation Q.

The resonance of the overall soil-mud column with the bedrock is shown in Figure 4.16, where the spectral ratio of the surface vs. the bedrock is shown for the 1971 San Fernando Earthquake. Spectral ratios are shown for different segments of the record.

Finally, Figures 4.17 and 4.18 show the basement and surface ground motions along with the ground motion assuming that most of the motion is due to vertically propagating S-waves, and that this motion can be computed using propagator matrices to "propagate" the motion from the basement to the surface.

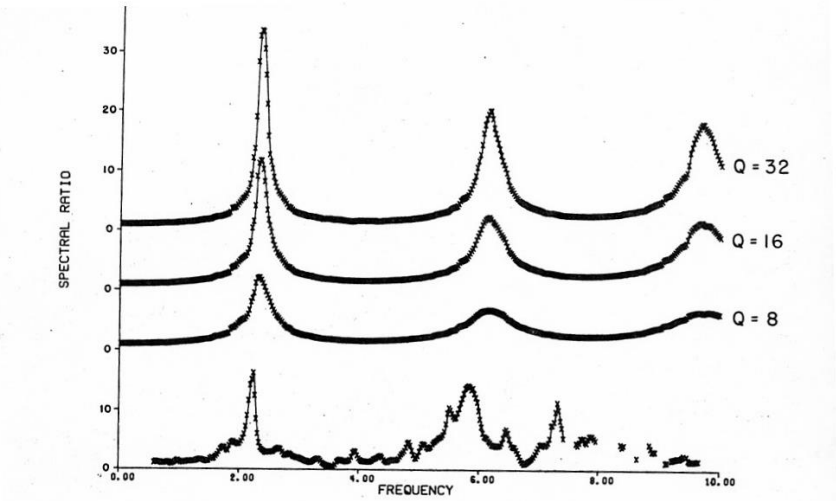


FIG. 7. Horizontal spectral ratios between the surface and 12 m. The *bottom curve* gives the observed ratio for the north component from the Pittsburg earthquake (magnitude 3.6, distance 66 km). The *other curves* give the predicted ratios computed for the model described in the text with different values of Q assumed for the material above 8 m.

Figure 4.15 from Joyner and others.

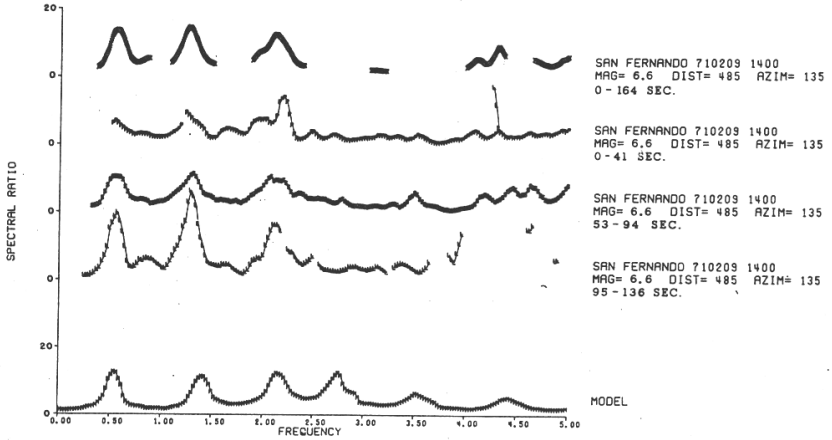


FIG. 9. Horizontal spectral ratios between the surface and 186 m. Computed ratio for the model is compared with observed ratios for the north component from different segments of the record of the San Fernando earthquake, as described in the text.

Figure 4.16 from Joyner and others

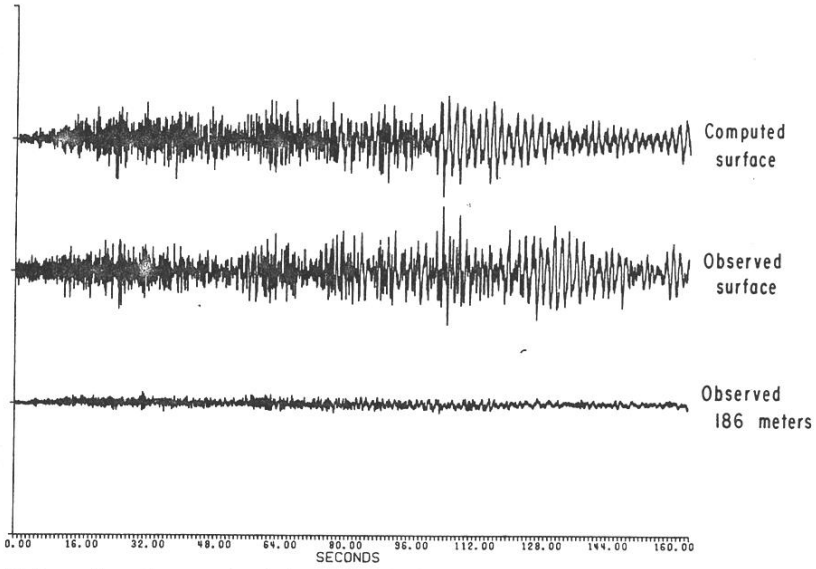


FIG. 14. Comparison of computed and observed motion for the north component from the San Fernando earthquake.

Figure 4.17 from Joyner and others

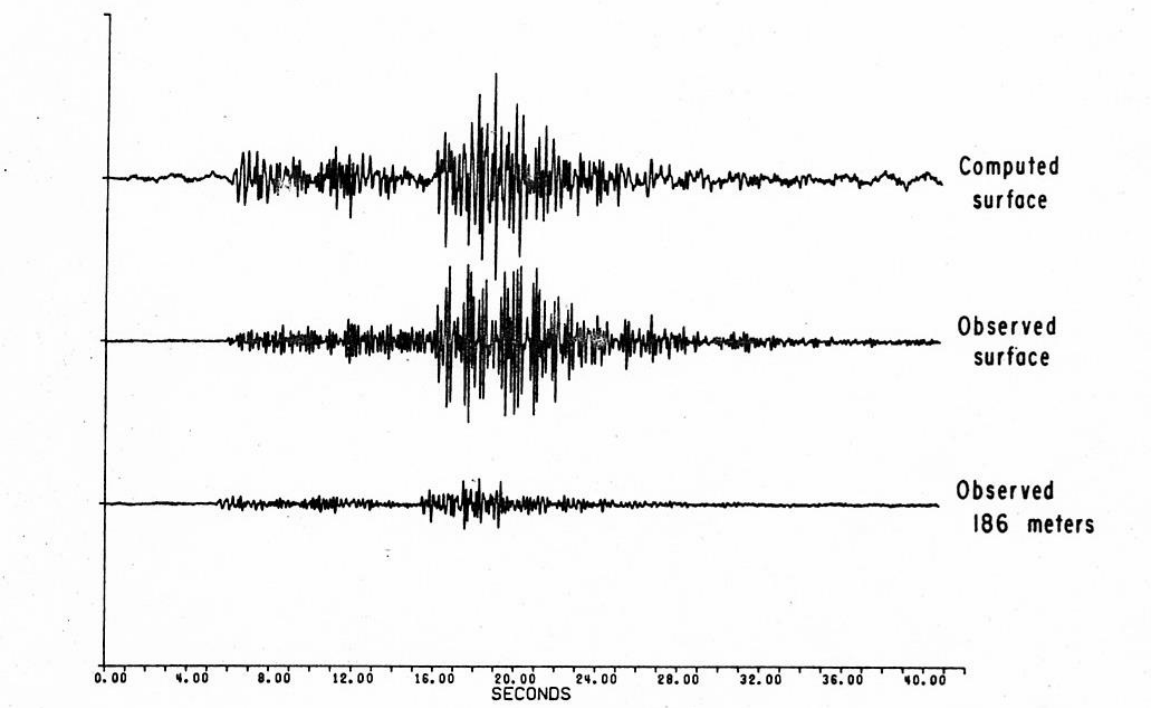


FIG. 16. Comparison of computed and observed motion for the north component from the Anzar earthquake (magnitude 3.1, distance 79 km).

Figure 4.18 from Joyner and others.

One of the most famous of all soil resonances occurred in the soft sedimentary lake deposits on which Mexico City is built. These deposits have S-wave velocities of about 80 m/sec and thicknesses of about 50 m. Figure 4.19 shows acceleration records from the 1985 Michoacan earthquake (M8.2). Notice how much larger the accelerations were in the sites in Mexico City.

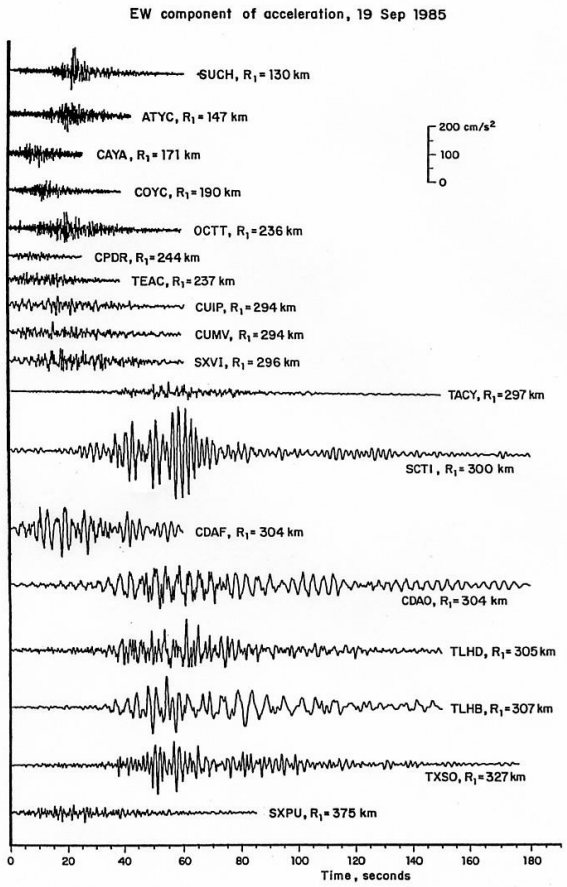


Figure 3 East-west component of acceleration from the 19 September 1985 Michoacán, Mexico, earthquake from selected sites. Records are shown in order of increasing distance from the rupture area. Sites CUIP, CUMV, SXVI, TACY, SCT1, CDAF, CDAO, TLHD, and TLHB are all in Mexico City. Note the variation in ground-motion characteristics for this earthquake (from Singh et al 1987; copyright Seismological Society of America).

Homework

4-1. Show that the reflection and transmission coefficients match the boundary conditions for an incident planar SH wave onto the interface between two Poissonian solids.

4-2. Write the vector solution for an incident planar SV wave onto the interface between two Poissonian solids. This is similar to equations (4.29) and (4.30).

Post-Critical Planar SH waves

In the last chapter I discussed how to solve the problem of a planar SH-wave incident from a half space onto the bottom of a layer welded on that half space. In this case, there were no post-critical angle reflections in the problem. However, there are instances in which plane waves can travel horizontally such that there is a critical reflection of the wave at the boundary between the layer and the half-space. In this case the waves are totally reflected at the top and the bottom of the layer and it becomes trapped within the layer. This problem is similar to the propagation of planar SH-waves in a plate that was also discussed in the last chapter. However, this problem differs in that the plate problem had traction free boundaries above and below, while the current problem has an elastic half-space below. In order to solve this problem, we must investigate the nature of critically reflected waves. Consider the problem of a planar harmonic SH-wave that is incident on the boundary of two welded spaces, as was shown in Figure 4.2 of the previous chapter. A general expression for such a wave can be written as

$$\mathbf{u}(\mathbf{x},t) = A\mathbf{d}\exp\left[ik\left(\mathbf{x}\bullet\mathbf{p}\right)-ct\right] \tag{5.1}$$

where the motion is either the real or imaginary part of the right hand side of (5.1) . A is a complex number whose modulus is the amplitude of the wave, **d** is a unit vector in the direction of the particle motion, **p** is a unit vector in the direction of propagation (perpendicular to the wavefront), and c is the phase velocity in the direction of **p**. k is the wavenumber, and it is related to the wavelength Λ, period T, and angular frequency ω, by

$$\omega = kc = \frac{2\pi}{T} \tag{5.2}$$

$$k = \frac{2\pi}{\Lambda} \tag{5.3}$$

Using this notation we can write the incident SH wave shown in Figure 5.1 as

$$u_2^I = A_I \exp\left[ik_I\left(x_1\sin\theta_1 + x_3\cos\theta_1 - \beta_1 t\right)\right] \tag{5.4}$$

The reflected wave is

$$u_2^R = A_R \exp\left[ik_I\left(x_1\sin\theta_1 - x_3\cos\theta_1 - \beta_1 t\right)\right] \tag{5.5}$$

and the transmitted wave is

$$u_2^T = A_T \exp\left[ik_T\left(x_1 \sin\theta_2 + x_3 \cos\theta_2 - \beta_2 t\right)\right] \tag{5.6}$$

If the reflection is post-critical, then $\sin\theta_2 = \left(\beta_2/\beta_1\right)\sin\theta_1 > 1$, and since $\cos\theta_2 = \sqrt{1-\sin^2\theta_2}$, then $\cos\theta_2$ is imaginary. In this case, the transmitted wave (5.6) is given by

$$u_2^T = A_T \exp\left(-bx_3\right)\exp\left[ik_I\left(x_1 \sin\theta_2 - \beta_2 t\right)\right] \tag{5.7}$$

where

$$b = k_I\sqrt{\left(\beta_2^2 \big/ \beta_1^2\right)\sin^2\theta_1 - 1} \tag{5.8}$$

(5.7) means that the disturbance into the second medium dies exponentially with the distance from the interface, with longer wavelength harmonic waves disturbing regions further into the medium than do short wavelengths. The reflection coefficient A_R becomes complex with a modulus of 1 for post-critical reflections. That is,

$$A_R = \frac{\mu_1 \cos\theta_1 - \mu_2\left(\beta_1 \big/ \beta_2\right)\cos\theta_2}{\mu_1 \cos\theta_1 + \mu_2\left(\beta_1 \big/ \beta_2\right)\cos\theta_2} \tag{5.9}$$

When $\cos\theta_2$ is imaginary (post-critical reflection), then the modulus of A_R is 1. The fact that A_R is complex means that a phase shift δ is introduced in the reflected wave given by (5.5). This phase shift is a number that depends on the velocity contrast and the incidence angle, but it does not depend on the wavelength. This means that the location of the crests of the incident waves at the boundary are offset by a constant percentage of the wavelength from the crests of the reflected waves as is shown in Figure 5.1. In a sense, it almost looks as if the wave is reflected at some virtual point beneath the boundary, where the depth of the virtual bounce point increases with the wavelength of the wave.

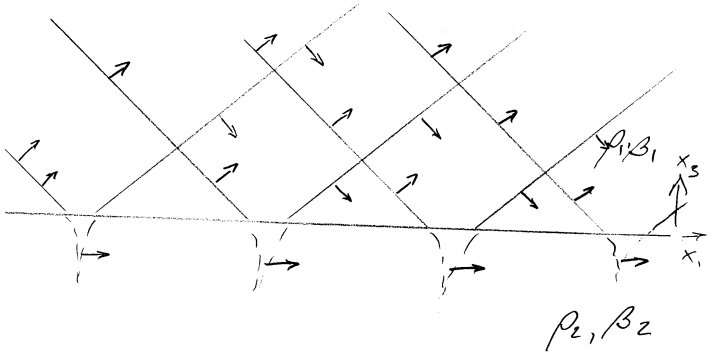


Figure 5.2 Critically reflected planar SH-waves.

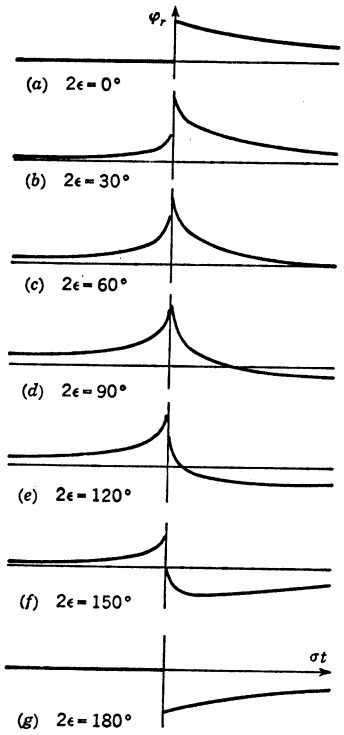


FIG. 3–17. Shape of reflected pulse for several values of phase change 2ϵ, computed from Eqs. 3–50. (*After Arons and Yennie.*)

Figure 5.3 This figure shows the effect of the phase lag introduced by the complex reflection coefficient on an incident wave that consists of a step in displacement. The reflected wave has a very different time behavior. At the time of the expected geometric reflection, the waveform locally looks like it is the time derivative of the incident wave.

While this constant phase shift is relatively easy to understand for a harmonic plane wave, things get more complex if we consider the case of an impulsive wave. Such waves can be considered to be the superposition of harmonic waves. However, the effect of phase-shifting the reflected harmonic waves can be rather dramatic. Figure 5.2 demonstrates the shape of the reflected wave for different phase lags; the top trace is the shape of the incident wave. Notice the rather surprising fact that the reflected wave actually starts at minus infinite time (long before the incident wave begins its motion). This is a rather unusual consequence of the fact that we have assumed a planar incident wave. Such waves can never really exist in nature since they require infinite spaces with waves that travel throughout all time. Nevertheless, there is always a disturbance in the lower medium that precedes a critically reflected incident wave. This is the inevitable result of the fact that there is no wave in the slower medium that can travel as slowly as the phase velocity of the incident wave along the interface.

Diffraction vs. Refraction

Waves that can be fully described by their ray paths (Snell's Law) are generally referred to as **refracted** waves; this is really the same thing as saying that the waves can be considered to be plane waves that are propagating through a plane layered medium without any critical angle reflections. In truth, nothing in the Earth actually does this, and in many cases wave propagation is very different from that of plane waves. Waves that **are not** refracted are called **diffracted**. In a sense, the phase lag introduced in a critically reflected wave is an example of diffraction.

Love Waves

If we now consider the case of SH-plane waves propagating in a low-velocity layer over a half-space, we see that we have a situation that is very similar to the plate-wave problem of Chapter 4, except that we now have a critical reflection to deal with at the bottom of the layer. Although the wave is totally reflected at the bottom of the layer, there is a phase lag associated with the critical reflection. The wavefronts are sketched in Figure 5.3. The phase velocity dispersion relation is similar to that given in the last chapter for a plate, but an additional lag must be introduced to account for the phase shift of the critical reflection. Love waves can only occur if the surface layer has a lower velocity than the whole space.

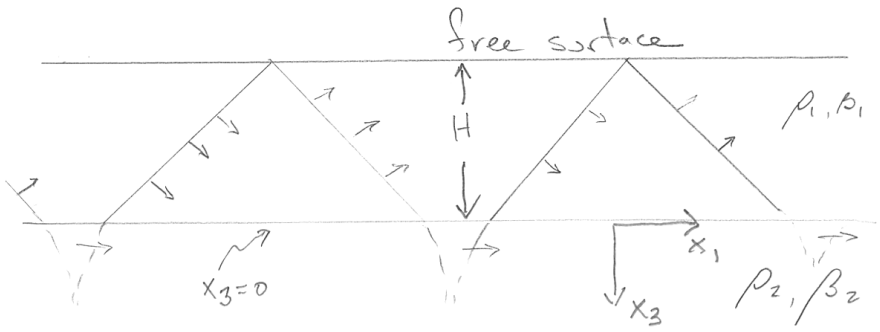


Figure 5.3. A love wave can be thought of as a plane wave that is trapped in a low velocity layer at the surface of the earth.

Love waves were actually observed on seismometers long before they were explained, by A.E.H. Love. Although it is possible to derive the solution as series of critically reflected SH waves, it is also possible to derive the solution to this problem by investigating solutions of the form given below. As it turns out, the solutions for a harmonic Love wave traveling in the x_1 direction in a layer of thickness H overlying a half space as shown in Figure 5.3, are given by

$$u_2 = A_2 \exp\left[-kx_3\sqrt{1-\frac{c^2}{\beta_2^2}}\right]\exp\left[ik\left(x_1 - ct\right)\right] \tag{5.10}$$

in medium 2 (i.e., $z>0$), and

$$u_2 = \left\{A_1 \exp\left[-kx_3\sqrt{1-\frac{c^2}{\beta_1^2}}\right] + A_1' \exp\left[kx_3\sqrt{1-\frac{c^2}{\beta_1^2}}\right]\right\}\exp\left[ik\left(x_1 - ct\right)\right] \tag{5.11}$$

in medium 1 (the layer). $u_1 = u_3 = 0$ everywhere in the medium. Unfortunately, these motions seem rather complex to just pick out of a hat. However, it can be shown that they do satisfy Navier's equation. If $c < \beta_2$, then we have a solution that dies exponentially in amplitude with distance below the interface. Now (5.10) and (5.11) are acceptable solutions to the geometry shown in Figure 5.3 if, i) the motion u_2 is continuous across the boundary at $x_3 = 0$, and ii) the stress σ_{23} is continuous across the boundary $x_3 = 0$, iii), and if the stress $\sigma_{23} = 0$ on the free surface at $x_3 = -H$. By imposing these conditions on (5.10) and (5.11), we can find conditions on $c, A_2, A_1,$ and A_1' which provide an acceptable solution. By following this procedure, one can show that for an acceptable solution, the phase velocity c depends not only on the intrinsic velocities β_1 and

β_2, but also on the frequency of the wave. That is, the wave is dispersive. It is also generally true that $\beta_1 < c < \beta_2$. Furthermore, low frequency Love waves tend to have phase velocities that approach those of the high-velocity half space, whereas short period love waves tend to have phase velocities close that of the low-velocity layer (at least that is true for the fundamental mode Love waves). As was the case for plate waves, there are also higher mode Love waves, but their derivation is more complex because of the critical reflection.

Love waves can be generated by a source located at a point (called a **point source**). In this case, Love waves are observed on the **transverse component** of motion Since Love waves have their motions near the top of the elastic medium, their energy spreads as an expanding circle along the surface. That is the energy flux associated with Love waves must be conserved for any outward traveling wave. Since the circumference of a circle grows as r, and since energy depends on $\dot{u}^2$, the amplitude must decay as $1/\sqrt{r}$ for Love waves radiated by a point source. Thus, as the observer distance becomes larger, the Love waves become larger relative to other waves whose amplitude decays as $1/r$. It is generally true that at large distances, surface waves tend to be larger than body waves.

Rayleigh Waves

The Rayleigh wave is a special solution to the equation of motion; it has the characteristic that it allows zero traction along a boundary. We will discuss the simplest example of a Rayleigh wave; that is, a two dimensional plane Rayleigh wave that propagates at velocity c in the x_1 direction as is shown in Figure 5.4.

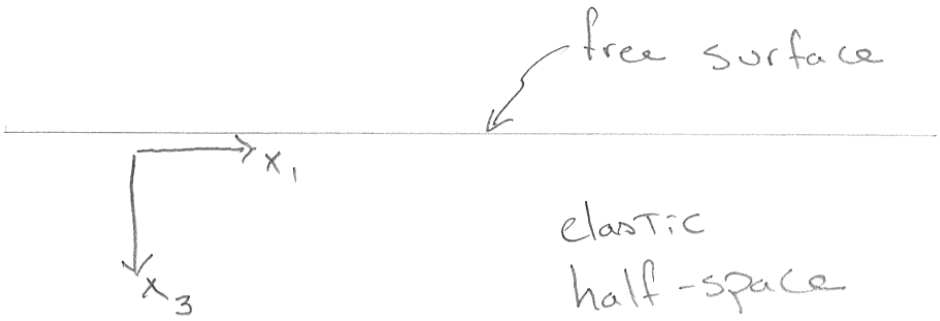


Figure 5.4

The material is an elastic half-space and the surface $x_3 = 0$ is traction free, or $\sigma_{13}\big|_{x_3=0} = \sigma_{23}\big|_{x_3=0} = \sigma_{33}\big|_{x_3=0} = 0$. Consider the motion

$$u_1 = \Re\left\{A\exp(-bx_3)\exp\left[ik(x_1 - ct)\right]\right\} \tag{5.12}$$

$$u_3 = \Re\left\{B\exp\left(-bx_3\right)\exp\left[ik\left(x_1 - ct\right)\right]\right\} \tag{5.13}$$

$$u_2 = 0 \tag{5.14}$$

where b is a real constant, A and B are complex constants, and where $\Re$ means to take the real part of the argument. Keep in mind that

$$\exp\left[ik\left(x - ct\right)\right] = \sin\left[k\left(x - ct\right)\right] + i\cos\left[k\left(x - ct\right)\right] \tag{5.15}$$

This solution represents a sine wave of wavelength $\Lambda = 2\pi/k$, which travels in the x_1 direction with a velocity of c. For this to be a valid solution for our problem, we need to find the appropriate values of b, A, B, and c. Since we are only considering the real parts of the solution in (5.12) and (5.13), we have four unknowns and two boundary conditions at the free surface($\sigma_{23}\big|_{x_3=0} = 0$ is satisfied trivially because of (5.14)). Therefore

$$\sigma_{33}\big|_{x_3=0} = 0 = \lambda\left(\frac{\partial u_3}{\partial x_3} + \frac{\partial u_1}{\partial x_1}\right) + 2\mu\frac{\partial u_3}{\partial x_3} \tag{5.16}$$

and

$$\sigma_{13}\big|_{x_3=0} = 0 = \mu\left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1}\right) \tag{5.17}$$

Furthermore, we must also satisfy Navier's equations, or

$$\rho\frac{\partial^2 u_1}{\partial t^2} = \mu\left(\frac{\partial^2 u_1}{\partial {x_1}^2} + \frac{\partial^2 u_1}{\partial {x_3}^2}\right) + \left(\lambda + \mu\right)\left(\frac{\partial^2 u_1}{\partial {x_1}^2} + \frac{\partial^2 u_3}{\partial x_1 \partial x_3}\right) \tag{5.18}$$

and

$$\rho\frac{\partial^2 u_3}{\partial t^2} = \mu\left(\frac{\partial^2 u_3}{\partial {x_1}^2} + \frac{\partial^2 u_3}{\partial {x_3}^2}\right) + \left(\lambda + \mu\right)\left(\frac{\partial^2 u_3}{\partial {x_3}^2} + \frac{\partial^2 u_1}{\partial x_1 \partial x_3}\right) \tag{5.19}$$

It is actually a rather laborious process to find the appropriate values of b, A, B, and c that satisfy these equations. However, the following solution does satisfy the conditions if the solid is considered to be Poissonian (i.e., Poisson's ratio $= 1/4$, or $\lambda = \mu$).

$$u_1 = D\left(e^{-0.8475kx_3} - 0.5773e^{-0.3933kx_3}\right)\cos\left[k\left(x_1 - c_R t\right)\right] \tag{5.20}$$

and

$$u_3 = D\left(-0.8475e^{-0.8475kx_3} + 1.4679e^{-0.3933kx_3}\right)\sin\left[k\left(x_1 - c_R t\right)\right] \tag{5.21}$$

where

$$c_R = 0.9194\sqrt{\frac{\mu}{\rho}} = 0.9194\beta \tag{5.22}$$

Figure 5.5 shows a schematic of the particle motion for a harmonic Rayleigh wave in a half-space.

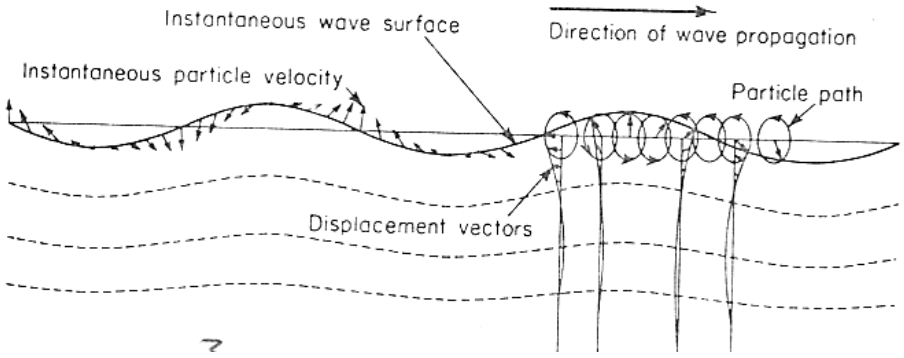


Figure 5.5 Snapshot in time of the particle motion for a harmonic Rayleigh wave in an elastic half-space.

Notice that the particle motion at the free surface is an ellipse for this harmonic wave. That is, the vertical component, which is about 150% larger in amplitude than the horizontal component, is a sinusoid and the horizontal component is a co-sinusoid. That is, the vertical and horizontal components are out of phase by $\pi/2$. Notice that, at the top of the ellipse, the particle is moving in the opposite direction from the direction of the wave propagation. This is referred to as "retrograde particle motion," and it is characteristic of Rayleigh waves. Also, notice that horizontal motions reverse their direction at a depth of 0.192 times the wavelength Λ of the Rayleigh wave (or they have a node at this depth). That is, the horizontal motion is zero at 0.192Λ and the particle motion is prograde at larger depths.

Notice that Rayleigh wave velocities in a homogeneous half space are independent of the wavelength. That is, Rayleigh waves are non-dispersive in a homogeneous half-space. This is different from Love waves, which are inherently dispersive. However, Rayleigh waves are only non-dispersive in a homogeneous half-space. Shear wave velocities generally increase with depth in the Earth. Since long-wavelength Rayleigh waves have motions at a significantly greater depth than do short wavelength Rayleigh waves (remember they die as

$\exp(-z/\Lambda)$), the velocities of Rayleigh waves generally increase with wavelength for the Earth.

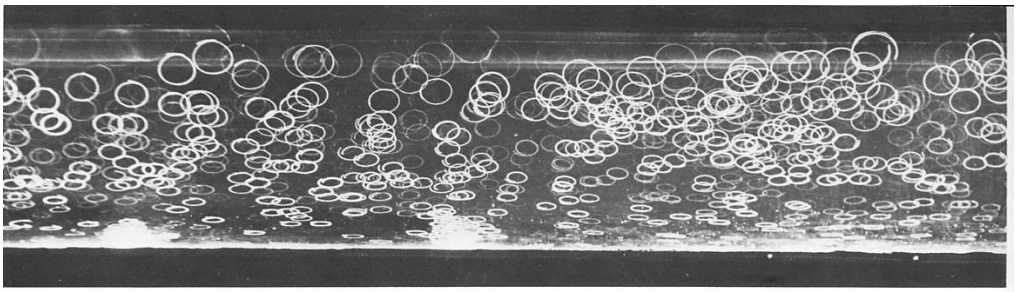
Just as was the case for Love waves, Rayleigh-wave energy generally spreads circularly from a point source and the amplitude of Rayleigh waves generally decrease as $1/\sqrt{r}$ with distance from the point source. In the case of point sources, Rayleigh waves are generally observed on both the **radial and vertical** components of a seismographic station.

The solution demonstrated above is called a fundamental Rayleigh wave, and there are no nodes in the vertical motion as a function of depth (there is 1 node in the horizontal component with depth). The fundamental Rayleigh wave is the only surface wave that can occur in a homogeneous half-space. Love waves require at least one low-velocity layer to be present. In addition to fundamental mode Rayleigh waves, there are solutions to the layered space that are similar to the Love wave case, but which involve P- and SV- waves. If the velocity of the upper layer is low enough so that both P- and SV-waves can be completely reflected (all reflections are post-critical), then there can be plate modes in the P-SV system just as there are in the SH system. These are referred to as **higher mode Rayleigh waves**. Just as with the Love wave case, they can be simplified into a harmonic wave with a horizontal phase velocity that is constant as a function of depth and also another function of depth (pseudo-harmonic) that describes the depth dependence of the mode. If some of the reflections (e.g. SV to SV) are post critical, but others are pre-critical (P to P), then there may be some wave energy that is continually radiated from the low-velocity layer. The system is no long a perfectly trapped system. These are referred to as **leaky modes**.

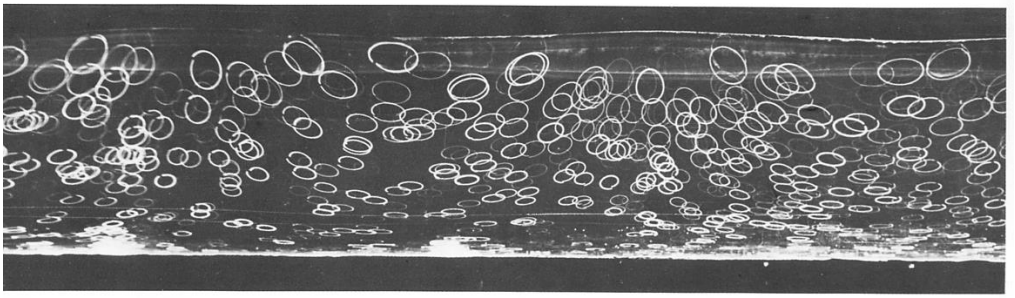
Excitation of fundamental mode Rayleigh waves is rather difficult to imagine. They cannot be generated by planar P- and S-waves incident on a planar free surface. As it turns out, the curvature of either a wavefront, or a free surface is critical to the generation of fundamental Rayleigh waves; the tighter the radius of curvature, the broader the frequency band of Rayleigh waves that can be generated. That means that deep earthquakes can only generate long-period fundamental Rayleigh waves, whereas shallow earthquakes generate broader-band surface waves. However, it is also true that fundamental P- and S-waves are much larger at high frequencies than Rayleigh waves.

This fundamental Rayleigh wave has some mathematical similarities to the solution for gravity waves in the ocean (the type that make you sea sick). However, oceanic water waves have prograde particle motions as is shown in Figure 5.6. As is the case with any two identical wavetrains that are traveling in opposite directions, standing waves are formed. The series of figures show long exposures of gravity water waves traveling towards the right. The white streaks

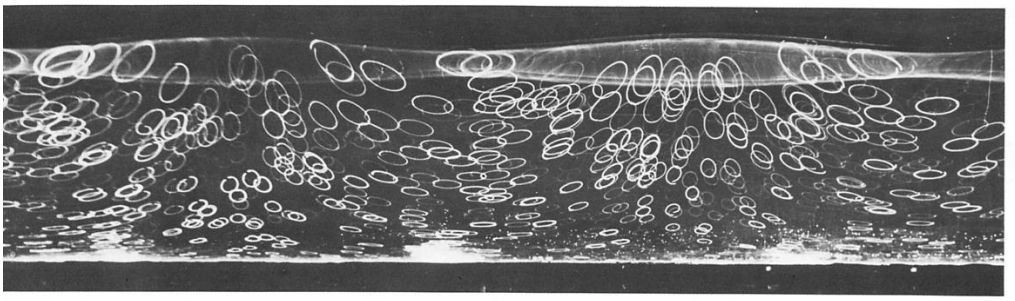
are white particles of neutral buoyancy. The ellipses are their particle motions. In each successive figure more left-traveling wave is added to the mix. In the final figure, one can see that there are purely standing waves. In this case the particle motions are no longer elliptical, but instead become purely linear. Notice that nodes for the horizontal motions are maxima for the vertical motions and vice versa.



No reflection: pure progressive waves

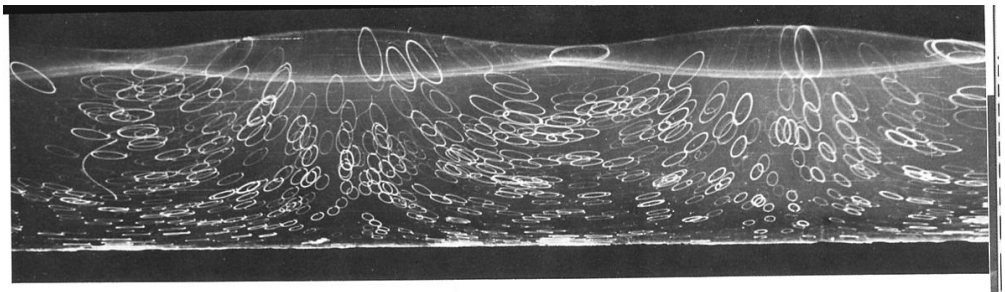


24% reflection

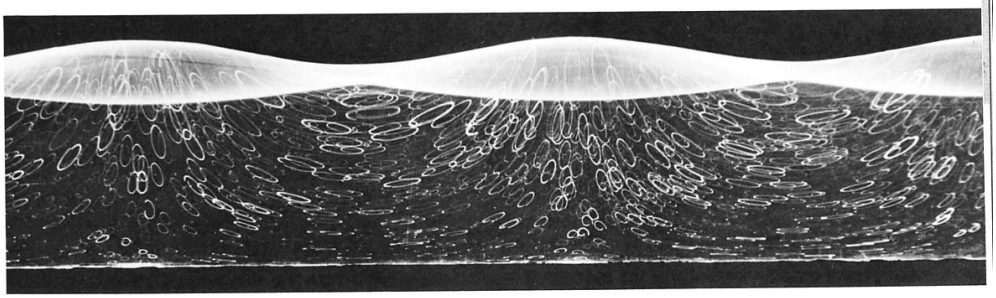


38% reflection

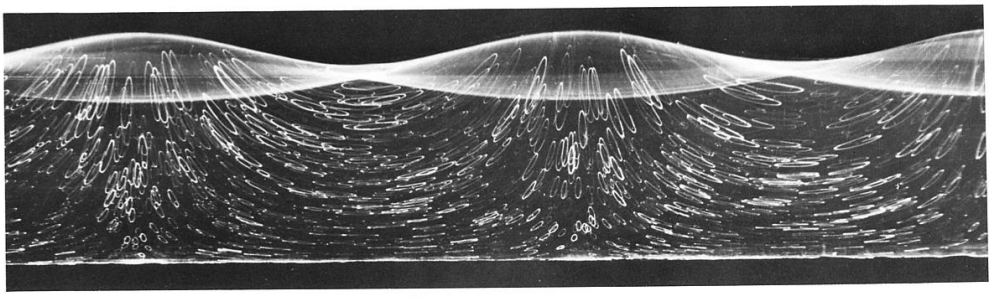
191. Particle trajectories in plane periodic water waves. Two wave trains of the same frequency traveling in opposite directions are produced by a progressive wave coming from the left that is reflected by a partially absorbent barrier. The top photograph shows the pure progressive wave with no reflection. Its amplitude is four per cent of the wavelength, and the water depth is 22 per cent. White particles suspended in the water are photographed during one period. Their trajectories are practically ellipses traversed clockwise, circular at the free surface and flattened toward the bottom. Some open loops indicate a slow drift to the right near the surface and left near the bottom. As the reflection is increased, the orbits become increasingly flattened and inclined. Complete reflection gives a pure standing wave in the last photograph, where the trajectories are streamlines. There the upper and lower envelopes of the water surface show that the vertical motion does not vanish at the nodes. *Wallet & Ruellan 1950, courtesy of M. C. Vasseur*



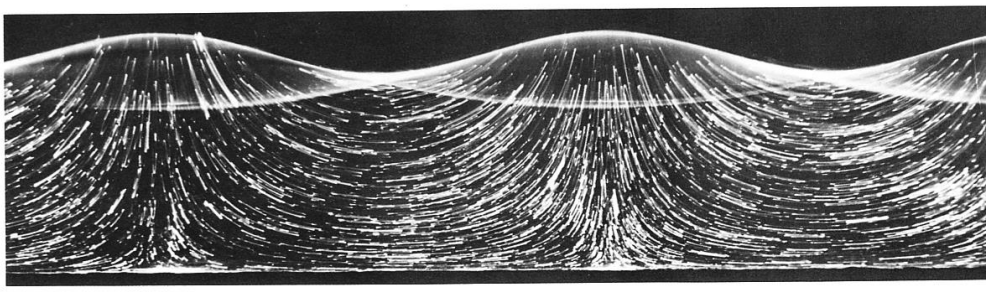
53% reflection



71% reflection



85% reflection



100% reflection: pure standing waves

Chapter 6 Building Response

In this chapter I explore the nature of deformation and forces in buildings during earthquake shaking. Of course, buildings are complex connections of columns, beams, floors and walls, the study of which, deserves an entire course in structural engineering. However, as an introduction, it is instructive to investigate the nature of forces and deformations that would occur in a solid body whose properties are similar to the average properties of a building. I begin with a simple description of different types of buildings and comment on their characteristics in earthquake shaking. In general, I will characterize buildings with the following parameters (refer to Figure 6.1).

- Average density of a building. Density is important because it is used to calculate inertial momentum. The density of buildings ranges from about 100 kg/m^3 (tall flexible frame buildings) to 200 kg/m^3 (stiff heavy shear wall buildings). Earthquake loads in buildings generally increase with the density of the building. These average densities are small compared to material densities because buildings are primarily air.
- Yield strength is the maximum horizontal load that can be applied to a building. It is expressed in units of acceleration if the yield force is normalized by the weight of the building. While increasing yield strength is generally desirable, it usually comes with the penalty of increasing stiffness.
- Stiffness is the horizontal force distributed throughout a building divided by resulting lateral shear strain in the building (usually called drift). Maximum stresses in a building generally increases with stiffness, so making a building stiff can lead to high stresses. While low stiffness has advantages, decreasing stiffness usually comes with the penalties of increasing shear strains and decreasing yield strength.
- Ductility refers to the ratio of the horizontal shear strain at which a building collapses divided by the strain at which a building begins to strain inelastically. Increasing ductility is always desirable, but it usually comes with the penalty of increasing cost.

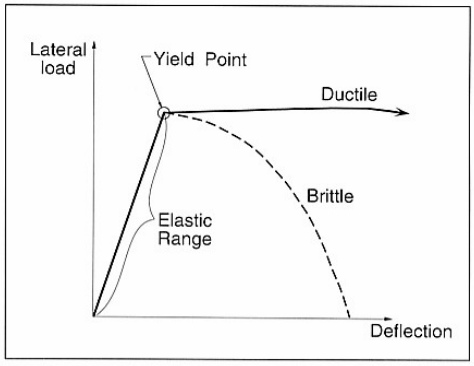


Figure 6.1 Idealized building response.

The following examples give some idea of different classes of buildings.

Concrete Shear-Wall Buildings

This is a common class of buildings that generally have at least some walls that consist of continuous slabs of concrete. These concrete walls are very resistant to in-plane shearing forces. Perpendicular shear walls are generally connected to each other through 1) the strong floor slabs, and 2) sometimes they are connected at corners of rooms. When a building consists of a rectangular concrete box with interior columns supporting the floor slab, then this is generally referred to as box/shear-wall construction. This type of construction is very common at Caltech. It has the advantage of very high yield strength. Furthermore, if the walls are properly reinforced, the ductility is also high. This type of construction has the disadvantages that it tends to lead to very stiff buildings with high average density. As we see later in the chapter, this can lead to high stresses in a building. It also has the disadvantage that the architecture of the building is fixed. That is, walls cannot be reconfigured once the building is constructed. Furthermore, because of their stiffness, it is impractical to build shear wall buildings taller than about 10 stories that also adequately resist earthquake loading.

Figure 6.2 shows two versions of the Olive View Hospital in the San Fernando Valley. The first version was a nonductile-concrete frame building that was completed just prior to its collapse in the 1971 San Fernando earthquake. The hospital was rebuilt as a shear wall structure (some of the shear walls were solid steel) and it experienced heavy shaking in the 1994 Northridge earthquake. In that case, there was no structural damage because of the very high yield strength of the building. I have heard structural engineers criticize this building for being "overbuilt." That said, the overbuilt design ended up saving the County of Los

Angeles a lot of money during the violent shaking in 1994 (pga > 2 g recorded on the roof).

Figure 6.3 shows an example of a Japanese concrete shear-wall apartment building after the 1964 M 7.5 Niigata earthquake (see the figure caption). One-story concrete shear-wall buildings are quite common for construction of mini malls (e.g., 7-11 stores). They are simple to build (typically constructed with steel-reinforced concrete blocks) and they are inexpensive to build. They are extremely resilient to earthquake damage, but they are generally disdained by architects.

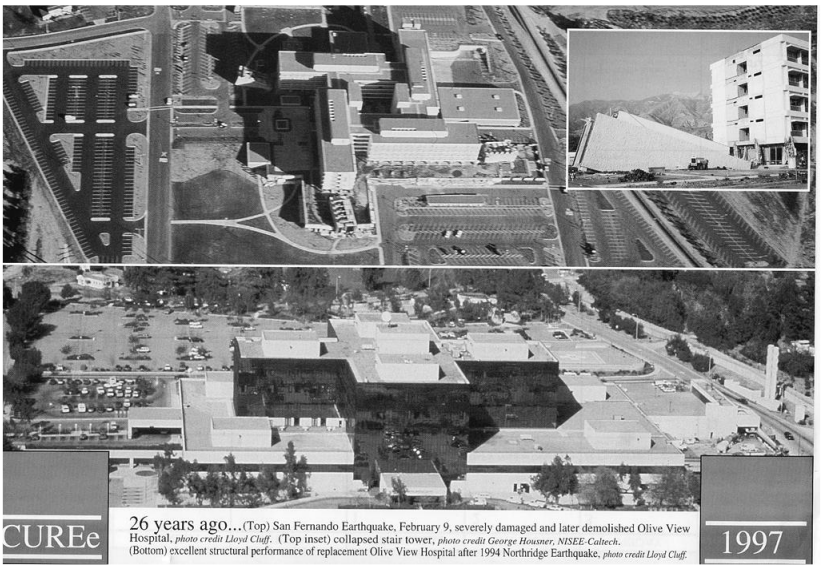


Figure 6.2. Two versions of the Olive View Hospital in the San Fernando Valley. The top picture shows the heavy damage that occurred to the first hospital (non-ductile concrete) in the 1971 M 6.7 San Fernando earthquake. The bottom picture shows the shear-wall structure that replaced the first one. This strong building had no structural damage as the result of the violent shaking in the 1994 M 6.7 Northridge earthquake.

Figure 6.3. Japanese concrete shear-wall apartment buildings after the 1964 M 7.2 Niigata earthquake. Despite the fact that the foundations of the buildings failed due to liquefaction, the building structures were undamaged and the buildings were later jacked back to an upright position and they were reoccupied.

The resilience of concrete shear-wall buildings was severely tested during the 2011 M 7.2 Cucapah-El Major Earthquake. Mexicali is a Mexican city of one million that is just south of the US border. Strong motion records show that the city was strongly shaken. Nevertheless, there were only four fatalities. The excellent performance of Mexicali contrasts with the enormous tragedy of the 2010 M 7.0 Haitian earthquake. There are many similarities between these two earthquakes. They were similar sized strike-slip earthquakes with shallow strike-slip ruptures that were about 15 km from densely populated cities. In the case of the Haiti earthquake, more than 100,000 people died because of the extensive collapse of nonductile concrete frame buildings that were built by amateurs. Although the Mexicali buildings were also constructed by amateurs using concrete, the Mexicali buildings were far more resilient because they made single-story reinforced concrete block shear wall buildings.

Moment Resisting Frame (MRF) Buildings

This is a very common class of buildings, whose structural system generally consists of a rectangular latticework of columns and beams (the frame), together with the relatively rigid floor slabs, which are typically made of reinforced concrete. The columns and beams can be either mild steel (SMRF) or reinforced concrete (RCMRF). Figure 6.4 shows an example of a SMRF. These buildings are popular with many architects since they are 1) inexpensive, 2) office space can be easily reconfigured, and 3) they can be quite tall (The Library Towers in downtown Los Angeles is 80 stories high).

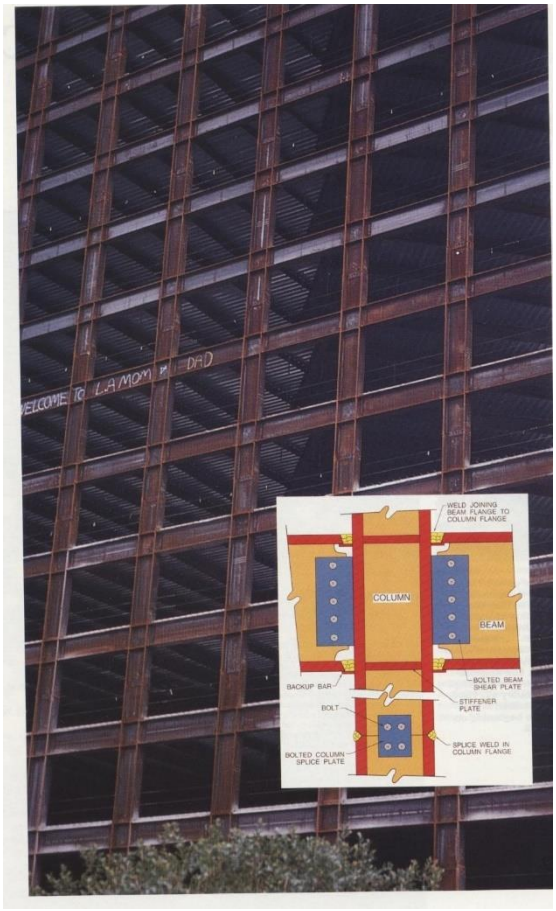


Figure 6.4. Example of a steel moment resisting frame. The connections between the beams and columns are typically welded (called a moment-resisting connection) to keep the elements perpendicular. Many of these critical connections were observed to fracture in the 1994 Northridge earthquake.

Figure 6.5 shows the basic physics of how an MRF resists lateral motion. As the frame is deflected horizontally, the beams and columns must bend if their

connections remain perpendicular. Note that in typical US buildings, not all of the beam-column connections are moment resisting. Many of the interior connections are simple connections, which act structurally more like a hinge. Although simple connections are adequate to support the weight of the floor slabs, they do not cause flexural bending of the beams. The moment-resisting beam-column connections are critical elements of a MRF since that is where the bending moment originates on a beam or column. It is critical that the MRF failure strength exceeds the flexural yield strength of the beams, since a building's ductility (high ductility is good) derives from the inelastic bending of beams (it's not good to inelastically bend columns since they carry gravitational loads).

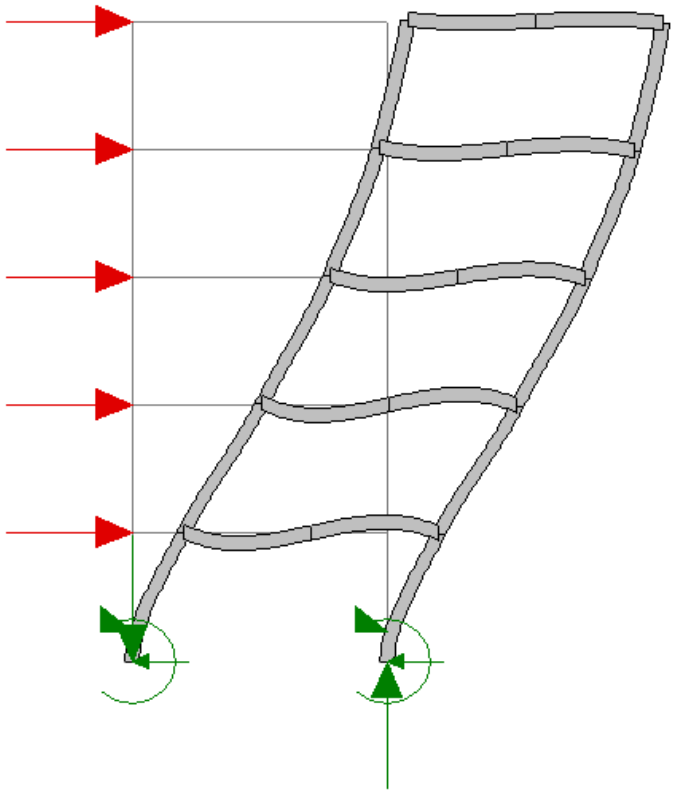


Figure 6.5. Cartoon showing how the flexural bending of beams and columns provides a resistance to lateral deformation for a moment-resisting frame structure. Note that only the connections on the exterior are moment frame connections, whereas the interior connection is a "simple" connection (unwelded) that acts more lie a structural hinge.

In the case of Steel MRF's, the moment resisting column-beam connections typically consist of welds between the flanges of the beams and columns (see Figure 6.4). These welded connections became popular in the 1960's since they

are inexpensive to construct, and they were thought to have high strength. However, many of these welded connections fractured during the 1994 M 6.7 Northridge earthquake, so many of the existing steel MRF's are not as ductile as designers thought when buildings were constructed. Figure 6.6 shows the design of moment-resisting steel beam-column connections that were used prior to the 1960's. For the connection to resist bending, it is critical that the flange of the beam (the top or bottom of the I-section) is firmly connected to the flange of a beam. For buildings with I-sections, that can only be accomplished at ½ of the connections (the connections where the beam flanges abut the column flanges). Prior to the 1960's, the flanges were usually connected to each other using either rivets or bolts.

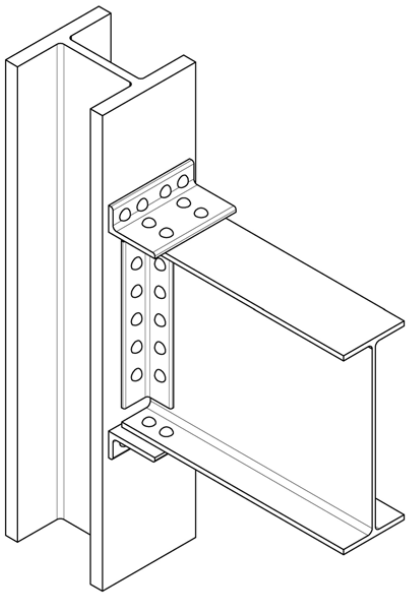


Figure 6.6. Schematic of a steel moment resisting connection that was common in older steel frame construction (pre 1960's). Connections used either rivets, or later, bolts._

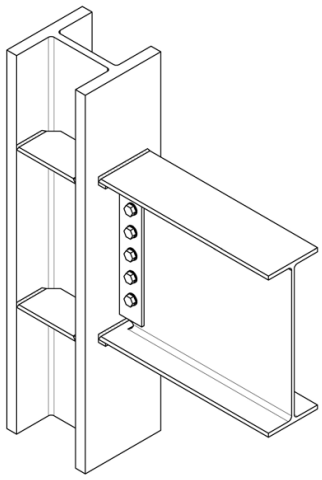


Figure 6.7. Schematic of a welded steel moment resisting connection that was common in steel frame construction (1960's to 1995). The web of the beam is simply connected using bolts. In contrast the flange of the beam is welded to the flange of the beam.

Beginning in the 1960's welded, moment-resisting connections were introduced (Figure 6.7). It was believed that welded connections had a higher strength than either bolted or riveted connections. Importantly, welded connections were less expensive to construct than bolted connections.

The switch from bolted connections to welded connections was partially justified on the basis of lab testing of the relative strengths of bolted and welded connections. It was found that either type of connection was strong enough to cause ductile bending of a beam. That is the desired behavior. However, the lab testing was typically conducted using beams and columns that were significantly smaller than is seen for real high-rise buildings. That is, it would have been very expensive to build a lab test to bend the very large-size steel sections of real buildings. It was believed that the small-scale tests showed that the weld strength exceeded the bending strength.

Critical real-world tests of the performance of welded-steel moment-resisting-connections occurred during the 1994 Northridge and 1995 Kobe earthquakes. For the first time steel frame buildings were shaken strongly enough to cause inelastic structural response. In both cases, inspectors expected to observe plastically bent steel beams. To their horror, they did not find any bent steel, but they did find numerous examples of fractured welds (see Figure 6.8). At first, researchers suspected poor welding. However, when full-scale welded connections were tested in the lab, the full-scale tests showed that the welds fractured before there was enough force to bend the steel. This result happened even when the welds were "high quality."

Figure 6.8. Example of a welded steel connection that fractured in the 1994 Northridge Earthquake. Notice that the weld between the flanges experienced a tensile crack that extended across the web of the steel column. Prior to this earthquake, it was generally believe that mild structural steel would ductily bend, but that it would never fracture. .

It now seems that there is a size effect. That is, the forces required to fracture the welds did not scale with the cross-sectional area of the beams and columns. This size effect is complex and interesting. I will come back to the effect of size on the failure of materials in Chapter 8.

Steel moment-resisting connections were extensively modified in the latter 1990's and Figure 6.9 shows an example of a modern bolted and welded moment-resisting connection. Unfortunately, it is very expensive to "fix" brittle welds in existing buildings. Some have argued that it is only necessary to fix welds if they have previously failed in some past earthquakes. However, the fact that plastic yielding of steel was never observed for buildings in Kobe and Northridge indicates that all pre-1994 steel welds should be repaired since they are unlikely to survive the forces necessary to cause plastic bending of the beams.

Figure 6.9. Post Northridge Special Moment Resisting connection (welded and bolted) in one direction and simple connection in the orthogonal direction

While steel MRF's have the advantage that they are very flexible, that comes with the penalty that they have very low lateral strength. Figure 6.10 shows a **pushover analysis** (finite-element analysis by John Hall) of a 20-story SMRF building that meets the 1992 Uniform Building Code (UBC) code for California. This analysis included numerous nonlinear effects on the deformation of the steel, as well as also explicitly including the effect of how gravitational forces act on the building for large finite displacements. That is, when the drift of the building becomes large, then ever-increasing lateral loads are put on the building by gravity (kind of like the Tower of Pisa). This is known as the **P-Δ** effect and it is an important collapse mechanism for buildings.

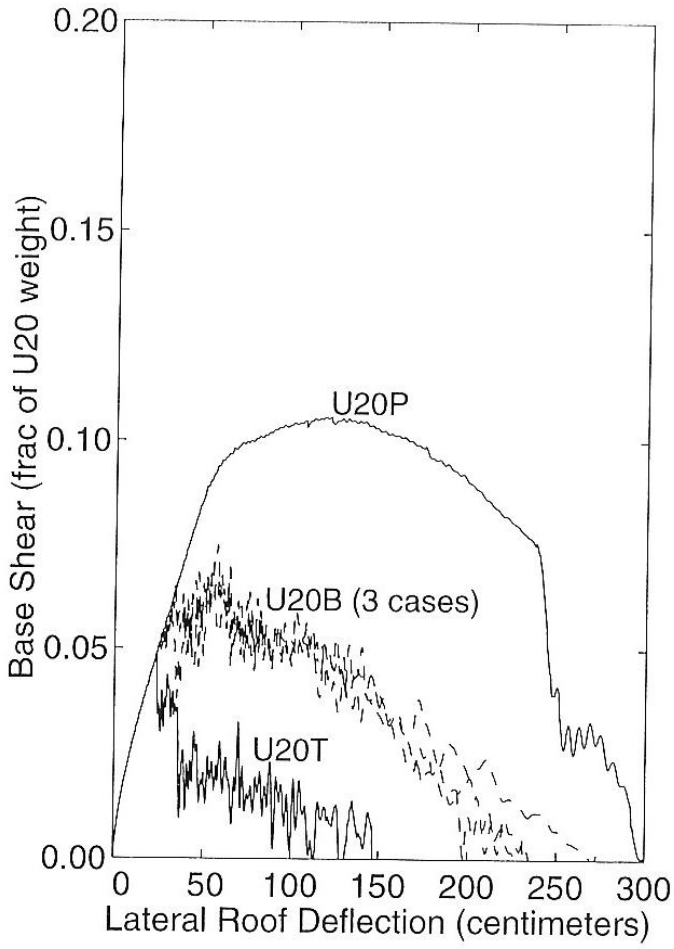


Figure 6.10 (from John Hall). Finite-element pushover analysis of a 20-story building that meets the 1992 US code standards for zone 4. P refers to the assumption that the moment frame connections do not fracture, B assumes that weld fractures occur randomly at stresses compatible with what was observe in the Northridge earthquake, and T assumes that the welds had even less fracture resistance.

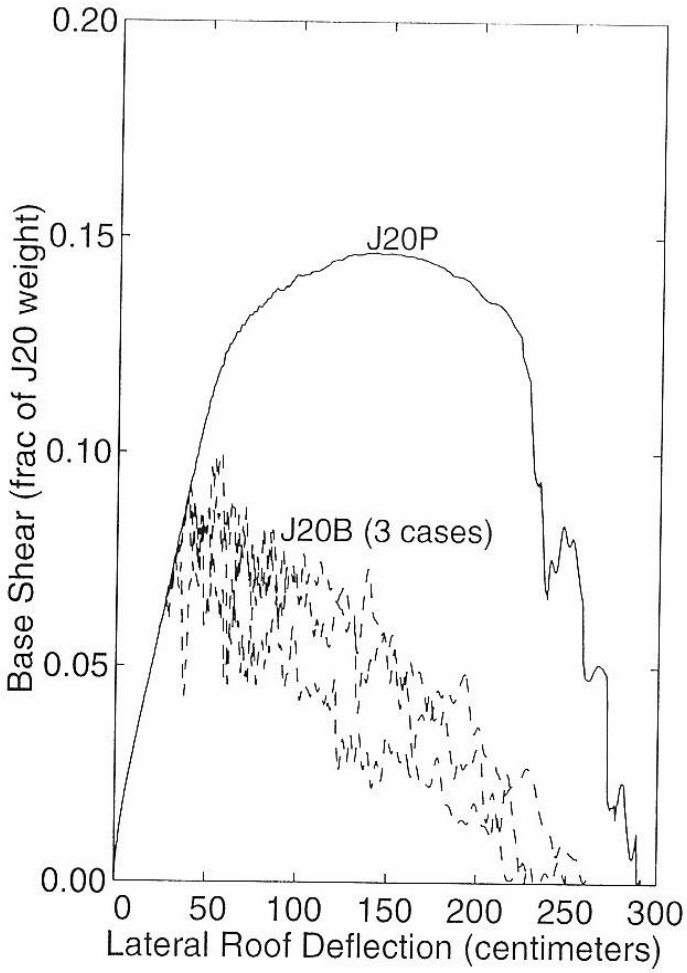


Figure 6.11. (from John Hall) Same as Figure 6.10., except for a 20-story steel frame building that meets Japanese codes in place in the 1990's. Notice the higher yield strength compared to the US building.

The curves U20P refer to a 20-story building that meets US 1992 zone 4 codes (highest code level), and for which the welded-moment resisting connections behave perfectly (no failures). The curve that is designated as B refers to allowing failure of the moment resisting connections assuming weld behavior consistent with observations in the 1994 Northridge earthquake. T refers to the assumption of "terrible" performance of the welds. Notice that weld failure significantly decreases both the yield strength and the ductility of the structure. Also notice that a horizontal force of only 7% of the weight of the building is necessary to push over a typical 20-story building in high seismic risk areas of the US.

Figure 6.11 shows a similar analysis, but it assumes that the building meets the building code used in Japan in the 1990's. Japanese construction tends to put

more emphasis on the yield strength of a structure, and it is common that all connections in a Japanese structure are moment-resisting connections (more costly than the US). Japanese steel buildings are commonly constructed with box section columns. This means that it is feasible to make all beam-column connections into moment-resisting connections.

As it turns out, the code required yield strength tends to increase as building height both increases and decreases from 20 stories. This is because design forces for wind loads increase as the building becomes taller, whereas design forces for earthquakes decrease as the building becomes taller (we'll visit this later). So buildings shorter than 20 stories are designed for earthquake loads and buildings taller than 20 stories are designed for wind loads. It is important that a building remains elastic for wind loads since wind loads are not oscillatory. If a building began to yield in a strong wind, then the force would remain, and the building would just blow over. In contrast, it is assumed that earthquake loads are oscillatory; that is, the building deformation will reverse direction before the yielding becomes too large to survive **P-Δ** effects.

Figures 6.12 and 6.13 show the pushover analyses of 6-story steel moment resisting frame buildings for 1990's US and Japanese codes, respectively. Notice that the 6-story buildings are required to have a greater yield strength than the 20-story buildings.

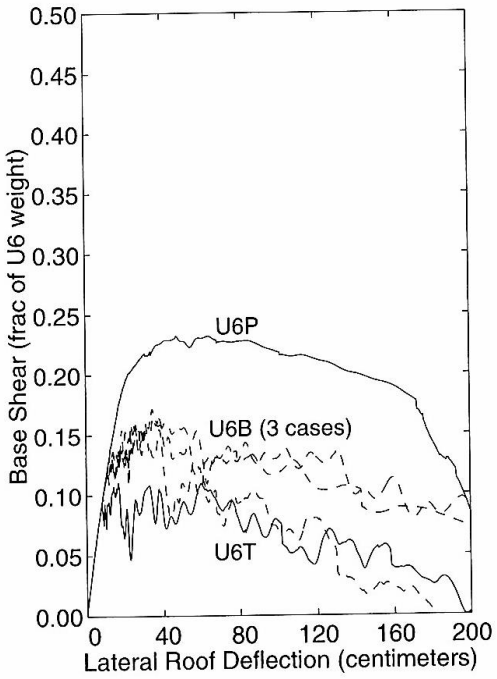


Figure 6.12 (from John Hall) Same as Figure 6.10, but for US-code, 6-story, steel-frame building.

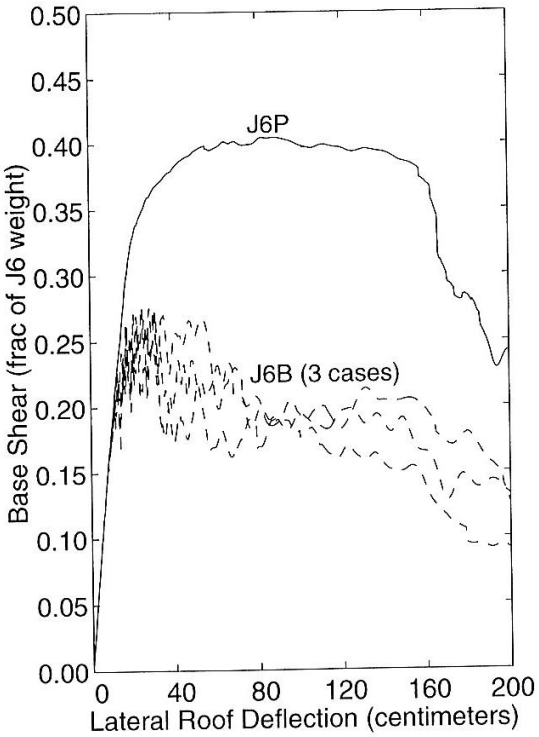


Figure 6.13 (from John Hall). Same as Figure 6.10, but for Japanese-code, 6-story steel frame building

Concrete Moment-Resisting Frame Buildings

Moment resisting frame buildings can also be constructed with reinforced concrete beams and columns. Concrete mrf's have similar flexibility to steel mrf's and the code requirement for lateral yield strength is also similar. Both types of mrf's (steel and reinforce concrete) are required to have high ductility (approximately a factor of 10), but this is achieved in different ways with concrete. While steel is naturally ductile in tensional strain, unreinforced concrete is naturally brittle in tension. Although it is common to speak of the shear strength of the solid parts of the Earth, beams and columns are slender and they generally carry only longitudinal forces. In many cases, this can be axial tension or compression. In other cases, the element is subjected to bending forces (think of the forces in a diving board). In the case of bending, one side of the beam experienced tension while the opposite side experiences compression. In order to construct concrete beams that support bending (e.g., a beam supporting a floor), steel reinforcing bars (rebar) are run longitudinally in concrete beams in order to greatly increase the tensional strength and ductility. While longitudinal rebar is very important, it is not sufficient to make a beam adequately ductile. This was discovered through the inspection of reinforced concrete beams and columns that failed in the 1971 San Fernando earthquake. When a column experiences brittle

failure, then the building collapsed vertically (referred to as a vertical collapse mechanism). An example of this type of failure is seen in Figure 6.14, which shows the failure of an 8-story concrete frame building in Mexico City from the M 7.2 1957 Acapulco earthquake. A more revealing photo of non-ductile failure of a concrete column is shown in Figure 6.15, which shows a freeway bridge column that failed during the 1994 Northridge earthquake. Notice that the column originally fractured because of horizontal shear loads on the column. Although the failure was caused by shear stress in the column (this happens when the thickness of the column increases relative to its length), the failures actually consist of diagonal tensile cracks.

Once the concrete in the column cracked, the concrete fell away from the column and the remaining rebar buckled into a mushroom shape. This is an example of **non-ductile concrete** behavior. This deficiency was rectified by requiring spiral reinforcing bars that serves to confine the concrete to the beam, even if it is fractured. Figure 6.16 shows how a concrete column can continue to carry significant loads even though it has been strained well beyond its yield point. Unfortunately, this parking garage suffered significant collapse because the elements of the building were insufficiently connected to each other. That is, the reinforcing bars must adequately tie the different elements together.

Figure 6.14. Collapse of an 8-story non-ductile concrete moment-resisting frame building in Mexico City from the 1957 Acapulco earthquake.

Non-ductile concrete frame buildings are recognized as a class of particularly dangerous structures. They have the particularly bad combination of having a low yield stress combined with a low ductility (they're brittle). The tremendous loss of life in the 1999 Izmet Turkey earthquake was an example of non-ductile concrete frame failures. These failures are often very disastrous since the building often pancakes into a pile of floor slabs (referred to as a vertical collapse mechanism; very nasty). Many concrete moment resisting frame buildings that were constructed in the United States prior to 1975 can also be classified as non-ductile concrete frames. Failures in the 1971 San Fernando earthquake resulted in a building code change in 1975 that significantly enhanced the ductility of buildings built after 1975. Unfortunately, repair of non-ductile concrete frame buildings is so expensive that there are very few ordinances that force a building owner to strengthen these buildings. Furthermore, most of the occupants of these buildings are not aware of the potential deficiencies of their building. Notably, the cities of Los Angeles and Santa Monica have passed laws that direct the Building Departments to publish lists of these buildings in those cities. Furthermore, the law directs owners to upgrade the buildings on a specified schedule that prioritizes actions according to the occupancy of the buildings. This work is not required to be completed for several decades.

It's important to distinguish concrete frame buildings from concrete shear wall buildings. That is, there are many old concrete shear wall buildings that are very resistant to failure in earthquakes, while most older concrete frame buildings may perform poorly in strong shaking.

Figure 6.15. Example of a nonductile concrete column failure on a freeway bridge during the 1994 Northridge earthquake. The column was fractured by horizontal shear, the concrete fell away, and then the weight of the bridge deck caused the rebar to buckle. This failure could have been avoided by adding more spiral reinforcing loops radially around the column to make it more ductile as is seen in Fig. 6.16.

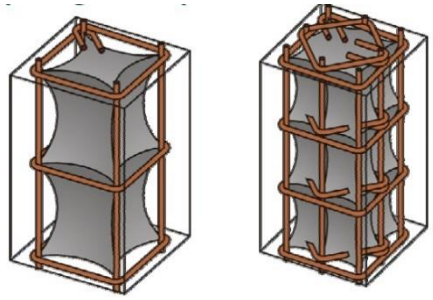


Figure 6.16. Schematic showing how the addition of additional circumferential reinforement increases the strength of concrete beams. This is referred to as "detailing."

Figure 6.17. Example of ductile deformation of concrete columns from the 1994 Northridge earthquake. Adequate spiral reinforcing resulted in more ductile behavior than was shown in Figure 6.14. Unfortunately, the structure had other inadequacies that led to collapse as is shown in Figure 6.17.

Figure 6.18. Despite the ductile behavior of the concrete columns, this parking structure collapsed because the floor slabs were not adequately connected to the rest of the structure. That is, the beams and columns were ductile, but the connections between these elements were not.

Braced Frame Structures

The lateral yield strength of a building can be increased by adding diagonal braces to a structure, as is shown in Figure 6.19. Recall that the restoring force of a beam is primarily bending. In contrast the restoring force for a diagonal brace is primarily uniaxial compression (or tension, depending on the direction of the inter-story shearing). Beams are much stiffer in uniaxial deformation than they are in bending. In general, diagonal braces increase the pushover yield strength, while they also increase the overall stiffness of a building. That is, there is a trade-off between the desired trait of high strength and the undesired trait of high stiffness. Furthermore, it can be difficult to make a braced frame that has high ductility. This is because the use of large bracing elements tends to result in very stiff braces that apply very large loads to their connections with the structure, thereby concentrating damage at these connections. However, the use of small diameter bracing elements can end up with braces that tend to have ductile extension (good), but they buckle in compression (bad). As a building undergoes cyclic loading, small braces become ineffectual, since they permanently extend for motions in one direction and then they buckle for motions in the opposite

direction (see Figure 6.20). Caltech's Broad Center is one of the first buildings in the United States to use a new style of brace called an unbonded **bucking-restrained brace** (**brb**). This consists of a small diameter steel brace that is jacketed in a concrete liner. There is a lubricating element between the concrete and the steel. The concrete jacket prevents the brace from buckling in compression and hence this brace is ductile in both extension and compression. Figure 6.21 shows an example of Broad Center's unbonded buckling-restrained braces.

Figure 6.19. Example of a braced steel-frame.

Figure 6.20 Large braces are stiff and they put large loads into a frame, but small braces can buckle in compression.

Unbonded Brace in the Broad Center

Figure 6.21. Broad Center bucking-restrained brace. The steel is ductile in tension and compression since it is jacketed by concrete (yellow) to prevent buckling in compression.

Base Isolated Structures

Throughout my career (starting in 1972), I have been asked about a common folk lore that buildings (especially high-rises) are built on rollers. I don't know where

this story originated, but I have never encountered any buildings on rollers. That said, the use of base isolators has become ever more popular (especially in Japan) since they were first introduced in New Zealand in the 1980's. The idea is simple; introduce a highly flexible layer between the ground and a building. Loads to the building are dramatically reduced by concentrated shearing at the isolators. The first practical designs consisted of rubber disks placed between the foundation and the ground floor (see Fig. 6-22). In order to ensure that the rubber disks could carry the vertical loads from the weight of the building, stainless steel disks were molded into the isolators. In some cases, the disks also had a lead plug at their center to provide plastic damping. As you can see in Fig.6-22, the diameter of the isolator must be comparable to the largest lateral design displacement. The isolator of Fig 6-22 is relatively large (150 cm??) and it is intended to accommodate differential displacement of up to the size of the gap between the ground floor and the foundation. It is typical to construct a massive reinforced concrete box that is buried one story into the ground. The walls of the box prevent the differential motions from becoming so large that the isolator either tears or becomes gravitationally unstable. These walls are referred to as safety stops. Obviously, it is undesirable for the building to exceed the gap displacement in a real earthquake.

Figure 6-22. Base Isolated structure. The large black cylinder is one of many rubber/steel plate composite isolators that supports the weight of the building. The isolators are stiff in compression and flexible in shear. The concrete wall behind the engineer is the safety stop. The ground floor of the building is a massive reinforced concrete element that ensure that the ground floor moves as a rigid block.

There are some tradeoffs in the design of isolated buildings. It is desirable that the isolators can withstand large isolator drifts, which means that the isolators should have large diameters. However, such large rubber disks are surprisingly stiff and they come with the penalty that there is less isolation. Making the isolators taller (thicker) means that they are more flexible, but it also means that they become vulnerable to $P-\Delta$ instability.

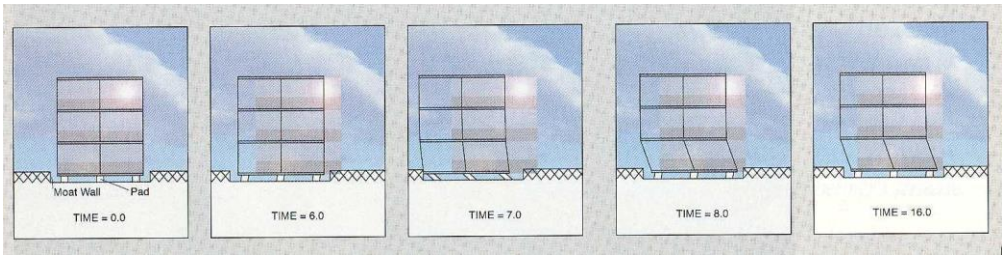


Figure 6-24. Simulation of a typical three-story base-isolated building in simulated near-source ground shaking from a hypothetical M 7.0 thrust earthquake.

The earliest generation of isolators typically had maximum differential displacements of about 35 cm. Following a paper by Heaton, Hall, Wald and Halling (1995, Response of high-rise and base-isolated buildings to a hypothetical M 7.0 blind thrust earthquake, Science, V 267, 206-211) that pointed out that near-source ground motions from M > 7 earthquakes could exceed these design limits (see Figure 6-23), there was a move to increase the design drift of isolated buildings (see Figure 6-24).

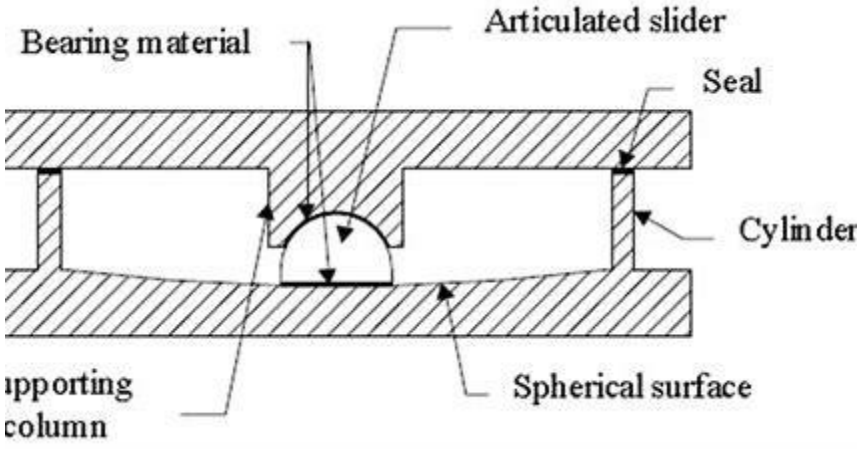


Figure 6-24. Schematic of a friction-pendulum bearing for base isolation. The sliding surfaces of the bearing are typically made of stainless steel that is coated to reduce sliding friction. The lateral restoring force for this system is gravity. The motion of the buildings

is identical to that of a pendulum of the length of the radius of curvature of the spherical surface.

A new type of isolator was developed; a friction-pendulum isolator consists of a metal bearing that slides in a low-friction hemispherical cup. If the radius of curvature of the isolator is large, then the isolator can accommodate large differential displacements. Unlike the rubber base isolated buildings that only displace laterally, the motion of the pendulum-isolated building is the same as the motion of a pendulum whose length l_p is the radius of curvature of the isolator surface. Recall that the free period of a pendulum is $T_0 \approx 2\pi\sqrt{l_p/g} = 2.0\sqrt{l_p}$, where the pendulum length is in meters and the period is in seconds. Now the maximum lateral displacement of the pendulum is just $x_{\max} = l_p \sin\theta_{\max} \approx l_p\theta_{\max}$, where $\theta_{\max}$ is the maximum rotation of the pendulum in radians. Therefore, $T_0 \approx 2.0\sqrt{x_{\max}/\theta_{\max}}$. It is clear that the maximum rotation of the pendulum cannot be more than about 0.1 radians (1.57 radians corresponds to rotating the base to a vertical orientation). Assuming that $\theta_{\max} = 0.1$, then $T_0 \approx 6.3\sqrt{x_{\max}}$. This means that designing for large bearing displacements implies that the isolator system has large natural periods; an isolator maximum displacement of 1 m has a free period of about six seconds. As we will see in the next chapter, the maximum isolator displacement grows with isolator period for near-source ground motions in large earthquakes ($pgd \sim 10^{M/2}$). In reality, it is difficult to estimate the maximum isolator displacement that is expected at a given site. Nevertheless, it is currently common for practicing engineers to ignore this issue when discussing base isolated buildings with the public and policy makers.

As long as the displacement of the ground is less than the maximum isolator displacement, then base isolated buildings are expected to have the benefit that the accelerations in the building will be substantially reduced compared to the ground acceleration. This is especially beneficial for historic buildings that are both fragile and socially important. In particular, the city halls of San Francisco, Los Angeles, Pasadena, and Salt Lake City have all been retrofitted with base isolation systems. Installing base-isolation in an existing building is remarkably expensive; it's probably less costly to build a new replica than it is to isolate an existing historic structure. Convincing political decision makers to commit public funds for these retrofits means that there is reluctance to admit that the systems will not survive large near-source ground motions.

Another benefit of base isolation is that the contents of isolated buildings (e.g., an expensive computer system) are less likely to be destroyed than for the same system in a traditional building since the peak accelerations in the building are typically reduced (that is, unless the building collides with its safety stops).

Wood-Frame Structures

Wood frame is the most common type of construction in California; most residences and many commercial structures are in this category. Since wood is thought of as a "flexible" material, you might think of a very flexible building when you think of a wood frame structure. This would be a mistake. In fact, most wood frame construction is extensively braced. Furthermore, continuous plywood panels, and sheetrock panels are typically fastened to either side of the wooden framing. Such walls may be best described as shear panels. These panels are geometrically connected into rectangular box shapes. In this sense, most wood frame construction might be better described as a shell type of structure. Another feature of wood frame construction is that the structure is relatively light; the dead load of the building is small compared with the weight of the contents (the live load). Because of the higher stiffness to mass ratio, wooden houses have high natural frequencies (typically 5 to 7 Hz).

Since it is not workable to construct houses that would crack their plaster walls when subjected to live loads (e.g., a piano, a graduation party), wood houses are also designed to have large yield stresses (compared to their weight). In fact, the pushover yield stress of most houses exceeds the weight of house. Furthermore, wood houses are designed to be stiff to limit the deflections caused by the live loads.

Wooden houses are also relatively ductile (the framing is redundant, and nails must be pulled out to disconnect elements). Because of these features (stiff, strong, and ductile), wood frame structures tend to perform very well in earthquakes. Even though these structures have been located in areas of violent shaking in past earthquakes, collapse of these structures is exceedingly rare. Figure 6.22 shows the Turnagain Heights housing development (wood frame) following the 1964 Alaskan Earthquake. Despite the tremendous damage caused by a massive landslide beneath the development, the wood frame houses essentially remained intact.

1964 Alaska Earthquake, Turnagain Heights

Figure 6.22. Wood frame houses that rode through the massive landslide triggered by the 1964 Alaskan earthquake.

Quantification of ductility of wooden houses is a little tricky since it involves some understanding of how damage occurs. When wooden houses do yield, the damage is most often concentrated in the region between the foundation and the floor of the first story. In California, this region is often called a cripple wall. I don't want to get into the details of construction styles, so if you want to speak about any particular house, then you had better go to more detailed documents (like the wood retrofit report of the Calif. Earthquake Authority).

Unreinforced Masonry

In the earlier part of the 20th Century, many buildings were constructed of unreinforced brick; that is the exterior walls are several courses of brick and mortar, whereas the inner walls, floors, and roof are wood frame construction. The exterior brick walls in this type of construction tend to be heavy and brittle. That is, the walls cannot sustain tension. URM's have the undesirable characteristics that they are heavy, stiff, and quite brittle. The inadequacies of unreinforced masonry (URM) became obvious in the 1933 Long Beach earthquake and many municipalities adopted building codes (between the mid 1930's and 1950, depending on the city) that required that these masonry construction buildings should be reinforced with steel. However, several cities have numerous examples of these historic structures. Following serious damage to URM's in the 1971 San Fernando earthquake, the cities of Los Angeles and

Long Beach adopted controversial legislation that required that all URM should have some strengthening. At a minimum, this involved making stronger connections between the wooden floor trusses and the brick walls. This tends to decrease the bending moments on the base of the brick walls for out-of-plane shaking. Some buildings are also reinforced by building another structural system (often steel) within the building. Although strengthened URM's are an improvement on the pre-existing structures, there is a widely held belief that they are still lacking in strength and ductility. Despite their obvious shortcoming, the interior walls of URM's often prevent the catastrophic pancaking of the floors seen in non-ductile concrete frame buildings.

While almost all un-retrofitted masonry buildings constructed prior to 1930 are probably dangerous for even moderate levels of shaking, modern steel reinforced masonry buildings are thought to be among the most resilient structures available. That is, if you see a brick building, don't immediately assume that it is vulnerable. In California, most brick buildings constructed since 1950 are reinforced with steel and they are best classified as reinforced concrete shear wall structures.

Table 6.1 is a list of different types of buildings along with a general assessment of the structural characteristics of each type. This table is very generalized and qualitative. The actual characteristics of real buildings vary tremendously (caveat emptor).

Figure 6.23. Example of an unreinforced masonry building (URM).

Table 6.1 Qualitative summary of the characteristics of different building types.

Building Type	Stiffness	Average density	Yield Strength	Ductility

Shear Wall	high	high	high	medium to high
Wooden House	high	low	high	high
Braced Frame	medium	low	medium	medium to high
MRF (frame)	low	low	low	Medium to high
Nonductile frame	low	low	low	low
Unreinforced Masonry	high	high	low	low
Base Isolated	low	variable	variable	Variable

Understanding Building Response

The next sections provide some methods for understanding the response of buildings in earthquakes. This is an enormous field of study and the analysis in these chapters is simply an introduction to the subject.

Building as a Rigid Block

Buildings are not rigid blocks! However, it is still instructive to investigate the forces in a rigid block that is subject to ground acceleration. This example has some application if the lowest natural frequency of the structure is high compared to the predominant frequency of the ground acceleration.

If both the building and the ground are considered to be rigid (how do you have earthquakes in a rigid earth?), then there are no waves and the problem can be solved by balancing force as follows. Consider a rigid building of height h, length and width w, and average density ρ that experiences a horizontal acceleration $\ddot{u}(t)$ as shown in Figure 6.24.

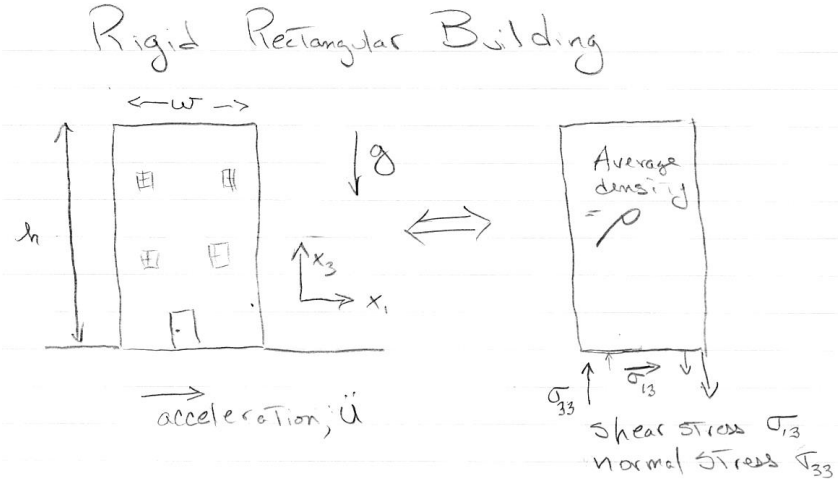


Figure 6.24. Forces acting on a rigid rectangular building that is sitting on a rigid earth. The cartoon on the right is known as a free body diagram. There are both shear stresses to horizontally accelerate the building, and also a moment that is applied to the base to keep the building from rotating.

The total momentum of the building is the sum of the translation of the center of mass of the building and also the rotational momentum of the building about an axis running through the center of the building. Since the ground is rigid, the building cannot rotate and the total momentum of the building is just

$$P_1(t) = \rho h w^2 \dot{u}(t) \tag{6.1}$$

where $P_1(t)$ is the momentum in the x_1 direction (that's all there is in this problem). The total horizontal force on the bottom of the building is

$$F_1(t) = \dot{P}_1(t) = \rho h w^2 \ddot{u}(t) \tag{6.2}$$

Therefore the shear stress on the bottom of the building is just the force divided by the cross sectional area, or

$$\sigma_{13}(t) = \rho h \ddot{u}(t) \tag{6.3}$$

So the shear stress at the bottom of a rigid building on a rigid earth just depends on the ground acceleration and the height of the building (assuming that the density is constant).

However, there is more to this simple problem. The shear stress at the bottom of the building would cause the building to rotate if there were no counteracting forces on the base of the building. That is the total moment applied to the base of the building must be zero, or

$$F_1(t)\frac{h}{2} - w\int_{-\frac{w}{2}}^{\frac{w}{2}} \left[\sigma_{33}(t,x_1) - \rho g h\right] x_1 dx_1 = 0 \tag{6.4}$$

where we assumed that the normal force on the base of the building consists of the weight of the building **plus** a moment that keeps the building form rotating. If we assume that the normal stress consists of constant compressional stress from the weight of the building $(\rho g h)$ plus another stress that varies linearly with distance along the base, then

$$\sigma_{33}(x_1,t) = c(t)x_1 + \rho g h \qquad -w < x_1 < w \tag{6.5}$$

where $c(t)$ is now only a function of time. In this case

$$F_1(t)\frac{h}{2} - wc(t)\int_{-\frac{w}{2}}^{\frac{w}{2}} x_1^2 dx_1 = 0 \tag{6.6}$$

Substituting (6.2) into (6.6) and performing the integration yields

$$\frac{\rho h^2 w^2 \ddot{u}(t)}{2} - \frac{w^4}{12}c(t) = 0 \tag{6.7}$$

or

$$\sigma_{33}(x_1,t) = \rho g h + 6\rho\left(\frac{h}{w}\right)^2 x_1 \ddot{u}(t) \qquad -w < x_1 < w \tag{6.8}$$

Therefore the normal stresses at the outer edges of the building are

$$\begin{aligned}\sigma_{33}(x_1 = \pm w,t) &= \rho g h \pm 3\rho\left(\frac{h}{w}\right)^2 w\ddot{u}(t) \\ &= \rho h\left[g \pm 3\left(\frac{h}{w}\right)\ddot{u}(t)\right]\end{aligned} \tag{6.9}$$

There are cases where engineering materials fail easily in tension. Tensional stresses in our rigid building occur at the outer edge when

$$\ddot{u} > g\,\frac{w}{3h} \tag{6.10}$$

Therefore, rigid buildings that are tall compared with their width can result in tensional stresses in the exterior part of the building.

If we can approximately model the dynamics of a building with a rigid block, then the stresses in the building are determined by the **peak acceleration**. However, it has been long known that the peak accelerations observed in earthquakes are

considerably larger than the nominal lateral strength of buildings that have survived those ground accelerations. As it turns out, the peak acceleration in most seismic records is strongly dependent on high-frequency parts of the motion (typically > 3 Hz), and the assumption that the building's fundamental frequencies are large compared to the ground acceleration does not apply.

Rigid Building on a Flexible Foundation (Rocking)

In this section I investigate what happens if we allow a rigid building to tilt due to flexibility of the foundation. This is not a very realistic problem to consider the ground to be far more flexible than the building, but it does illustrate how forces in the base of a building can be modified by the elasticity of the soil. This is called a **soil-structure interaction** (there are other effects that enter into this problem, but this is probably the most important). The problem is sketched in Figure 6.25. However, for the current discussion, we only ask the simpler question of how the rotation of the building changes the forces on the base of the building as compared with the previous section of this chapter.

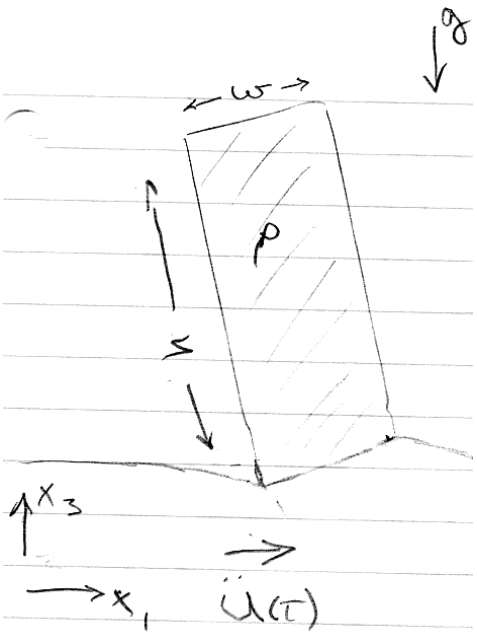


Figure 6.25. Sketch of a rigid building on a flexible foundation

Solving for the forces in this problem involves separating the motion of the building into two parts; 1) the rectilinear motion of the center of mass of the building, and 2) the rotational motion about it axis of angular momentum. This decomposition is shown in Figure 6.26.

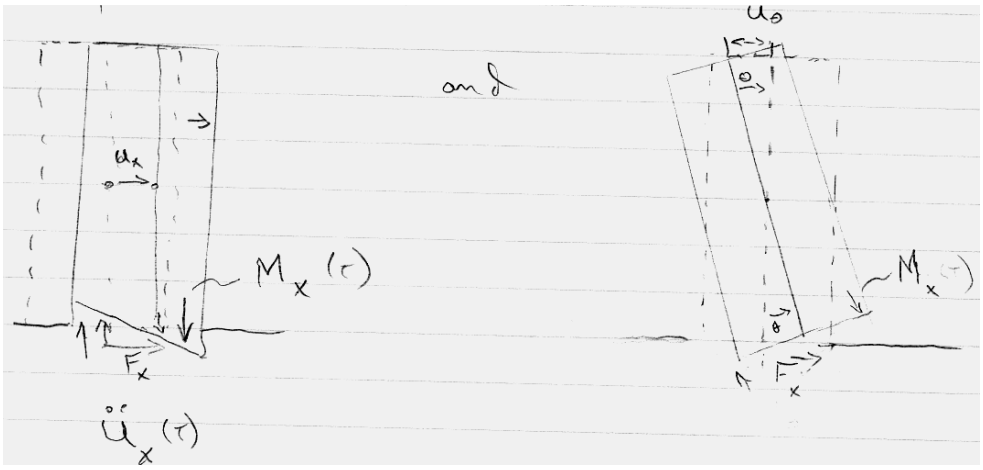


Figure 6.26 The motion of the rigid building can be viewed as the sum of a translation of the center of mass together with a rotation about the principal axis of inertia.

Let $u(t)$ be the motion of the base of the building in an inertial frame, $u_x(t)$ be the linear motion of the center of mass, and $u_\theta(t)$ be the motion of the base due to pure rotation of the building about the center of rotation which is located at the midpoint of the building. Then

$$u(t) = u_x(t) + u_\theta(t) \tag{6.11}$$

and

$$\ddot{u}(t) = \ddot{u}_x(t) + \ddot{u}_\theta(t) \tag{6.12}$$

now

$$u_\theta(t) = \frac{h}{2}\sin\theta(t) \approx \frac{h}{2}\theta(t) \qquad \theta \ll \pi \tag{6.13}$$

then

$$\ddot{u}_\theta(t) \approx \frac{h}{2}\ddot{\theta}(t) \tag{6.14}$$

The force system on the base of the building is the sum of the horizontal force $F_1(t)$ (these are unbalanced) and a resisting moment $M(t)$ caused by the distribution of vertical stresses on the base of the building. The rectilinear momentum of the building $P(t)$ is just

$$P(t) = \rho w^2 \int_0^h \left(\dot{u}(t) - \frac{1}{2}x_3\dot{\theta}(t) \right) dx_3 = m\left(\dot{u}(t) - \frac{1}{2}h\dot{\theta}(t) \right) \tag{6.15}$$

where $m = \rho h w^2$ is just the mass of the building. The force on the base of the building is then

$$F_1(t) = \dot{P}(t) = m\left(\ddot{u}(t) - \frac{1}{2}h\ddot{\theta}(t)\right) \tag{6.16}$$

The shear stress on the base of the building is then

$$\sigma_{13}(t) = \frac{F_1(t)}{w^2} = \frac{\dot{P}(t)}{w^2} = \rho h\left(\ddot{u}(t) - \frac{1}{2}h\ddot{\theta}(t)\right) \tag{6.17}$$

At this point, we do not yet know the shear stress, since we do not know $\theta(t)$. We can calculate the rotation of the building as follows,

$$M(t) + F_1(t)\frac{h}{2} = I\ddot{\theta}(t) \tag{6.18}$$

where I is the moment of inertia. For a rectangular prism,

$$I = \frac{1}{12}m\left(h^2 + w^2\right) \tag{6.19}$$

Combining (6.16), (6.17), and (6.19) yields

$$\ddot{\theta}(t) = \frac{12}{m\left(4h^2 + w^2\right)}\left[M(t) + \frac{mh}{2}\ddot{u}(t)\right] \tag{6.20}$$

Now the moment at the base of the building is

$$M(t) = w\int_{-w/2}^{w/2}\sigma_{33}(t)\,x_1\,dx_1 \tag{6.21}$$

If we assume that the normal stress is proportional to the vertical deflection, then

$$\sigma_{33}(t) = k\theta(t)\,x_1 \qquad -w/2 < x_1 < w/2 \tag{6.22}$$

where k is a type of stiffness with units of stress per unit of displacement (it differs from a regular spring constant, which has units of force per unit displacement). Therefore,

$$M(t) = wk\theta(t)\int_{-w/2}^{w/2}x_1^2\,dx_1 = \frac{1}{12}w^4k\theta(t) \tag{6.23}$$

Combining (6.20) and (6.23) gives us the equation for a single degree of freedom forced oscillator

$$\ddot{\theta}(t) - \frac{kw^4}{m\left(4h^2 + w^2\right)}\theta(t) = \frac{6h}{\left(4h^2 + w^2\right)}\ddot{u}(t) \tag{6.24}$$

We immediately recognize that this is an un-damped forced linear oscillator with a natural period of

$$\omega_0 = w^2\sqrt{\frac{1}{4h^2+w^2}}\sqrt{\frac{k}{m}} \tag{6.25}$$

The full solution to this force oscillator is given in Chapter 1 (see equation 1.40 with damping equal to zero) as

$$\theta(t) = \frac{6h}{\left(4h^2+w^2\right)}\ddot{u}(t) * \left[H(t)\frac{\sin\omega_0 t}{\omega_0}\right] \tag{6.26}$$

While we have modeled the building as an un-damped sdof, the full solution to this problem is quite complex since the oscillations of the building would excite waves in the elastic medium. The excitation of the waves would cause kinetic energy in the building to be radiated as wave energy into the surrounding medium. This would be a form of radiation damping of the oscillations of the building. Substituting (6.26) into (6.17) gives

$$\begin{aligned}\sigma_{13}(t) &= \rho h\left(\ddot{u}(t) - \frac{3h^2}{\left(4h^2+w^2\right)}\ddot{u}(t) * \frac{\partial^2}{\partial t^2}\left[H(t)\frac{\sin\omega_0 t}{\omega_0}\right]\right) \\ &= \rho h\ddot{u}(t) * \left(\delta(t) - \frac{3h^2}{w^2\sqrt{4h^2+w^2}}\sqrt{\frac{m}{k}}\frac{\partial^2}{\partial t^2}\left[H(t)\sin\omega_0 t\right]\right)\end{aligned} \tag{6.27}$$

We can also calculate the vertical compressive stresses at the outer edges of the building by substituting (6.26) into (6.22).

$$\begin{aligned}\sigma_{33}(t; x_1 = \pm w) &= \rho g h \pm kw\theta(t) \\ &= \rho g h \pm kw\frac{6h}{\left(4h^2+w^2\right)}\ddot{u}(t) * \left[H(t)\frac{\sin\omega_0 t}{\omega_0}\right] \\ &= \rho g h \pm \frac{6h}{w\sqrt{4h^2+w^2}}\sqrt{mk}\ddot{u}(t) * \left[H(t)\sin\omega_0 t\right]\end{aligned} \tag{6.28}$$

These are fairly complex relationships. However, we can get some idea of the effect of building tilting by investigating the initial response of the building to a impulse in acceleration, $\ddot{u}_{\max}$. In this case (6.27) becomes

$$\sigma_{13}^{I}(t) = \rho h \ddot{u}_{\max}\left(\delta(t) - \frac{3h^2}{\left(4h^2+w^2\right)}\delta(t) * \frac{\partial^2}{\partial t^2}\left[H(t)\frac{\sin\omega_0 t}{\omega_0}\right]\right)$$

$$= \rho h \ddot{u}_{\max}\left(\delta(t) - \frac{3h^2}{\left(4h^2+w^2\right)}\frac{\partial}{\partial t}\left[\delta(t)\frac{\sin\omega_0 t}{\omega_0} + H(t)\cos\omega_0 t\right]\right)$$

$$= \rho h \ddot{u}_{\max}\left(\delta(t) - \frac{3h^2}{\left(4h^2+w^2\right)}\left[\dot{\delta}(t)\frac{\sin\omega_0 t}{\omega_0} + 2\delta(t)\cos\omega_0 t - H(t)\omega_0\sin\omega_0 t\right]\right)$$

$$= \rho h \ddot{u}_{\max}\delta(t)\left(1 - \frac{6h^2}{\left(4h^2+w^2\right)}\cos\omega_0 t\right) - \rho h \ddot{u}_{\max}H(t)\omega_0\sin\omega_0 t$$

(6.29)

Or we can write the shear response for an arbitrary acceleration as

$$\sigma_{13}^{I}(t) = \rho h \ddot{u}(t)\left(1 - \frac{6h^2}{\left(4h^2+w^2\right)}\cos\omega_0 t\right) - \rho h \ddot{u}(t) * H(t)\omega_0\sin\omega_0 t$$

(6.30)

Therefore, **the shear maximum shear stress is decreased for a rocking building compared with a rigid base** (that is, for an impulse of acceleration). To fully understand the effect of this rocking, we would have to know the actual ground acceleration time history. If the ground motion was harmonic with the same period as the natural frequency of rocking, then the rocking building would resonate with the ground. If the duration of the ground motion was large enough, then the rocking building would develop even larger shears than the rigid building on the rigid foundation.

Notice that the acceleration impulse response for the outer edges of the building can be derived from (6.28), and is

$$\sigma_{33}(t; x_1 = \pm w) = \rho g h \pm \frac{6h}{w\sqrt{4h^2+w^2}}\sqrt{mk}\,\ddot{u}_{\max}\delta(t) * \left[H(t)\sin\omega_0 t\right]$$

$$= \rho g h \pm \frac{6h}{w\sqrt{4h^2+w^2}}\sqrt{mk}\,\ddot{u}(t) * \left[H(t)\sin\omega_0 t\right]$$

(6.31)

Notice that as $k \to 0$, a very, very flexible foundation, then

$$\sigma_{33}(t; x_1 = \pm w) = \rho g h \qquad t \ll \omega_0 \tag{6.32}$$

That is, as the foundation becomes more flexible, the normal stresses on the columns from the overturning torque decrease. However, depending on the convolution term, there may be resonances and the compressive stresses may actually increase.

Building as a single-degree-of-freedom oscillator; Response Spectra

In the previous discussion, I discussed stresses in very stiff buildings that could be approximated as accelerating rigid blocks. Of course, buildings cannot be damaged if they do not deform. In reality, all buildings are flexible and more complex models are required to understand the vulnerabilities of flexible buildings. I now discuss the properties of a building modelled as a single-degree-of-freedom linear oscillator. This is identical to the problem of a simple seismometer that was discussed in Chapter 1 (Figure 1.1). I will shortly describe how this problem can be generalized to multi-degree-of-freedom linear oscillators. For now, though, I will discuss the dynamics of a base-isolated building. In particular, I will assume that the building is approximately rigid and that it sits on a linearly elastic isolator that also has viscous damping (linear damping). Real isolators are nonlinear at large offsets, but I'll ignore that for now. Consider the sdof shown in Fig. 2.27.

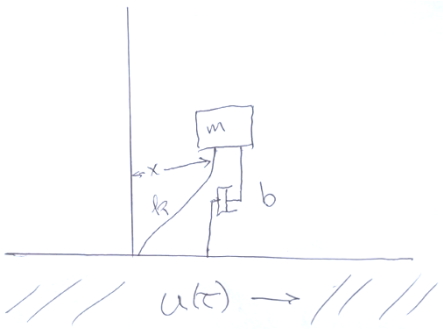


Figure 6.27 Building modeled as a single-degree-of-freedom damped oscillator. This model is useful for base-isolated buildings.

The system in Fig. 6.27 is exactly analogous to that shown in Fig. 1.1, except that the motions are horizontal. In particular, the ground moves horizontally with displacement $u(t)$; the horizontal motion $x(t)$ of a mass m is resisted by a spring of stiffness k; and there is a viscous damper that resists the relative velocity

$\dot{x}$ of the mass with respect to the ground with force $-b\dot{x}$. The force on m is $-kx - b\dot{x}$ and the inertial force on m is $m(\ddot{u}+\ddot{x})$. The equation of motion of the system is then

$$m(\ddot{u}+\ddot{x})+b\dot{x}+kx=0 \tag{6.33}$$

which can be rewritten as

$$\ddot{x}+2\beta\dot{x}+\omega_0^2 x=-\ddot{u} \tag{6.34}$$

where $\omega_0=\sqrt{k/m}$ (the undamped natural frequency) and $\beta \equiv \frac{b}{2m}$ (the damping constant), which is related to the fraction of critical damping, ς, by $\beta = \omega_0 \varsigma$.

It's important to recognize that $u(t)$ is motion with respect to an inertial frame, but $x(t)$ is a **noninertial coordinate.** However, $x(t)+u(t)$ is an inertial coordinate for the mass.

(6.34) can be easily solved using techniques in either the time domain, or alternatively, the frequency domain (see Chapter 1). For example. we can find the response to any ground motion by use of the convolution operator, which is defined as follows.

$$x(t)=\ddot{u}(t)*G(t)\equiv\int_{-\infty}^{\infty}\ddot{u}(\tau)G(t-\tau)\,d\tau \tag{6.35}$$

where

$$G(t)=H(t)\frac{1}{\omega_1}e^{-\beta t}\sin\omega_1 t \tag{6.36}$$

where

$$\omega_1\equiv\sqrt{\omega_0^2-\beta^2} \tag{6.37}$$

G(t) is a Green's function for this system; it is the response of the system to an impulse of ground acceleration. An impulse of acceleration induces a step in inertia, so this solution is identical to the transient response of the system with initial conditions of zero displacement and initial velocity of 1 m/s. The Green's function for 20% damping is shown in Figure 6.28.

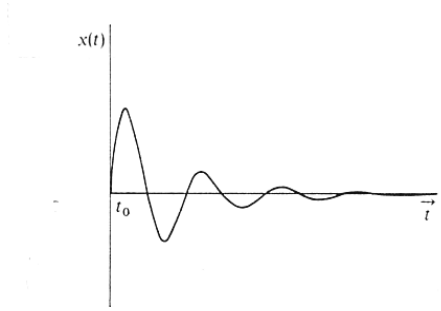


Figure 6.28. Response of a 20% damped sdof to an impulse of acceleration.

If there was no damping in the system, then the mass would move as a sinusoid indefinitely. Interestingly, the amplitude of the Fourier spectrum of x at ω_0 is identical to the amplitude of the steady state harmonic motion that persists after the ground acceleration is over.

When analyzing the dynamic response of elastic buildings, we are more interested in understanding the maximum deflections and forces than we are in the steady vibration. Furthermore, there is always dissipation (damping) in real structures. Suppose that we are interested in estimating the maximum isolator displacement for a base-isolated building. Then this is simply the **response spectral displacement**, which is defined as

$$S_d(\varsigma, \omega_0) \equiv \max_t \left[x(t; \varsigma, \omega_0) \right] \tag{6.38}$$

where ξ and ω_0 are the % critical damping and undamped natural frequency, respectively. Similarly, the **response spectral velocity** is defined as

$$S_v(\varsigma, \omega_0) \equiv \max_t \left[\dot{x}(t; \varsigma, \omega_0) \right] \tag{6.39}$$

and finally the **response spectral acceleration** is defined as

$$S_a(\varsigma, \omega_0) \equiv \max_t \left[\ddot{x}(t; \varsigma, \omega_0) + \ddot{u} \right] \tag{6.40}$$

Notice that spectral acceleration is defined with respect to the inertial coordinate $x+u$, which means that the response spectral acceleration measures the maximum force applied by the spring and damper (normalized by the mass). Figure 6.29 shows an example of response spectral velocity for a range of frequencies and damping. Response Spectra are ubiquitous in earthquake engineering, but please pay attention to the units. Tragically, even today it is common to see British units (inches and feet), which complicates any discussions between engineers and scientists. While any damping can be used, 5% damping is most commonly assumed; this is a pernicious mistake that I will discuss later.

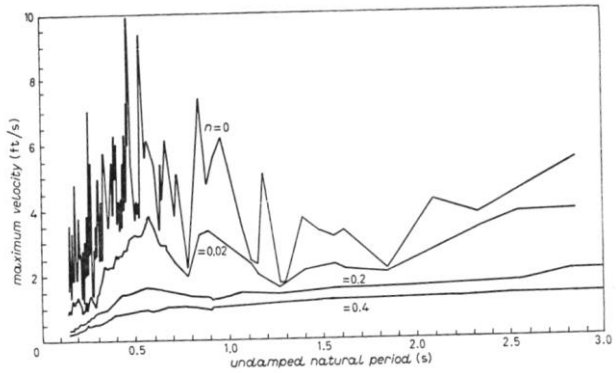


Figure 6.29. A very old figure (circa 1960's) showing response spectral velocities for a variety of damping (0, 2%, 20%, and 40%).

Response spectra are similar to Fourier spectra since they both characterize the amplitude as a function of frequency. Figure 6.30 compares the $S_v(\varsigma = 0)$ (solid line) with the Fourier amplitude spectrum (dotted line) of a ground motion record. Notice that while theses spectra are similar, there are differences. In particular, the response spectrum is always greater than, or equal to the Fourier amplitude spectrum. That is because the response spectrum measures the maximum transient amplitude, while the Fourier spectrum measures the undamped steady state amplitude after the ground has stopped moving. The difference between these two spectra becomes more pronounced as the damping is increased. This difference is obvious in Figure 6.29, where it is seen that the response spectral amplitude decreases with damping.

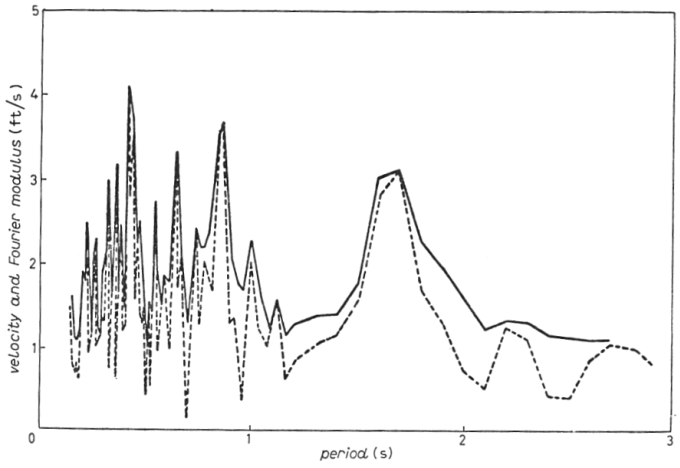


Figure 6.30. Comparison between 0%-damped response spectrum (solid line) and the Fourier amplitude spectrum (dotted line) for a strong motion record.

It is important to recognize that although the Fourier transform is a linear operator, the response spectral operator is not. That is because response spectra requires finding the maximum amplitude; there is no phase (timing) information in response spectra. Recall that linear operators have the property that they can be performed in any order. This is not the case for response spectra.

Response spectra were first introduced in the 1930's by Romeo Martel, who was a Caltech Professor of Mechanical Engineering. The application of response spectra was expanded by George Housner who was a Caltech Professor of Civil Engineering; Housner has been referred to as the "father of response spectral analysis."

The development of response spectral analysis preceded the development of digital computers and the calculation of spectra like that shown in Fig. 6.29 required many hours of human effort. Accordingly, researchers employed approximations that decreased numerical calculations. For example, when calculating $\max_t(x)$, the velocity $\dot{x} \approx 0$ when x is a maximum. Substituting into (6.34) gives

$$x_{\max} \approx \frac{(\ddot{u} + \ddot{x})_{\max}}{\omega_0^2} \tag{6.41}$$

This means that $S_a \approx \omega_0^2 S_d$. Also notice that when $\dot{x}$ is a maximum, then $x \approx 0$, so

$$\dot{x}_{\max} \approx \frac{(\ddot{u}+\ddot{x})_{\max}}{2\beta} = \frac{(\ddot{u}+\ddot{x})_{\max}}{2\varsigma\omega_0} \tag{6.42}$$

which means that $S_a \approx 2\varsigma\omega_0 S_v$. These approximations mean that you only need to calculate one of these response spectra (typically S_v) and then the other spectra can be approximated by simple multiplication or division with ω_0. In the 1960's it was common to determine response spectra using these simple approximations. In this case, the approximated spectra are called **pseudo response spectra**. For example, $PS_d \equiv {}^{2\varsigma}\!/_{\omega_0}\, S_v$ and $PS_a \equiv 2\varsigma\omega_0 S_v$. Figure 6.31 shows how these approximations can be used display all three response spectra on a single plot which is called a tripartite diagram. It starts with a log-log plot of S_v. Diagonal logarithmic axes are added so that PS_a and PS_d can be read off of the S_v curve. This is a very clever way to show lots of information. Unfortunately, it is rarely used these days. It is most common to see S_a since it measures the maximum base shear force (mS_a).

Response spectral analysis is the language of modern earthquake engineering. It is useful to determine the excitation of the modal vibrations of buildings and it is used extensively in structural design. It's important to remember that although this theory works exceedingly well for buildings that have linearly elastic force-deformation, it must be modified to provide realistic answers for buildings that experience damage. Before I deal with the issue of inelastic, nonlinear behavior, I will introduce the theory of normal modes, which is a crowning achievement of applied mathematics. Furthermore, response spectra play a central role in this theory.

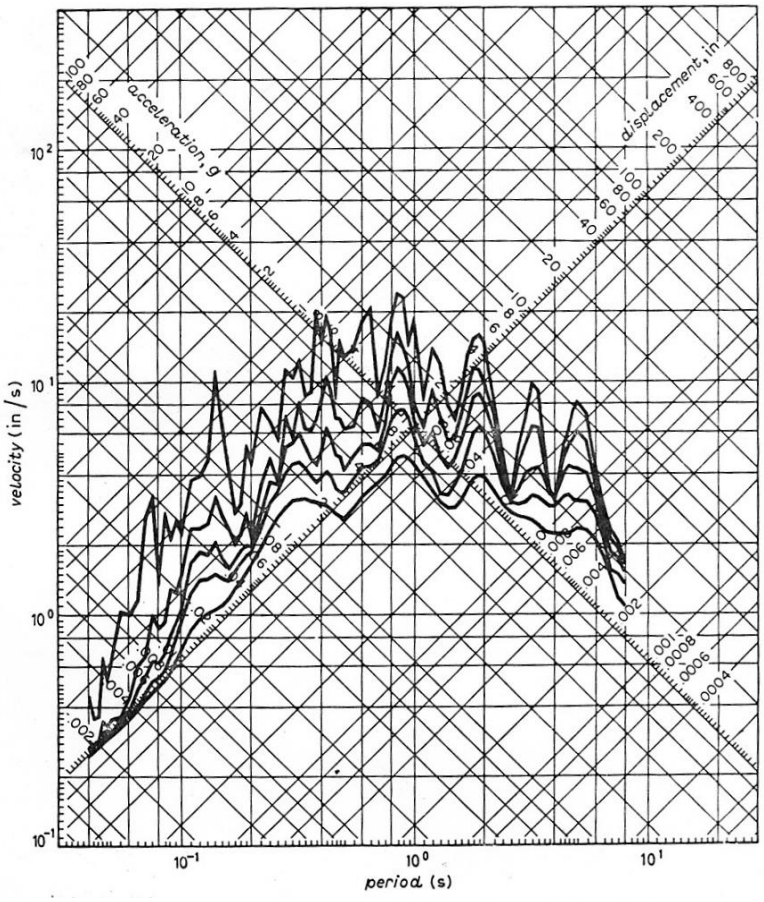
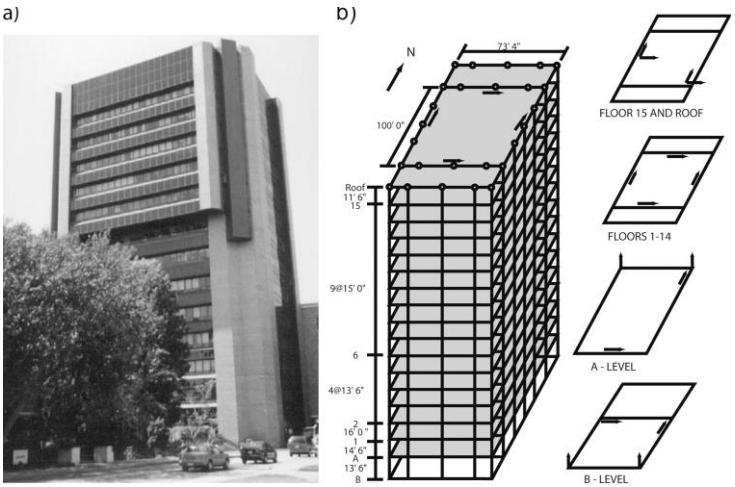


Figure 6.31. Tripartite diagram showing S_v, PS_a, PS_v for a single strong motion record and a variety of damping.

Modal Coordinate Frame

There are several different methodologies to analyze the deformations of buildings that are subjected to earthquake shaking. The situation is analogous to different approaches for solving the forced, single-degree-of-freedom, linear oscillator problem that I discussed in Chapters 1 and 2. In particular, I showed how the system responds to harmonic forces (frequency domain) or alternatively to temporal impulses (time domain). While these two approaches provide identical solutions (provided that all parts of the solutions are included), the two approaches can provide different insights into the behavior of the system. To begin, look at the displacements that were obtained by double integration of accelerations recorded in the 19-story Factor Building (UCLA, SMRF) for a M 3.6 earthquake that happened nearby in the Santa Monica Bay (Figure 6.32). Notice that the displacements in the building are primarily harmonic in nature. This is

typical of all motions in buildings that experience weak shaking that can be modeled as linearly elastic.



The Factor building (a) and its seismic array (b). Arrows show polarities of the single component sensor.

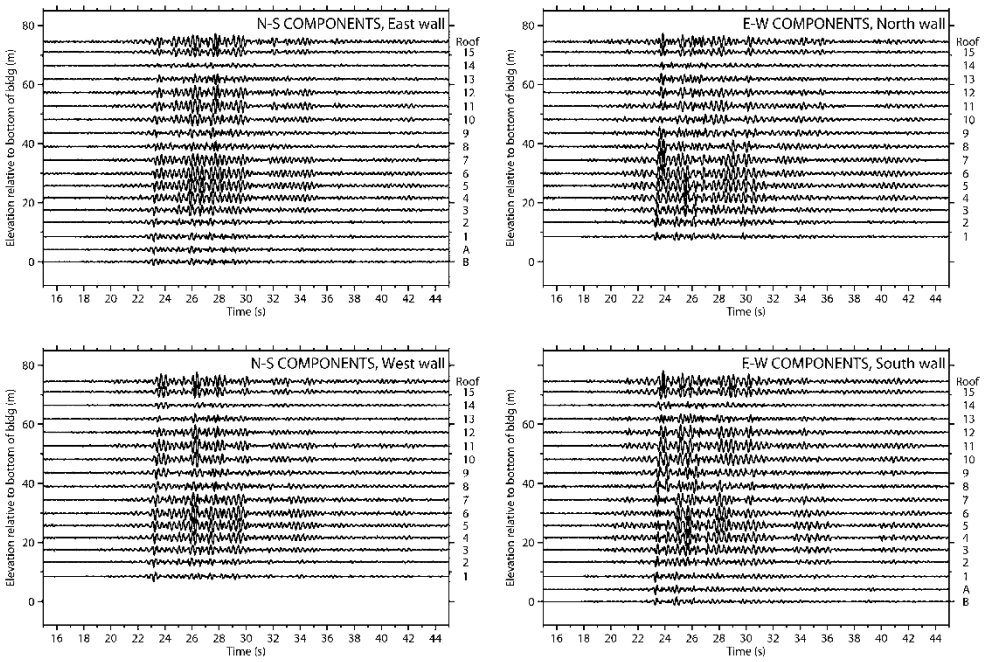


Figure 6.32 (from Kohler and others, 2007) Displacement records for the 16 December 2004 Santa Monica Bay(*M*L 3.6) earthquake. The figure shows the north–south components for sensors on the east and west walls (left), and east–west components for sensors on the north and south walls (right) except for the subbasement sensor which is on the west wall. 1 for location of sensors. Vertical numbering on the right indicates floor number with "A" for basement and "B" for subbasement.

If you look closely, you will see that the motions are harmonic in both time (dominated by 3 Hz) and space (nodes at floors 1, 8, and 14). The Factor Building is unusual in that it has continuously recording 24-bit accelerometers at every floor of the building. The Fourier amplitude spectra of the one-day recordings of ambient vibrations recorded on each floor are shown in Figure 6.33 (from Kohler, M. D., T. Heaton, and C. Bradford, Propagating waves recorded in the steel, moment-frame Factor building during earthquakes, Bull. Seis. Soc. Am., 97, 1334-1345, 2007). Notice that the noise at the roof level is largest in amplitude at 0.6, 1.8, and 3.0 Hz. These are the first three modes of the building. That is, the building resonates at these frequencies. Furthermore, the amplitude response of the spectral peaks is reminiscent of the amplitude response of an under-damped linear oscillator (see Chapter 1). Just as the width of the resonant peak

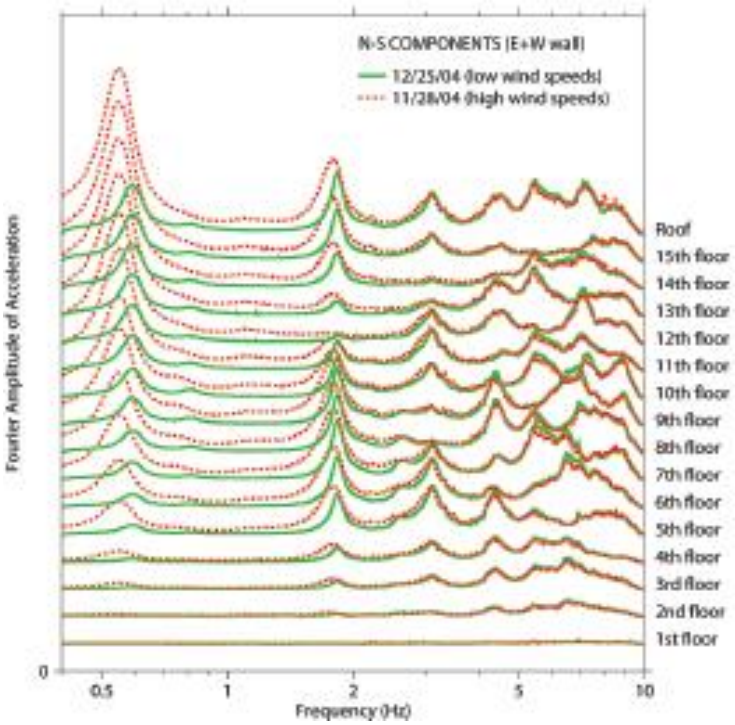


Figure 6.33. Fourier amplitude spectra of 24 hours of ambient vibrations recorded at each floor of the UCLA Factor Building.

of a sdof can be used to measure its Q (or more commonly, its fraction of critical damping), the width of the spectral peaks can be used to measure the damping of each mode of the building.

Figure 6.34 shows the amplitude of the spectral peaks as a function of the floor on which it was recorded. Notice that the first mode's amplitude as a function of height $A_1(z)$ is approximately ¼ of a wavelength of a sinusoid with a total wavelength that is four times the building's height h. That is

$$A_1(z) \approx \sin 2\pi \frac{z}{4h} \tag{6.43}$$

Furthermore, this can be generalized to the higher modes $A_n(z)$,

$$A_n(z) \approx \sin 2\pi \frac{(2n-1)z}{4h} \tag{6.44}$$

where the resonant frequency ω_n of an overtone is related to the of the frequency of the first (aka, fundamental) mode by

$$\omega_n \approx (2n-1)\omega_1 \tag{6.45}$$

These relationships are exact for the modes of a uniform beam that experiences pure shear. While they are useful for quick estimates of actual buildings, real buildings are neither uniform, nor do they experience pure shear; there is always some degree of overall bending.

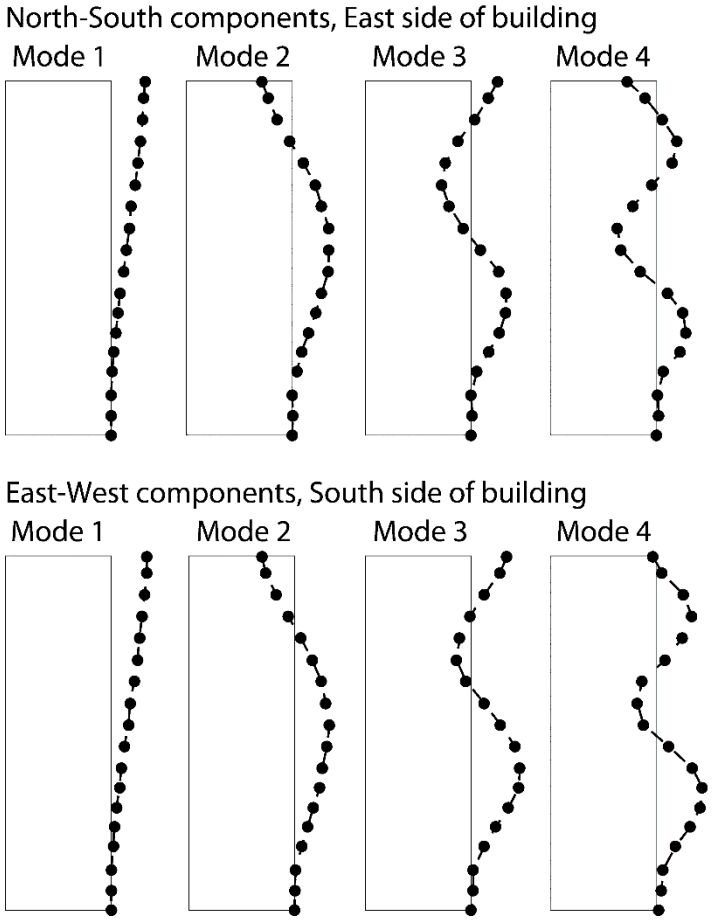


Figure 6.34. Mode shapes determined from narrowband filtered displacements recorded from the 16 December 2004 Santa Monica Bay earthquake. Filled circles represent actual sensor locations along the height of the building that contributed to the mode shape measurements.

We can gain insight into the response of a simple two-story mrf by using the simple model shown in Figure 635. I will assume that the horizontal motion of the base is $u(t)$ and that the first floor and the roof are rigid bodies of mass m. I will assume that the motions of the 2nd floor and the roof with respect to the base of the building is given by $x_1(t)$ and $x_2(t)$, respectively. Now assume that the columns have negligible mass and that when the 2nd floor moves relative to the base, then the columns flex and then create a horizontal restoring force of kx_1. Similarly, assume that when the roof moves relative to the second floor, then there is a restoring force of $k(x_2 - x_1)$. I can now use Newton's 2nd law to write the equations of motion as

$$m\left(\ddot{x}_1+\ddot{u}\right)=-kx_1+k\left(x_2-x_1\right)$$
$$m\left(\ddot{x}_2+\ddot{u}\right)=-k\left(x_2-x_1\right) \tag{6.46}$$

Which can be rewritten

$$m\ddot{x}_1+2kx_1-kx_2=-m\ddot{u}$$
$$m\ddot{x}_2+kx_2-kx_1=-m\ddot{u} \tag{6.47}$$

This gives us two 2nd order coupled linear differential equations together with forcing terms $-m\ddot{u}$ and $-m\ddot{u}$. Just as was the case for a linear sdof, we find that the solution is comprised of the solution to the homogeneous equations plus the particular solution for the forcing terms. The homogeneous equations are

$$m\ddot{x}_1+2kx_1-kx_2=0$$
$$m\ddot{x}_2+kx_2-kx_1=0 \tag{6.48}$$

It is convenient to rewrite (6.48) as a 2-vector equation as follows.

$$\mathbf{M}\ddot{\mathbf{x}}+\mathbf{K}\mathbf{x}=\mathbf{0} \tag{6.49}$$

where $\mathbf{x}\equiv\left(x_1,x_2\right)$, and

$$\mathbf{M}\equiv\begin{bmatrix}m & 0\\ 0 & m\end{bmatrix}\qquad \mathbf{K}\equiv\begin{bmatrix}2k & -k\\ -k & k\end{bmatrix} \tag{6.50}$$

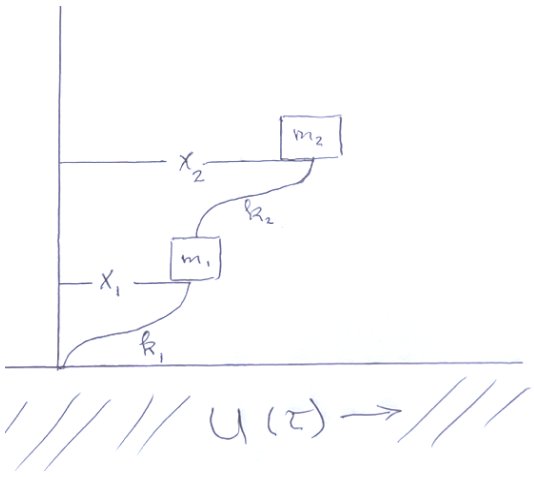


Figure 6.35. Simple model of a two-story building with lateral springs providing the restoring force of the concrete floors.

I can now guess that there are harmonic solutions. That is, assume that there is a solution in which both masses are oscillating with frequency ω, or assume that

$$\mathbf{x} = \mathbf{a}e^{i\omega t} \tag{6.51}$$

where $\mathbf{a} \equiv (a_1, a_2)$ is a 2-vector that describes the amplitude. Differentiating (6.51) twice with respect to time gives

$$\ddot{\mathbf{x}} = -\omega^2 \mathbf{a}e^{i\omega t} \tag{6.52}$$

Inserting (6.51) and (6.52) into (6.49) gives

$$-\omega^2 e^{i\omega t}\mathbf{M}\mathbf{a} + e^{i\omega t}\mathbf{K}\mathbf{a} = \mathbf{0} \tag{6.53}$$

This simplifies to

$$\left(-\omega^2 \mathbf{M} + \mathbf{K}\right)\mathbf{a} = \mathbf{0} \tag{6.54}$$

We now have two coupled linear equations with two unknowns. This is an eigenvalue problem that can be studied in most texts on linear algebra. (6.54) has nontrivial solutions if and only if

$$\det\left(-\omega^2 \mathbf{M} + \mathbf{K}\right) = 0 \tag{6.55}$$

This equation is known as the **characteristic equation** which can be used to find the resonant frequencies. Inserting (6.52) into (6.55) gives

$$\det\begin{bmatrix} 2k - \omega^2 m & -k \\ -k & k - \omega^2 m \end{bmatrix} = \omega^4 m^2 - 3k\omega^2 m + k^2 = 0 \tag{6.56}$$

The quadratic formula can be used to solve (6.56) for ω^2. This gives

$$\omega^2 = \left(\frac{3 \pm \sqrt{5}}{2}\right)\frac{k}{m} \tag{6.57}$$

The fundamental mode is the lowest frequency solution and it is

$$\omega_1 = \sqrt{\frac{3-\sqrt{5}}{2}}\sqrt{\frac{k}{m}} = 0.618\omega_0 \tag{6.58}$$

The 2nd frequency, the overtone frequency is

$$\omega_2 = \sqrt{\frac{3+\sqrt{5}}{2}}\sqrt{\frac{k}{m}} = 1.628\omega_0 \tag{6.59}$$

where $\omega_0 \equiv \sqrt{k/m}$, which is the frequency of a simple sdof using the same spring and mass. (6.58) can be inserted into (6.54) to solve for $\mathbf{a}_1$, which tells us the relative motion of the two masses when they are in free vibration (homogeneous solution) at ω_1. That is,

$$\left(-\omega_1^2\mathbf{M}+\mathbf{K}\right)\mathbf{a}_1=\mathbf{0} \tag{6.60}$$

or

$$\begin{bmatrix} 2k-\omega_1^2 m & -k \\ -k & k-\omega_1^2 m \end{bmatrix}\left[a_{11}, a_{12}\right]=\begin{bmatrix}0\\0\end{bmatrix} \tag{6.61}$$

substituting (6.58) into (6.61) gives

$$\begin{bmatrix} 2k-\dfrac{3-\sqrt{5}}{2}k & -k \\ -k & k-\dfrac{3-\sqrt{5}}{2}k \end{bmatrix}\left[a_{11}, a_{12}\right]=\begin{bmatrix}0\\0\end{bmatrix} \tag{6.62}$$

So

$$\left(\frac{1+\sqrt{5}}{2}\right)a_{11}=a_{12} \qquad a_{11}=0.618a_{12} \tag{6.63}$$

Which means that, when the system is resonating at the first mode, the amplitude of the top floor is 1.6 times larger than the 2nd floor; the amplitude of the base is, by definition zero. Since $\mathbf{a}_1$ is real, there are no phase lags between the floors and the floors move completely in phase.

The second mode can be found in a similar way. Inserting (6.59) into (6.54) gives

$$\begin{bmatrix} 2k-\dfrac{3+\sqrt{5}}{2}k & -k \\ -k & k-\dfrac{3+\sqrt{5}}{2}k \end{bmatrix}\left[a_{21}, a_{22}\right]=\begin{bmatrix}0\\0\end{bmatrix} \tag{6.64}$$

or

$$\frac{1-\sqrt{5}}{2}a_{21}=a_{22} \qquad -0.618a_{21}=a_{22} \tag{6.65}$$

Which means that the 2nd floor moves opposite to the roof. Notice that

$$\mathbf{a}_1 \bullet \mathbf{a}_2 = a_{11}a_{21}+a_{12}a_{22}=0 \tag{6.66}$$

Which tells us that the amplitude vectors (the eigenvectors) are orthogonal. This is a general property that is true as long as the system is purely elastic. I can now write the general solution to (6.49)

$$\mathbf{x}(t) = a_1 \begin{bmatrix} 1 \\ 1.618 \end{bmatrix} \cos(\omega_1 t - \delta_1) + a_2 \begin{bmatrix} -0.618 \\ 1 \end{bmatrix} \cos(\omega_2 t - \delta_2) \tag{6.67}$$

where $a_1, a_2, \delta_1, \delta_2$ are four constants that allow the solution to fit the initial conditions (velocity and displacement for each mass). Notice that I can define a new coordinate system (ξ_1, ξ_2) where

$$\begin{aligned} \xi_1 &\equiv 0.851(0.618x_1 + x_2) \\ \xi_2 &\equiv 0.851(0.618x_1 - x_2) \end{aligned} \tag{6.68}$$

ξ is called the modal coordinate frame and it decouples two coupled ode's for (x_1, x_2) into two uncoupled ode's. These coordinates have been normalized to have unit length and they are are orthogonal, which means that they are basis vectors. Using this new modal coordinate frame results in two equations that are a linear sdof that is extensively discussed in Chapter 1.

Of course, a model of a building that consists of two identical floors is of limited practical value. However, the approach to the problem can be generalized to cover any system of n point masses that are connected by linearly elastic springs and linear viscous dampers. In this case, the problem is written as

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = -\mathbf{M}\ddot{u}\mathbf{i} \tag{6.69}$$

Where $\mathbf{x}$ is a Cartesian n-vector that describes the motion of the discrete masses, $\mathbf{M}$ is the $n \times n$ mass matrix, $\mathbf{C}$ is the $n \times n$ damping matrix, and $\mathbf{K}$ is the $n \times n$ stiffness matrix. In many cases, the mass and stiffness matrices can be constructed using induction from simpler problems. However, sometimes these matrices are not obvious. In these instances, it is best to use a Lagrangian (kinetic energy minus the potential energy) to derive these matrices.

Once the building is described as an n-degree-of-freedom linear system (i.e., (6.69)), then there are several methods to solve the eigen-value problem for the homogeneous solution. Since the damping forces are not known *a priori*, it is best to begin with the approximation that $\mathbf{C} \approx \mathbf{0}$. The eigenvalues $(\omega_1, ..., \omega_n)$ are then computed. Since there is no damping, these eigenvalues are all real valued, which means that all of the masses vibrate in phase (except for possible π phase shifts). Once the n eigenvalues (modal frequencies) are found, then the n corresponding eigenvectors (the mode shapes) can be determined. These eigen vectors are then normalized to have a length of 1. The set of eigenvectors

are all orthogonal to each other and the normalized eigen vectors can be used as a set of basis vectors $(\mathbf{a}_1,...,\mathbf{a}_n)$ to describe any solution to the problem. At that point, the problem is then transformed into

$$\mathbf{x}_i(t) = \mathrm{Re}\left(\mathbf{a}_i e^{i\omega_i t}\right) \tag{6.70}$$

where $\mathbf{a}_i$ is an $n\times 1$ vector that satisfies

$$\left(\mathbf{K}-\omega_i^2\mathbf{M}\right)\mathbf{a}_i = \mathbf{0} \tag{6.71}$$

There is always a rotation matrix ,$\mathbf{Q}$, that will transform $\mathbf{K}-\omega_i^2\mathbf{M}$ into a diagonal matrix

$$\mathbf{Q}\left(\mathbf{K}-\omega_i^2\mathbf{M}\right)\mathbf{Q}^T\mathbf{a}_i = \begin{bmatrix} \xi_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \xi_n \end{bmatrix} \tag{6.72}$$

If this equation is rotated into the modal coordinate system, then the n coupled equations transforms into n decoupled equations.

While the mass and stiffness can be derived from the dimensions of the structural elements, the damping is far more problematic. If a building is driven by an impulse to its free vibrations, then the building would oscillate forever if there was no damping. Obviously, there is damping in real buildings. The simplest way to estimate building damping is using the modal coordinate frame. In particular, the width of the modal resonant peaks gives an estimate of the damping for the mode. This is fairly straightforward for the fundamental modes whose resonant frequencies are well separated from the higher modes. Unfortunately, it can be tricky to identify the characteristics of the resonant peaks of higher modes since these frequencies may be closely spaced.

The use of normal modes to simulate building vibrations assumes that damping can be described as linear viscous. That is, we assume that there is a damping matrix $\mathbf{C}$. The introduction of this damping matrix complicates the eigenvalue problem. While it's still a problem in linear algebra, the eigenvalues (resonant frequencies) are now complex numbers, which means that motions decay exponentially in time. Unfortunately, the eigenvectors also have complex amplitudes, which means that there are phase lags between the masses that oscillate at an eigenfrequency. However, if $\mathbf{C} = c_1\mathbf{M} + c_2\mathbf{K}$, then the eigenvectors are real-valued even though the eigenvalues are complex. The assumption that the damping matrix can be written as a linear sum of the

mass and stiffness matrices is common and it's referred to as **Rayleigh damping**.

Although damping is an important issue, it is rarely measured. In particular, the small-amplitude damping of the fundamental modes of tall buildings is probably about 1% to 2%. However, it is standard that 5% damping is assumed for almost all design calculations. The general idea is that damping increases with the amplitude of the excitation, and that 5% is a reasonable approximation of reality. I will shortly show that 5% is a gross underestimate for a building that undergoes significant damage. Unfortunately, underestimated damping is an endemic error with profound implications for the design of buildings. I will cover this issue when I discuss the behavior of inelastic, nonlinear systems. Before I get to that problem, I want to discuss another method to analyze elastic linear buildings. This technique is based on propagating waves and it's more familiar to seismologists.

Flexible Building as a Continuous Cantilevered Beam

The problem of the dynamic motions of a continuous prismatic beam can give us some insight into the deformation of buildings. To solve such a cantilevered beam problem in full generality is exceedingly complex. To begin with we recognize that, if the building is considered as a continuum, it would often be anisotropic. For example, consider a moment-resisting frame (e.g., Fig. 6.5). Assume that the 1 and 2 axes are horizontal and that the 3 axis is vertical, The building stiffness associated with inter-story drift $(\varepsilon_{13} \text{ and } \varepsilon_{23})$ would be much less than the stiffness associated with in-plane shearing of the actual floor slabs ε_{12}. Likewise the stiffness associated with extension along columns (ε_{33}) is different than for extension along the floor slabs $(\varepsilon_{11} \text{ and } \varepsilon_{22})$. However, we will assume that only the parts of the strain tensor that are important to describe the deformation of the building are the inter-story shear strain and the extensional strain along the columns. We can thus approximate the building as being isotropic since the other elastic moduli are not important. We must also recognize that the material in our imaginary continuous beam may have a very unusual Poisson's ratio. That is, the building is very stiff in compression along the columns and very flexible in inter-story shear. That is, $E \gg \mu$. In fact, for a tall mrf (either concrete or steel), the velocity of longitudinal-waves up the building $\left(\sqrt{E/\rho_{eff}}\right)$ is 10 to 20 times more than the inter-story S-wave velocity $\left(\sqrt{\mu_{eff}/\rho_{eff}}\right)$.

While the complete solution for an isotropic elastic cantilevered beam is itself quite complex, there are two end-member cases where the solution is more tractable (see figure 6.36).

The first is the case in which the height of the building is large compared to the length of the base (a tall skinny building). In this case the building deforms primarily by bending, which is the term used for extensional strains in the columns (extension and compression in the columns at the opposite sides of the building). There is a well-developed theory that allows beam problems to be solved with the assumption that the shear strains are approximately zero (the technical theory of bending). This is called a **bending beam**. The stiffness of a bending beam is determined by its flexural rigidity, EI, which scales with dimension as

$$EI = E\int_0^w\int_0^w x_1^2 dx_1 dx_2 = \frac{E}{3}w^4 \tag{6.73}$$

where E is the Young's modulus, and the building is assumed to have a square cross section of width w. The building becomes very stiff against flexure as w becomes large. It's easier to bend a thin rod than it is a pipe having the same total mass per unit length.

In the second case the building is assumed to be much wider than it is tall. When the ground beneath the building moves horizontally, this is identical to the problem of having an SH wave propagate vertically in a layer of building; the bending is approximately zero in this case. This is called a **shear beam.** In this case the total stiffness against shear is just the shear modulus times the cross sectional area $\left(\mu w^2\right)$. Therefore, it is easy to see why a wide building is dominated by shear and not by bending.

While actual buildings are neither a true bending beam nor a shear beam, we can gain some useful insight by looking at these approximate modes of deformation.

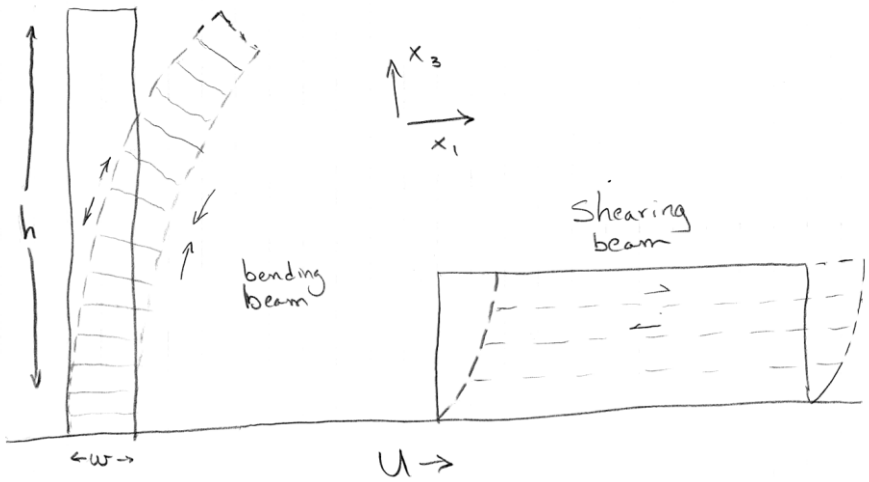


Figure 6.36. Sketch on the left is a bending beam (negligible inter-story shear strains). Most of the deformation occurs because of compression and extension of the outer columns. Sketch on the right is a shearing beam (negligible extension/compression on the columns). The floor slabs remain horizontal.

Flexible Building as a Shearing Beam

Figure 6.37 shows the horizontal accelerations that occurred on different floors of a 52-story steel mrf building in downtown Los Angeles during the 1994 Northridge earthquake. Notice the prominent pulse of acceleration that occurs at the base of the building at about 14 seconds into the record. This pulse can be observed to propagate up the building and it arrives at the top about 1.5 seconds later. Also notice that the pulse is twice as large on the roof as it is in the rest of the building. You can even see a hint that the pulse travels back down the building after it reflects off the top. This type of behavior is exactly what we expect from a shear beam. It is identical to the problem of a vertically propagating SH wave in a plate with a rigid boundary at the bottom and a free boundary at the top. I already extensively discussed this problem in Chapter 4.

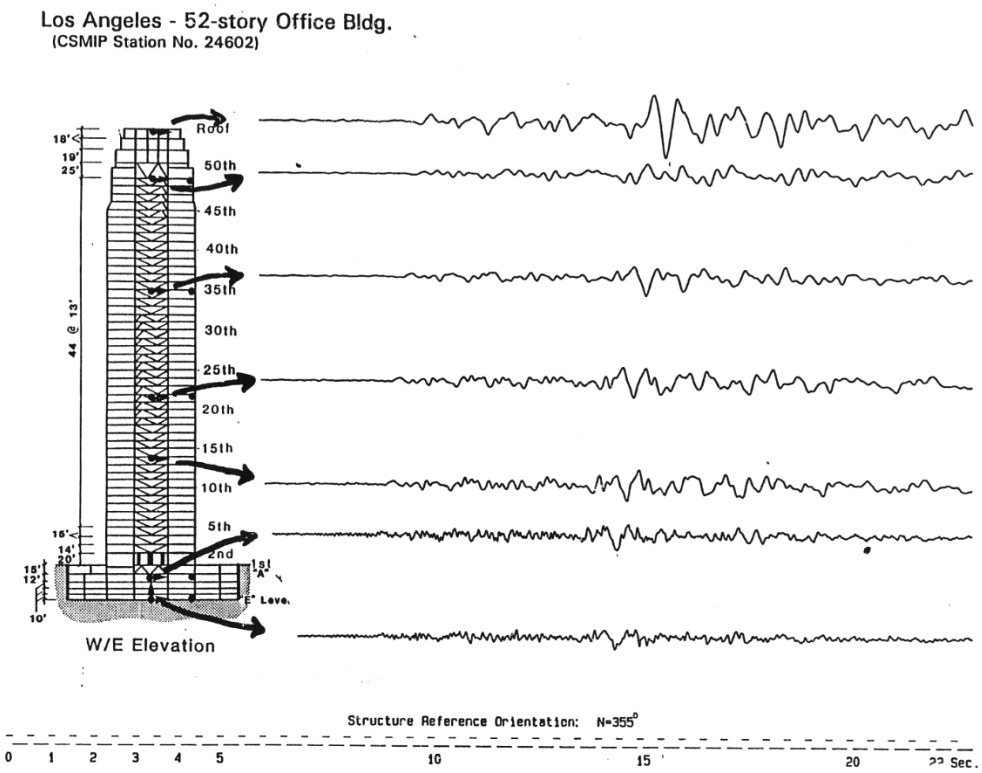


Figure 6.37. Horizontal accelerations in a steel mrf during the 1994 Northridge earthquake. Notice the vertically propagating pulse.

Figure 6.38 shows four other examples of SH-waves propagating vertically in a steel-frame building. The motions recorded at the ground level have been deconvolved from all of the records to show the response of the building to an impulse of horizontal displacement at the base.

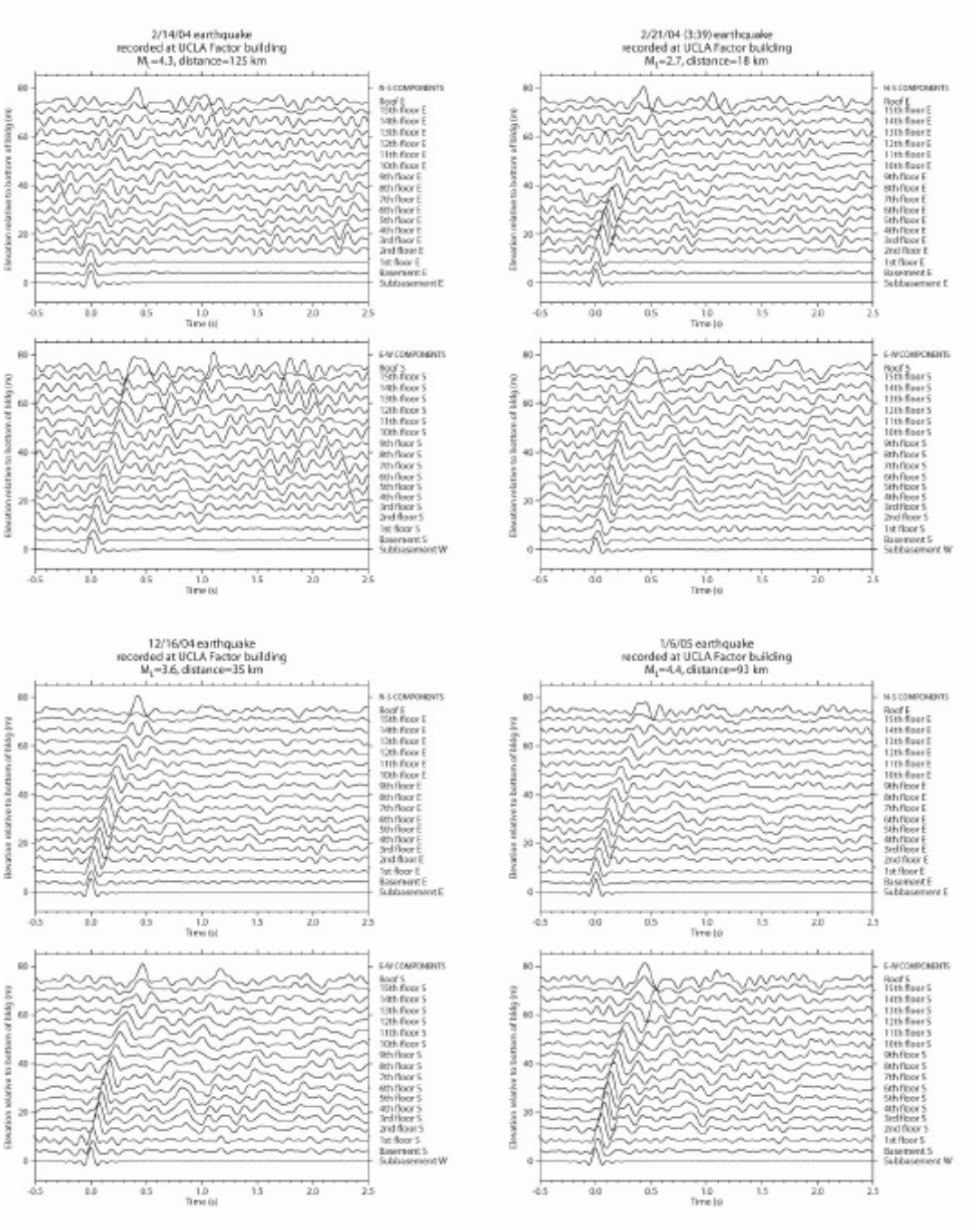


Figure 6.38. Individual impulse response functions for four earthquakes recorded in the UCLA Factor Building. (a) 14 February 2004, (b) 21 February 2004 (3:39 UTC time), (c) 16 December 2004, and (d) 6 January 2005. See Table 1 for more information about these earthquakes. Each earthquake's pair of plots shows the north–south components for sensors on the east wall (top) and east–west components for sensors on the south wall, except for the subbasement sensor which is on the west wall. For clarity, we show only those walls that had a subbasement sensor.

In chapter 4, we saw that the solution to this problem can be written as a sum of reflecting SH-wave pulses. The motion in the building is given by

$$u_1(t,x_3)=u_1^g\left(t-\frac{x_3}{c}\right)+u_1^g\left(t-\frac{2h}{c}+\frac{x_3}{c}\right)-u_1^g\left(t-\frac{2h}{c}-\frac{x_3}{c}\right)$$
$$-u_1^g\left(t-\frac{4h}{c}+\frac{x_3}{c}\right)+u_1^g\left(t-\frac{4h}{c}-\frac{x_3}{c}\right)+.... \tag{6.74}$$

where c is the shear-wave velocity in the building and $u_1^g(t)$ is the horizontal motion of the ground at the base of the building. Notice that this sequence repeats with periodicity $\frac{4h}{c}$, which is the fundamental period of the building oscillation. We are particularly interested in the drift in the building ε_{13}, which we can calculate from (6.74).

$$\varepsilon_{13}(t,x_3)=\frac{1}{c}\left[-\dot{u}_1^g\left(t-\frac{x_3}{c}\right)+\dot{u}_1^g\left(t-\frac{2h}{c}+\frac{x_3}{c}\right)+\dot{u}_1^g\left(t-\frac{2h}{c}-\frac{x_3}{c}\right)\right.$$
$$\left.-\dot{u}_1^g\left(t-\frac{4h}{c}+\frac{x_3}{c}\right)-\dot{u}_1^g\left(t-\frac{4h}{c}-\frac{x_3}{c}\right)+....\right] \tag{6.75}$$

We are especially interested in the drift at the base of the building $(x_3=0)$, or

$$\varepsilon_{13}^b(t)=\frac{1}{c}\left[-\dot{u}_1^g(t)+2\dot{u}_1^g\left(t-\frac{2h}{c}\right)-2\dot{u}_1^g\left(t-\frac{4h}{c}\right)+....\right] \tag{6.76}$$

That is at the base of the building, the up- and down-going waves interfere destructively to give zero displacement (remember it's a rigid base), but the associated strains interfere constructively to give twice as large a drift. Notice that the drift at the top of the building is zero, even though the motion is twice as large as in the rest of the building.

The actual drift in the base of our shear beam depends on the nature of the ground velocity $\dot{u}_1(t)$. Although strong shaking in earthquakes can take a wide variety of forms, it is common that ground displacements near large ruptures have motions described by a pulse of displacement (sometimes referred to as the "killer pulse"), or in other cases they may be dominated by the permanent offset of the ground with respect to an inertial reference frame. Consider the simple ground motions shown in Figure 6.22. The ground accelerations consist of a sequence of positive and negative constant steps. This results in ground velocity that consists of a number of linear ramps. The solution to this problem is simple as long as the duration of the pulse is shorter than the time required for the wave to travel up the building and then return $\left(\frac{2h}{c}\right)$. Things get more complex when the duration

of the ground motion becomes large. Of interest is the maximum shear strain at the base of the building, which we can write as

$$\left|\varepsilon_{13}^{b}\right|_{\max} = \frac{\left|\dot{u}_g\right|_{\max}}{c} A \tag{6.77}$$

where A is an amplification factor. For ground motions A and B, A depends on T_p/T_1, where $T_1 = 4h/c$ is the fundamental period of the building. The factor A reaches 2 for ground motion A (when $T_p \le T_1$) and 4 for ground motion B (when $T_p = T_1$). Plots of $\varepsilon_{13}^{b}(t)$ for $T_p = T_1$ are shown in Figure 6.39. The configurations of the building at different times are shown in Figure 6.40.

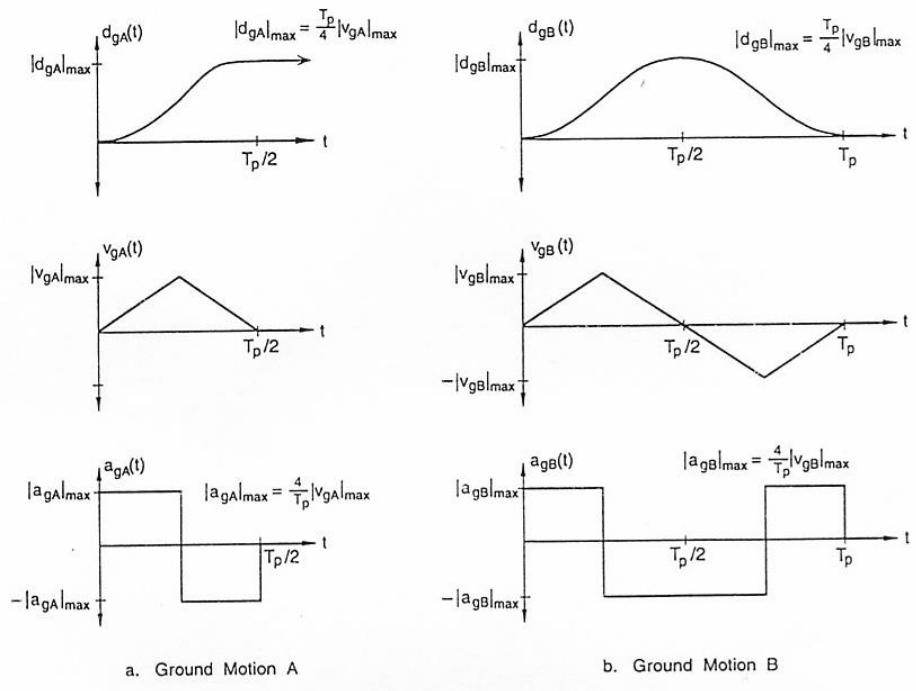


Figure 6.39. Simple ground motions that consist of a simple static displacement (case A) and a pulse of displacement (case B).

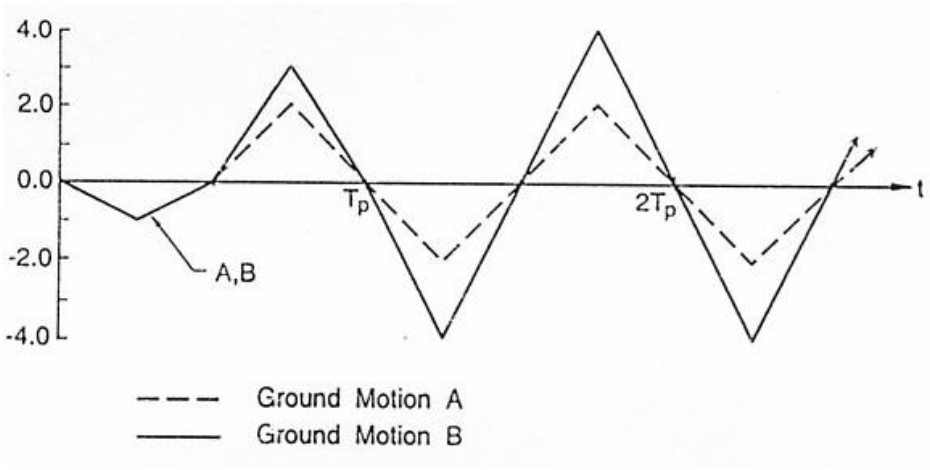


Figure 6.40. Shear strain in the base of the building. One unit on the vertical axis corresponds to a strain of $\frac{c\varepsilon_{13}^{b}(t)}{|\dot{u}_g|_{\max}}$.

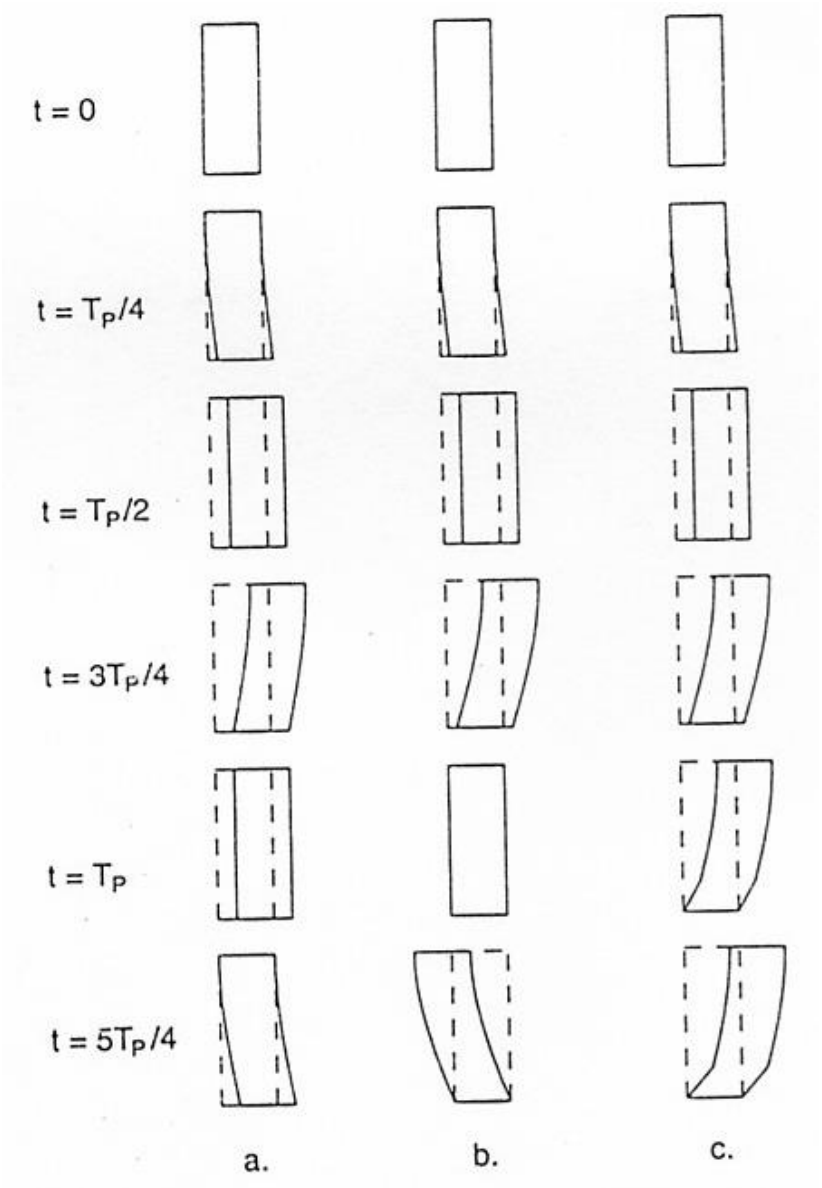


Figure 6.41. Configuration of a multi-story building at time intervals of $T_p/4$ for the case $T_p = T_1$. a) Elastic shear beam building for ground motion A. b) Elastic shear beam building for ground motion B. c) Inelastic shear beam building for ground motion B (qualitative depiction).

The shear strain $\left|\varepsilon_{13}^{b}\right|_{\max}$ can be large enough to be well into the inelastic range. With $\left|\dot{u}_g\right|_{\max} = 1\,\mathrm{m/s}$, and $c = 100\,\mathrm{m/s}$ (a typical value for a tall building), we can use (6.77) and Figure 6.39 to calculate that $\left|\varepsilon_{13}^{b}\right|_{\max} = 0.02$ for ground motion A

and 0.04 for ground motion B, when $T_p = T_1$. These are large values and can greatly exceed the yield strain at the base of the building (e.g., check out Figure 6.10).

The dynamics problem becomes far more complex when the building experiences yielding. That is the governing equations are no longer linear and it is generally necessary to perform a careful finite-element analysis to understand the deformation of the structure. Such an analysis was performed by Hall and others (1995), and the typical results are shown in Figure 6.41. Figure 6.42 shows the location of weld failures (moment resisting connections) in the structure.

When the building yields, it tends to develop a permanent bend in the structure. Once a tall frame building is permanently bent, there is really no practical way to straighten it again, and it is a total loss. Furthermore, if the bending exceeds several percent locally, then there may be a real fear of collapse due to **P-Δ** effects.

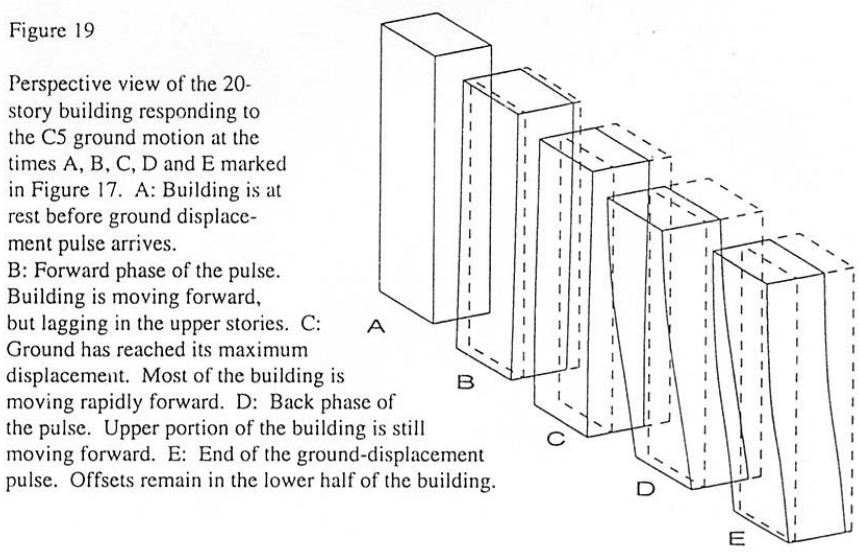


Figure 6.42 From Hall and others (1995).

One way to gain some insight into the behavior of a yielding beam is to consider the building as if it were linear, but with a local stiffness that changes with the amplitude of the local drift. For example, the slope of the force-drift curve in Figure 6.6 is called the tangent stiffness, and it rapidly decreases when the building begins to yield (it even changes sign). Since the stiffness is critical in determining the velocity at which a deformation propagates up a building, loss of stiffness due to yielding means that deformations tend to slow their propagation up a building. That is, once yielding begins, deformations tend to localize in these yielding zones. Perhaps an example of this is shown in Figure 6.44 in which an 8-story

building lost the 6th story during the 1995 Kobe earthquake. That is, when ground motions are propagating in a building, then once yielding begins at a particular place, then that is where the majority of the strain will occur.

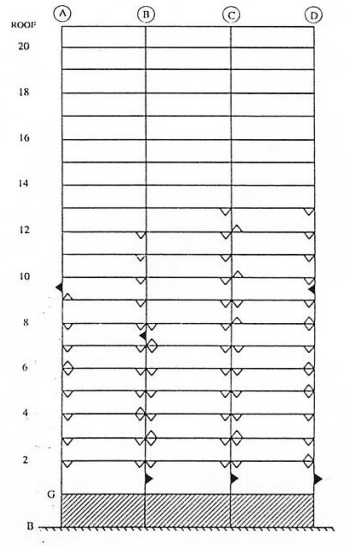


Figure 6.43. Distribution of weld fractures in a 20 story building caused by a 2 meter displacement pulse as reported by Hall and others (1995). A dark triangle locates a cracked column-flange weld at a column splice due to tension in the column. These can be very serious, since if the column separates in tension, but does not come back together properly, the column may fail to carry the weight of the building (very bad). Open triangles locate failed moment-frame connections, which causes a loss of ductility, but is not as serious as failure of a column splice.

Figure 6.44. Once yielding begins at some location, then the loss of stiffness at that location can tend to localize the deformations to that location. This may

have been the cause of the collapse of the 6th floor of this 8-story building in Kobe in 1995.

Prediction of the collapse of structures is extremely difficult. That is, the failure of a critical component (such as a weld) may cause loads carried by a structural element to be transferred to others structural elements, which may cause a cascade of failures. Assessing the likelihood of failure means obtaining an accurate understanding of all of these interrelations (highly nonlinear and perhaps chaotic). Consider the three identical 21-story steel mrf towers in Mexico City that experienced the 1995 Michoacan earthquake (Figure 6.45 and 6.46). This was not a failure due to the propagation of a displacement pulse. Instead, it was due to amplification of 2-second ground motions by the shallow sediments beneath Mexico City (see Chapter 4).

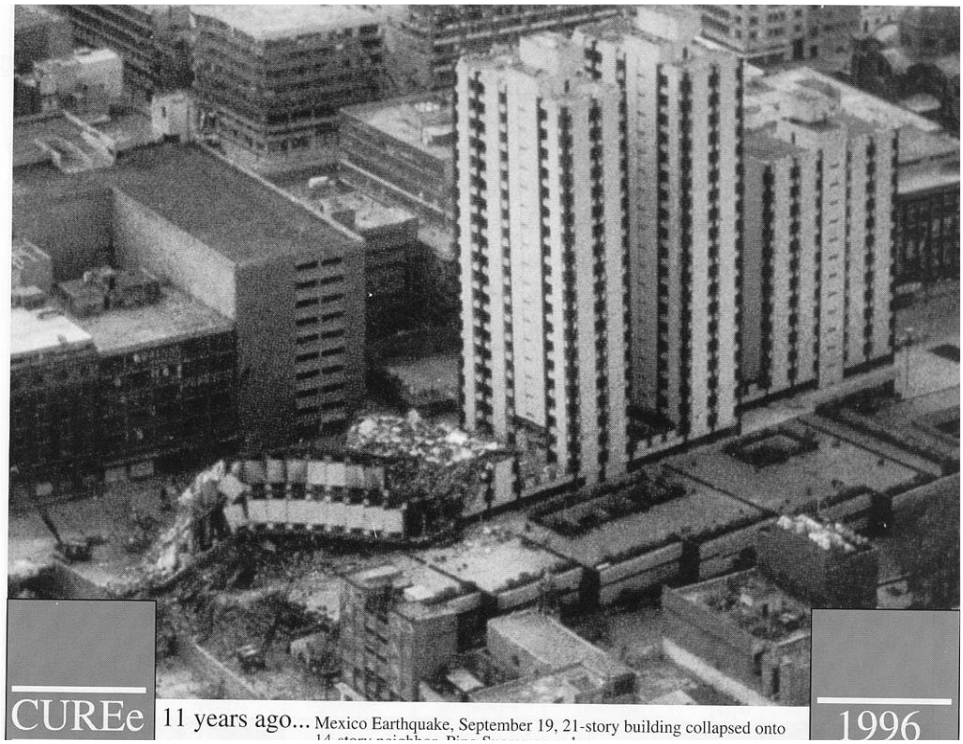


Figure 6.45. Three 21-story steel mrf's that were designed to be identical. The left tower collapsed, the middle tower had a 1-meter permanent drift of its roof, and the right tower suffered no apparent damage. (Mexico City following the 1985 Michoacan earthquake).

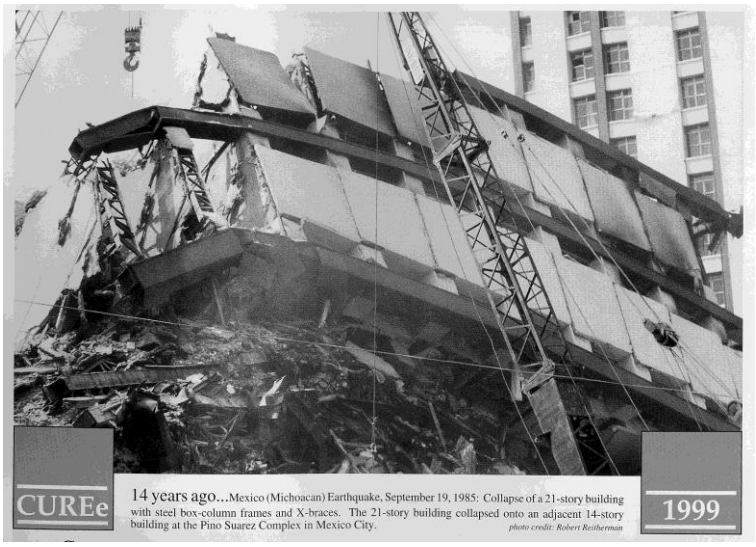


Figure 6.46. Close-up of the left tower in Figure 6.45.

Although the three towers were designed to be the same, their behavior was quite different, despite their close proximity to each other. The tower on the right suffered no apparent damage as a result of the shaking. The tower on the left collapsed, while the tower in the middle had a permanent roof drift of about 1 meter. It would be very difficult to explain this difference by current state of the art in numerical modeling of these buildings.

Base Shear from fundamental mode of a shear beam

The displacement as a function of position and time for a cantilevered shear beam that is vibrating in its fundamental mode is

$$u(z,t) = A\cos(\omega_0 t)\sin(k_0 z) \tag{6.78}$$

Where A is the amplitude of the roof motion, $\omega_0 = 2\pi/T_0$, $k_0 = 2\pi/\lambda_0 = \pi/2h$, where T_0 is the vibration period, k_0 is the wavenumber, and λ_0 is the wavelength of the fundamental mode, which is four times the height of the building, h. If the average density (total building weight divided by the total volume) is ρ, then the total momentum $P(t)$ is

$$
\begin{aligned}
P(t) &= \rho \int_{z=0}^{z=h} \dot{u}(t,z)dz \\
&= -\rho\omega_0 A\sin(\omega_0 t) \int_{z=0}^{z=h} \sin(k_0 z)dz \\
&= -\rho\omega_0 A\sin(\omega_0 t)\left|\frac{-\cos(k_0 z)}{k_0}\right|_{z=0}^{z=h} \\
&= \rho\frac{\omega_0}{k_0} A\sin(\omega_0 t) \\
&= \rho\frac{4h}{T_0} A\sin(\omega_0 t)
\end{aligned}
\tag{6.79}
$$

The force on the base is.

$$
\begin{aligned}
F_{base}(t) &= \dot{P}(t) \\
&= \rho\frac{4h}{T_0}\omega_0 A\cos(\omega_0 t) \\
&= \rho\frac{8\pi h}{T_0^{\,2}} A\cos(\omega_0 t)
\end{aligned}
\tag{6.80}
$$

The weight of the building is just $W = \rho hg$, so the base shear can be written,

$$
\frac{F_{\max}}{W} = \frac{8\pi}{gT_0^{\,2}} A \tag{6.81}
$$

We can now estimate the base shear yield force if we know the roof amplitude that causes yielding. I will assume that yielding occurs when the shear strain at the base exceeds $\varepsilon_{yield} \approx 8\times10^{-3}$.

The shear strain is just

$$
\varepsilon = \frac{\partial u}{\partial z} = A\cos(\omega_0 t)k_0\cos(k_0 z) \tag{6.82}
$$

The maximum at the base is then

$$
\varepsilon_{base} = Ak_0 = \frac{A}{h}\frac{\pi}{2} \tag{6.83}
$$

So

$$
A = \frac{2}{\pi} h\varepsilon_{base} \tag{6.84}
$$

If $\varepsilon_{base} \approx 8\times10^{-3}$, then $A \approx 0.0051h$. Therefore

$$\frac{F_{\max}}{W} = \frac{8\pi}{gT_0^{\,2}} 0.0051h = 0.0128\frac{h}{T_0^2} \approx \frac{\#ofstories}{20}\frac{1}{T_0^2} \tag{6.85}$$

Soil-Structure Interaction for a Shear Beam

If we are treating the shear-beam building as if it was a low-velocity, low-density layer on the top of the Earth, we can gain some insight into how the building interacts with the soil layers. For instance, we can ask 1) how much does the presence of the building change the ground motion at its base compared to the ground motion that would have occurred without the presence of the building (a free field site), and 2) how much of the motion of the building is transmitted through the base of the building when a wave reflects off the top of the building and then transmits through the base of the building?

From Chapter 4 we know that for an SH-wave vertically incident on the base of a layer of buildings,

$$\frac{A_B^T}{A_G^I} = \frac{2\mu_G\beta_B}{\mu_G\beta_B + \mu_B\beta_G} = \frac{2\beta_B\rho_B}{\beta_G\rho_G + \beta_B\rho_B} \approx 2\left[1 - \frac{\beta_B\rho_B}{\beta_G\rho_G} + \left(\frac{\beta_B\rho_B}{\beta_G\rho_G}\right)^2 + \ldots\right] \tag{6.86}$$

where A_B^T is the amplitude of the wave transmitted into the building and A_G^I is the amplitude of the incident wave from the ground. Let us suppose that the shear wave velocity in the soil is approximately twice that in the building (a fairly soft soil), and that the density of the building is 5% of the density of the soil, then

$$\frac{A_B^T}{A_G^I} \approx 2\left(1 - 0.025\right) = 1.975 \tag{6.87}$$

But the amplitude of the motion without the building is 2.0 because it is an SH reflection off of the free surface. Therefore, the building causes the amplitude of the motion to be decreased by 1 ¼ % relative to the ground motion that would have occurred without the presence of the building. This soil-structure interaction effect seems to be far less important than the effect of allowing the building to rock on its foundation (discussed earlier in this chapter).

We can also compute the size of the wave reflected off the base of the building A_B^R compared with the amplitude of the down-going wave in the building A_B^I. That is

$$\begin{aligned}\frac{A_B^R}{A_B^I} &= \frac{\beta_B\mu_G - \beta_G\mu_B}{\beta_B\mu_G + \beta_G\mu_B} \\ &= \frac{\mu_G^2\mu_B\rho_G - \mu_B^2\mu_G\rho_B}{\mu_G^2\mu_B\rho_G + \mu_B^2\mu_G\rho_B + 2\mu_G\mu_B\sqrt{\mu_G\mu_B\rho_G\rho_B}}\end{aligned} \tag{6.88}$$

If the density of the ground is large compared to the building $\left(\rho_G \gg \rho_B\right)$, then

$$\begin{aligned}\frac{A_B^R}{A_B^I} &\approx \frac{\mu_G^2\mu_B\rho_G - \mu_B^2\mu_G\rho_B}{\mu_G^2\mu_B\rho_G} \\ &= 1 - \frac{\rho_B\mu_B}{\rho_G\mu_G} \\ &= 1 - \left(\frac{\rho_B\beta_B}{\rho_G\beta_G}\right)^2\end{aligned} \tag{6.89}$$

Therefore, most of the wave is reflected off the base of the building. For the previous case, the reflected wave is 99.94% the amplitude of the incident wave.

Bending Beam

It is not possible to solve the bending beam problem in the same way that we did the shearing beam. For the technical theory of bending, the horizontal displacements of the building obey the Bernoulli-Euler equation, which is

$$EI\frac{\partial^4 u_1\left(x_3,t\right)}{\partial x_3^4} = -\rho S\frac{\partial^2 u_1\left(x_3,t\right)}{\partial t^2} \tag{6.90}$$

where S is the cross-sectional area of the building ($=w^2$ if the building has a square cross section). Whereas Navier's equation was a second order equation, the Bernoulli-Euler bending beam equation is a fourth order equation. Fortunately, this is still a linear equation. However, it is no longer true that there is a unique velocity such that $u_1 = f\left(t - x_3/c\right)$ solves this equation. We can try a traveling harmonic wave; that is assume that

$$u_1 = \sin\left(k_n x - \omega_n t\right) \tag{6.91}$$

Direct substitution indicates that (6.83) is a solution to (6.82) if

$$\frac{\omega_n}{k} = \sqrt{\omega_n}\sqrt[4]{\frac{EI}{\rho S}} \tag{6.92}$$

But ω_n/k_n is just the phase velocity c_n of this traveling harmonic wave. Therefore, we see that the phase velocity of a traveling harmonic waves increases as the square root of the frequency of the wave. Since (6.82) is linear, we can form a more general solution of the form

$$u_1 = \sum_n C_n \sin k_n \left(x - c_n t\right) \tag{6.93}$$

where the c_n's are constants. Since the different frequencies travel at different velocities, the waveform will change as the wave propagates. This is known as dispersion. There is probably some dispersion that occurred in the propagation of the pulse in the building shown in Figure 6.37. This may explain why it becomes difficult to distinguish the pulse after it has propagated a long distance in the building.

Misunderstanding of modal damping

As can be seen in Fig. 6.29, the amplitude of a response spectra varies strongly with the assumed damping. While undamped response spectra are similar to Fourier amplitude spectra, they grow ever more different with increasing damping. Remember that there is no information about the arrival of phases contained in Fourier amplitude spectra. Even if the time-domain record has the arrival of a pulse that has a broad frequency content, the spectrum of the entire record usually continues to vibrate for a long time (we call it coda). This coda is often comprised of resonances in the low-velocity materials near the Earth's surface. As the damping increases, the response spectrum becomes more sensitive to the largest transients (e.g. pulses). Attempts to measure the small amplitude damping of tall buildings indicates that 2 % is typical (reference). However, it is widely believed that inelastic dissipation (i.e. damping) increases when a building vibrates in the nonlinear range (inter-story strain above 10^{-3}). I don't know why, but it is commonly assumed that 5 % damping is a good assumption when designing an earthquake resistant building. In fact, the current standard relationships to predict response spectral amplitude as a function of distance and earthquake magnitude assume 5 % damping.

Damping of wooden houses

In order to see how damping affects the analysis of a building, I will consider the problem of a wooden house. I discussed wooden houses earlier in this chapter and I argued that they are very resilient in strong shaking. I argued that they are lightweight and that they have high stiffness to minimize deflections from the gravity load of contents (often called the live load). Forced vibration tests of wooden houses typically show that their fundamental mode resonance is about 7 Hz.

The California Earthquake Authority (CEA) sponsored a quantitative study of the vulnerabilities of wooden houses in earthquakes. They tested key structural elements to understand their yielding properties. There are numerous observations of wooden houses in strong shaking and collapse of foundation cripple walls is probably the most important class of failures. Cripple walls are short walls between the concrete foundation and the first-story floor (Figure 6.47 shows an example). Prior to building code updates in the mid-20th century, cripple walls were designed to provide a crawl space beneath a house, and they were designed to support the vertical weight of a house; their lateral strength was often low. New houses are required to have cripple walls that are braced against shear forces. Typically, cripple walls include plywood panels that resist in-plane shearing.

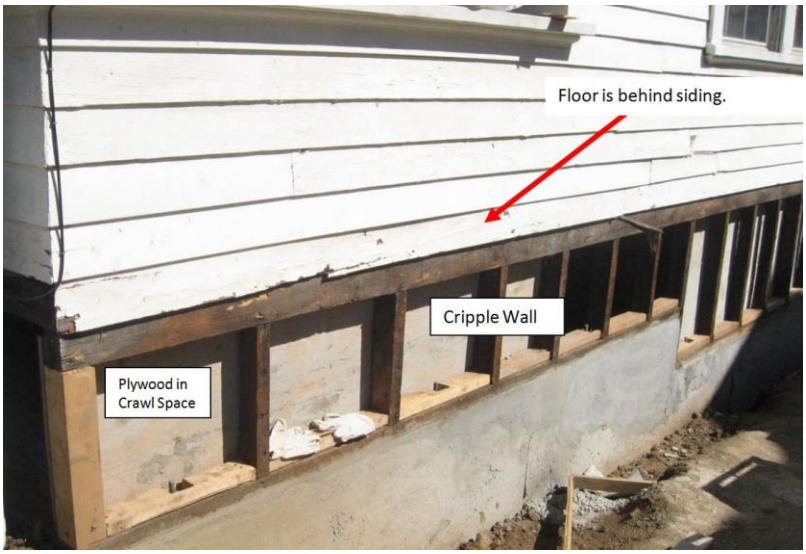


Figure 6.47. Example of a cripple wall on an old wooden house. The left end of the cripple wall has been retrofitted with plywood that resists shearing that is parallel to the wall. The solid wooden floor is strong and rigid and it ensures that there is no weak direction as long (as all the cripple walls are retrofitted).

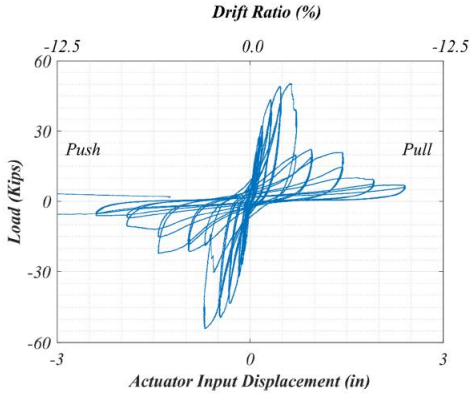


Figure 6.48. Force vs shear displacement for a cripple tested by the Pacific Earth Engineering Research Center (PEER) for the California Earthquake Authority (CEA). A hydraulic ram forces the cripple wall through a series of back and forth displacements whose amplitude grows with each cycle. As the wall is sheared, the force is also recorded. This shows the force displacement relationship for an unretrofitted cripple wall.

Figure 6.48 is the force/deflection curve for an unretrofitted cripple wall based on experimental test data. Note that the system is very far from elastic for any deflections more than 1.0 cm. While this system is also far from perfectly plastic, it can be thought of as an elasto-plastic system (very approximately) with strains up to about 10% (deflection of 7 cm).

Before delving into nonlinear dynamics, I just want to remind everyone that the relation between maximum acceleration and maximum displacement of a 7-Hz harmonic oscillator is $\ddot{u}_{\max} = \left(2\pi \cdot 7\text{Hz}\right)^2 u_{\max}$. So if the maximum acceleration is 1 g then the maximum displacement is 0.5 cm. That is, if the structure was oscillating at 7 Hz with a maximum displacement of 2.5 cm (about 1 inch), then the maximum associated acceleration would be on the order of 5 g. Clearly, the natural frequency of 7 Hz cannot apply to cripple walls that are deforming by an inch or more.

Figure 6.49 shows the envelope of the force/deflection cycles shown in Figure 1, and it is probably a good approximation to a static pushover analysis of the unretrofitted cripple wall. Since the point of the analysis is to characterize the vulnerability of old unretrofitted houses, I will focus on the dynamics of unretrofitted test specimens.

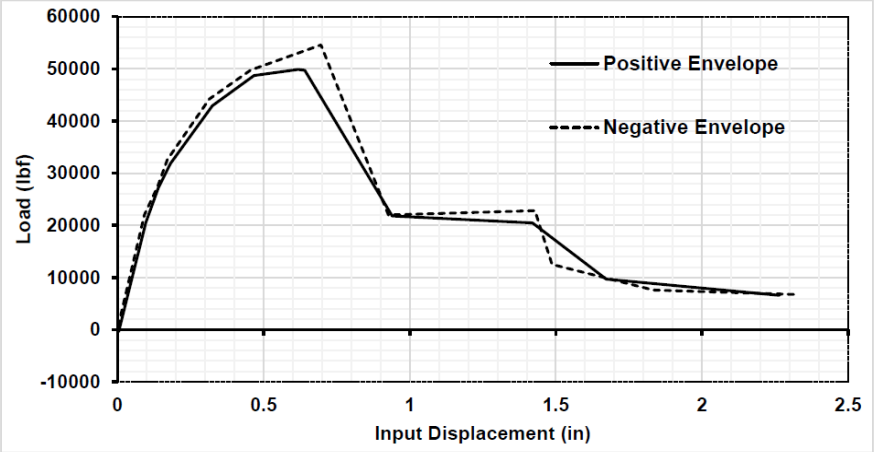


Figure 6.49 is also copied from the report of Working Group 4 and it shows that envelope of the force deflection curve in Figure 6.48.

Although an important objective of this study was to use experimentally consistent hysteretic models of force/deflection for dynamic analysis of wooden structures, I can gain some insight into the physics of the in-plane deformation of this cripple wall by using a very simple model of an elasto-plastic single degree of freedoom oscillator (sdof) that is driven harmonically (see Figure 6.50).

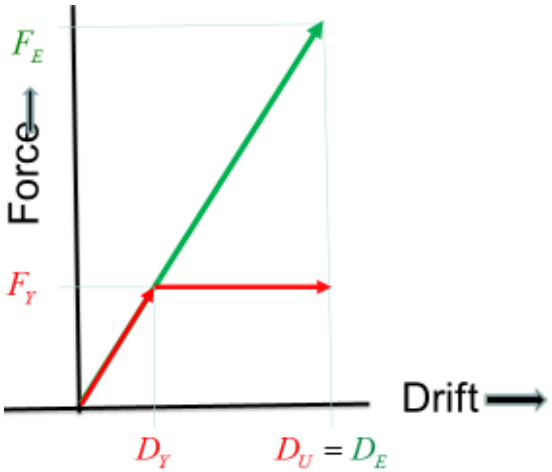


Figure 6.50. Definition of parameters for an elasto-plastic single-degree-of-freedom oscillator.

Consider an elastic-plastic single-degree-of-freedom oscillator with mass m, and a yield force F_Y that occurs when the deflection exceeds D_Y (see Figure 6.50). The elastic period of the sdof is

$$T_E = \frac{1}{2\pi}\sqrt{\frac{m}{k}} = \frac{1}{2\pi}\sqrt{m\frac{D_Y}{F_Y}} \tag{6.94}$$

Let D_U represent the maximum displacement of the system. If the deformation is large enough to yield the system, that is, if $D_U > D_Y$, then the approximate period of the plastic system is

$$\begin{aligned} T_P &= \frac{1}{2\pi}\sqrt{\frac{m}{k_P}} \\ &\approx \frac{1}{2\pi}\sqrt{m\frac{D_U}{F_Y}} \end{aligned} \tag{6.95}$$

where k_P is the equivalent (secant) stiffness of the elastic-plastic system. Now define the ductility of the system to be $R \equiv \frac{D_U}{D_Y}$, which is also sometimes called the "R-factor" in the design of elastic-plastic structures. In this case, D_U, is the maximum stable (i.e., ultimate) deflection of the system. Substituting into (6.87) shows that,

$$T_P \approx \frac{1}{2\pi}\sqrt{m\frac{RD_Y}{F_Y}} = T_E\sqrt{R} \tag{6.96}$$

That is, the plastic period increases as the square root of the ductility. In the case of Figure 6.49, the system shows strong strain-weakening behavior (progressive loosening of the nails) and the secant stiffness is more than a factor of 4 lower than for a perfectly plastic case. That is, for deflections greater than 4 cm (drift about 6%), the effective stiffness drops by more than a factor of 16, which would increase the natural period by a factor of 4. In the case of the PEER-CEA structures and drifts larger than 6%, $T_p \approx T_E R$. That means that the effective period of these houses would drop from 0.15 s. to about 0.6 s. Or put in another way, when a typically stiff house is shaken strongly enough to produce damage, then its effective period lengthens to be similar to that of a typical six-story elastic building.

Although the change in effective period is important, the change in the effective damping is even more important. Linear damping is most commonly interpreted as viscous damping; it's a kind of energy dissipation where the total system energy steadily decreases by a fixed fraction with every cycle of vibration. In that case, the fraction of critical damping ς of an sdof can be approximated by estimating the energy lost to dissipation in a cycle (See Chapter 1).

$$\begin{aligned}\varsigma &\approx \frac{1}{4\pi}\frac{\text{Energy dissipated in a cycle}}{\text{Elastic energy in cycle}} \\ &= \frac{1}{4\pi}\frac{\text{plastic work in a cycle}}{\text{Elastic energy in cycle}} \\ &\approx \frac{1}{4\pi}\frac{\text{plastic work in a cycle}}{\text{maximum kinetic energy}}\end{aligned} \tag{6.97}$$

Using the perfectly plastic model of Figure 3, I can estimate the plastic work in a deformation cycle as four times the product of the yield displacement, $\left(D_U - D_Y\right)$, with the plastic yield force F_Y. I can also approximate the sum of the kinetic and potential energies in the cycle to be the maximum kinetic energy during the cycle (this is strictly true for a linearly elastic system in free vibration). I will estimate the maximum kinetic energy by assuming that the maximum velocity of the mass is approximately the maximum displacement divided by the plastic period, or $\dot{x}_{\max} \approx 2\pi\left(\frac{D_U}{T_P}\right)$. (6.89) can then be rewritten as

$$\begin{aligned}\varsigma &\approx \frac{1}{4\pi}\frac{4\left(R-1\right)D_Y F_Y}{\frac{1}{2}\mathrm{m}\dot{\mathrm{D}}_U^2} \\ &\approx \frac{1}{4\pi}\frac{4\left(R-1\right)D_Y F_Y}{\frac{1}{2}\mathrm{m}\left(\frac{2}{\pi T_P}D_U\right)^2} = \frac{T_P{}^2}{\pi}\frac{\left(R-1\right)D_Y F_Y}{2\mathrm{m}\mathrm{D}_U^2} = \frac{T_P{}^2}{\pi}\frac{\left(R-1\right)F_Y}{2\mathrm{m}\mathrm{R}^2\mathrm{D}_Y} \\ &= \frac{T_P{}^2}{\pi}\frac{\left(R-1\right)}{2\mathrm{m}\mathrm{R}^2}\frac{m}{T_E^2} = \left(R-1\right)\frac{1}{2\pi}\left(\frac{T_P}{T_E}\right)^2 \\ &\approx \frac{R\left(R-1\right)}{2\pi}\end{aligned} \tag{6.98}$$

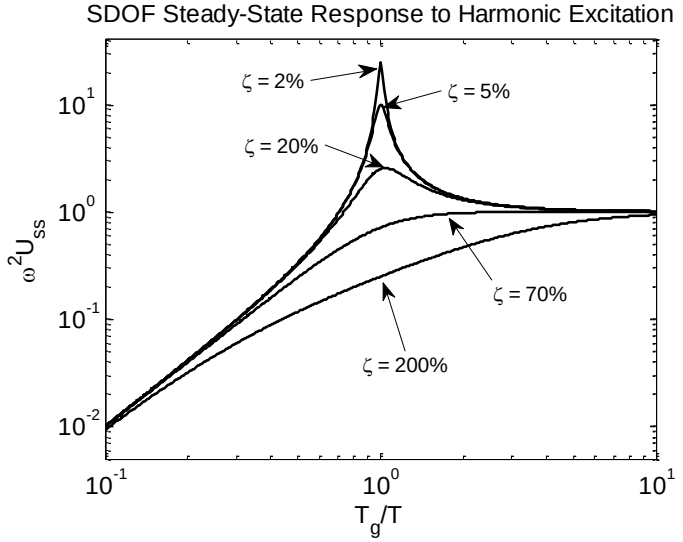


Figure 6.51. Response curves for different damping. There is no resonant peak for damping greater than 71% of critical. Notice that the slope of the log (acceleration response) vs log (period) is approximately linear with increasing driving period. This means that amplitude of the displacement response (i.e, the drift) scales linearly with ground velocity. Also, notice that when the damping is very large, the natural period of the oscillator is not very important. That is, the higher the damping, then the broader the frequency range that response measures velocity.

As we go to large ductility and long driving period, the fraction of critical damping goes way up. In fact, for the hysteretic loops shown in the CEA Wood report, the system is effectively overdamped. If the ductility in a damaging earthquake is 6.0, then the effective plastic period of a 0.2-s house would increase to about 0.8 s. That means that driving periods longer than 0.5 s would have the most importance to cause plastic drift. In this case the fraction of damping would be about 400%. While I admit that this is a very crude approximation, it seems clear to me that the effective damping is very large.

Lightly-damped systems (e.g. 5%) have peak responses that are relatively insensitive to the phase spectrum of the driving force (the ground acceleration); that is, for lightly damped response spectra, it doesn't matter too much whether all of the energy arrives in one pulse or whether the energy arrives randomly throughout the time period. Of course, the phase does have some influence response spectra; otherwise, we would use Fourier amplitude spectra.

Figure 6.51 shows the response curves of linear sdof's for different damping. As an example of the importance of damping, consider that modern seismometers are designed to be 71% damped sdof's. This is chosen so that there is no resonance at any period, thereby optimizing the bandwidth of motions that the seismometer records. This 71% damping also means that seismometers do a good job of recording seismic phases, which tend to arrive as pulses. Here is the thing about time-domain pulses, a pulse is inherently a broad-band phenomenon; pulses occur when a wide range of frequencies arrive in phase.

The use of highly damped response spectra to characterize motions that damage ductile buildings was suggested by Shiyan Song in his 2013 Caltech PhD thesis. Figure 6.52 is taken from his PhD thesis (Song, S., 2013, A New Ground Motion Intensity Measure, Peak Filtered Acceleration (PFA), to Estimate Collapse Vulnerability of Buildings in Earthquakes, Ph.D. Thesis, California Institute of Technology, Pasadena, CA.) and it shows a comparison between a variety of parameters and the base shear force in a simulated 6-story Steel MRF that is experiencing strong plastic yielding. This figure was the outcome of incremental

dynamic analysis. That is, the input ground acceleration was scaled to be just large enough to cause collapse ($P-\Delta$ instability). The top five traces are all on the same scale (force/weight, or % g). The second (yellow) trace is a 2%-damped sdof at the natural frequency of the building; the peak of this trace is identical to the response spectral acceleration for the record that is shown in the top (blue) trace. The true base yield force is the red trace. Notice that this base shear looks like a low-pass filtered version of the acceleration. Further, notice that the amplitude of the base shear (the red trace) is limited to be less than 23% of the weight of the building (this is the plastic pushover yield stress). Finally, the purple trace is a linear sdof with a natural period of 1.5 times the elastic period and with 71% damping. Notice that the purple and red traces are very similar when there is significant structural yielding. This result is anticipated by equations (6.86) and (6.88). This example shows why strong damping is a better way to parameterize strong shaking than is 5% damping. Coincidentally, the procedure for calculating a 71 %-damped response spectrum is identical to finding the maximum amplitude of a ground motion record that has been low-pass filtered with a 2-pole Butterworth filter (see Chapter 2). Therefore, Song referred to his damage parameter as *peak filtered acceleration* (pfa)

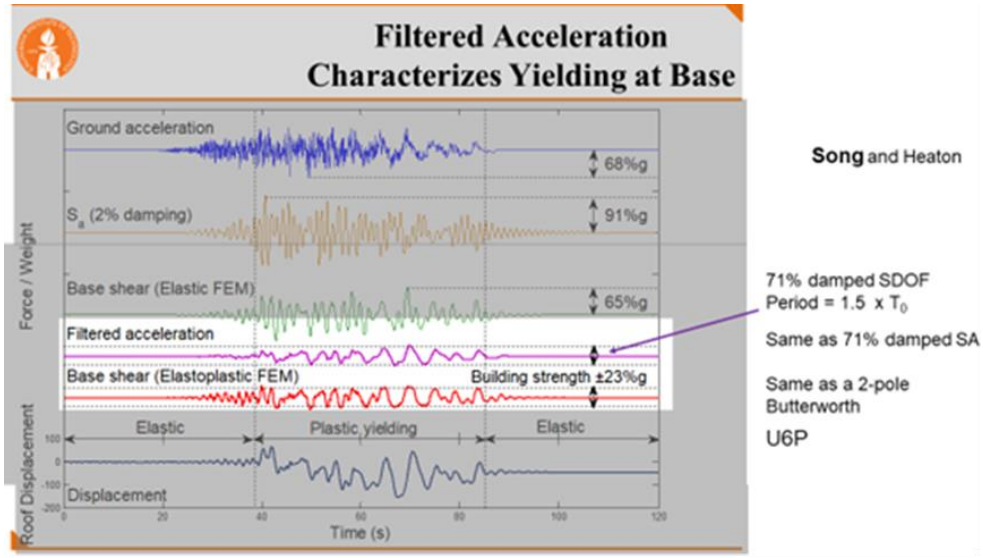


Figure 6.52. Is from the PhD dissertation of Shiyan Song (2013, Caltech) and it compares different time series that describe the response of a 6-story steel MRF building that experiences large plastic deformation (just smaller than $P-\Delta$ collapse). Whenever appropriate, all traces are on the same acceleration amplitude scale. The second (yellow) trace is the mass acceleration of a 2% damped linear sdof with the elastic free period of the building. The red trace is the base shear force of the yielding building given as a % of its weight. This base shear stress is limited to

23% of g, which is the plastic yield stress derived from push-over analysis.

Based on this insight, Song developed a procedure to predict the $P-\Delta$ collapse of buildings using parameters derived from a pushover analysis. Using full nonlinear analysis of steel mrf's, tested which simple parameters best predicted collapse reported by the complex analysis, He used a procedure called incremental dynamic analysis. In this procedure, a ground motion is used to compute the nonlinear response of a building model. The ground motion is then increased by a scale factor and the nonlinear response is calculated again. This procedure is repeated until the ground motion is large enough to cause simulated $P-\Delta$ collapse. This subject is more completely described in a recent paper by Buyco and Heaton (2019, 70%-Damped Spectral Acceleration as a Ground Motion Intensity Measure for Predicting Highly Nonlinear Response of Structures, Earthquake Spectra 35 (2), 589-610).

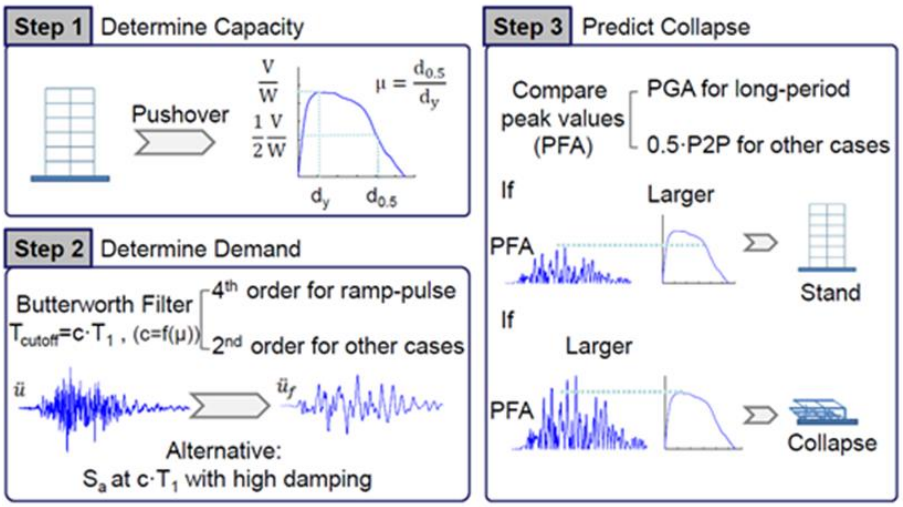


Figure 6.53 Procedure to predict the $P-\Delta$ collapse of a building from a nonlinear pushover analysis. The procedure is slightly different for motions that are dominated by a step in ground displacement.

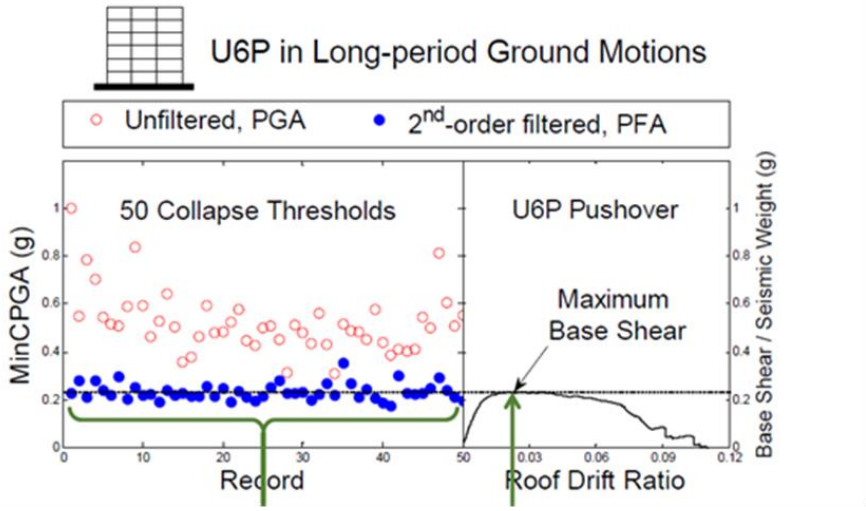


Figure 6.54 From Song's PhD dissertation. Vertical axis is horizontal acceleration in g. The red dots are peak accelerations of records that were just large enough to cause $P-\Delta$ collapse. The blue dots are the peak accelerations of the same records after filtering with a 2nd order low-pass Butterworth with a corner period of 1 ½ times the natural period. This is identical to 71% damped response spectral acceleration. The graph on the lower right shows the nonlinear pushover curve for the building. Notice that the amplitudes of the filtered accelertions that cause collapse are close to the peak static pushover acceleration of the building

.

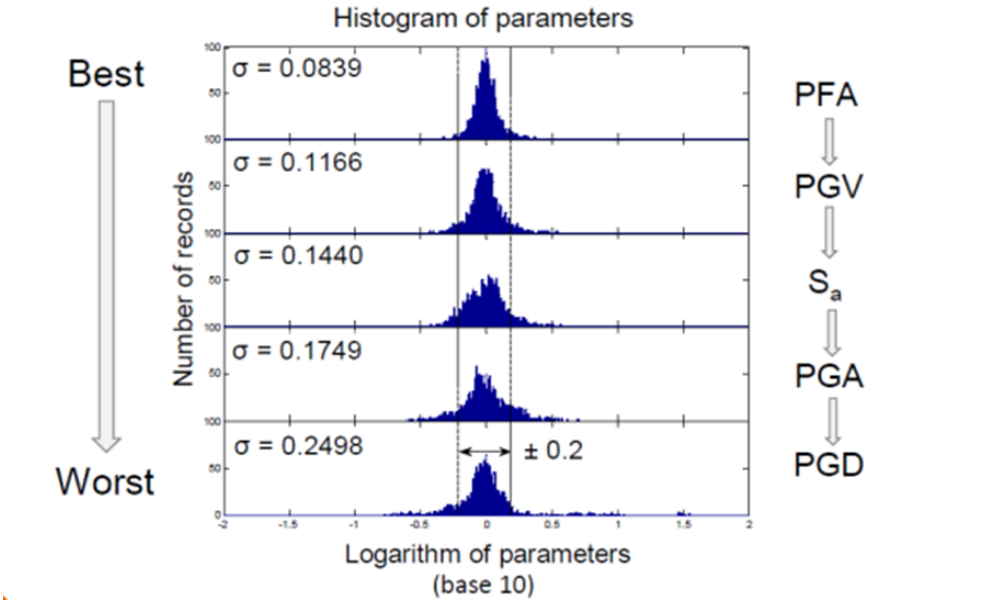


Figure 6.55. Dispersion of intensity measures that caused simulated collapse of a 6-story steel MRF. The intensity measures are ordered from best performance to worst. Increasing the damping to 71% (Peak Filtered Acceleration, PFA) decreases the dispersion by almost a factor of 2. σ is the standard deviation of the log normal distribution fit to each correlation.

Although 71% damped S_a at 1.5 times the natural period is the best predictor, pgv is a reasonable alternative to 71% damped spectra. This is verified in Song's study of collapse parameters (Figure 6.55). That is, pgv is the second-best intensity measure.

The correlation of damage (primarily to wooden houses) and pgv was also studied by Wald, Quitoriano, Heaton, and Kanamori (Spectra 1999). Modified Mercalli intensities (MMI) determined from observations of damage at particular sites was plotted against both pga (Figure 6.56) and pgv (Figure 6.58) for accelerograms recorded less than 3 km from the damaged structure. There is a lot of scatter in these plots and one could easily despair that there is nothing important to see. However, from the point of view of understanding structural damage, I believe that there is something to be learned from the sites that experienced MMI of VIII or greater. Sites that experience significant cripple wall damage would be assigned MMI's of VIII or greater. Notice that in Figure 6, the median pga for VIII and IX is virtually the same (about ½ g). Importantly, ½ g is the level of saturation of pga for all crustal earthquakes and it occurs primarily in the very near-source region (Joyner-Boore distance < 10 km) of earthquakes larger than M 6. For

example, Figure 6.57 shows all available pga's from very near-source stations in M>6 earthquake (as of about 2005). These data are well described by a log-normal distribution with a geometric mean of 47% g.

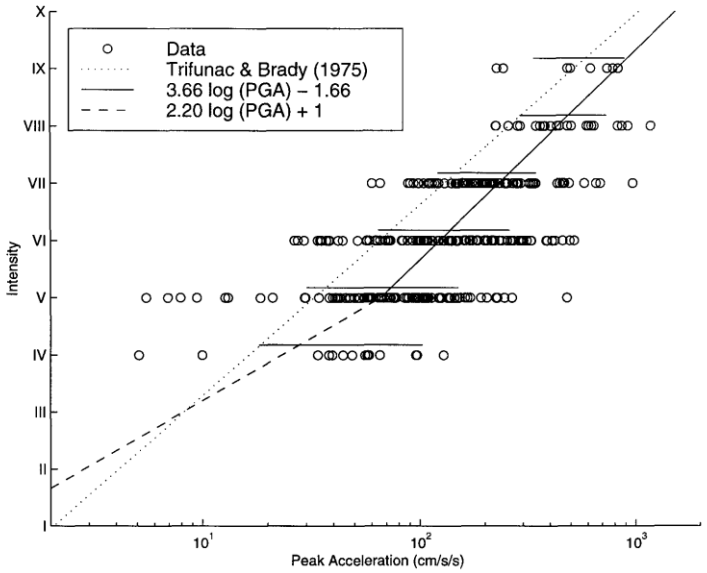


Figure 3. Modified Mercalli intensity plotted against peak ground acceleration for all events combined. Circles denote data; horizontal lines above data depict the range of the geometric mean, plus and minus one standard deviation. The solid line is regression from this study, the dashed line is assigned (see text for details). The dotted line is that of Trifunac and Brady (1975).

Figure 6.56. This is copied from Wald, Quitoriano, Heaton, and Kanamori (Spectra, 1999).

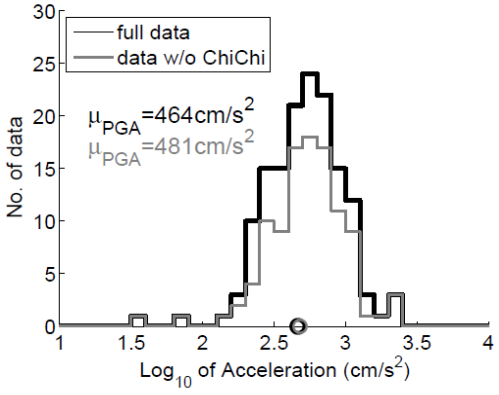


Figure 6.57. This is copied from Yamada, Olsen, and Heaton and it shows the pga for all near-source recordings (JB distance less than 10

km) of earthquakes larger than M 6 for all available data at the time of publication (2004). The grey histogram includes all data prior to the M 7.7 1999 Chi-Chi earthquake, and the black histogram includes the Chi-Chi data. Notice that the addition of Chi-Chi data had a neglibable effect on the geometric mean. That is pga saturates at a geometric mean of about ½ g; It seems likely that near-source pga from future earthquakes will also be log-normally distributed about ½ g.

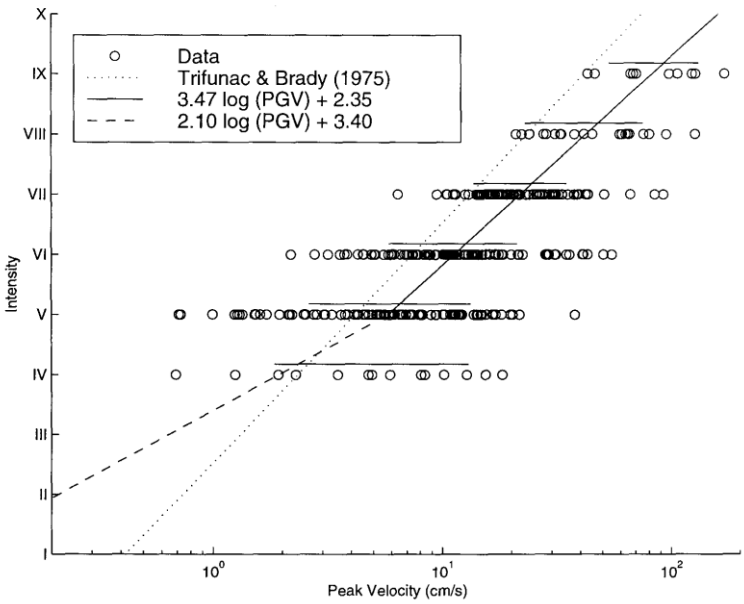


Figure 4. Modified Mercalli intensity plotted against peak ground velocity for all events combined. Circles denote data; horizontal lines above data depict the range of the geometric mean, plus and minus one standard deviation. The solid line is regression from this study, the dashed line is assigned (see text for details). The dotted line is that of Trifunac and Brady (1975).

Figure 6.58. Same as Figure 6.56, but for pgv.

In contrast, Figure 6.58 is the same as Figure 6.56, but for pgv. Notice that the median pgv for intensity IX is about 100 cm/s, whereas it is about 40 cm/s for intensity VIII. This is clear evidence that pgv is a better predictor of damage to wooden houses than pga is (or 7 Hz response spectral acceleration).

Reason for the Importance of pgv

During the early part of my career, peak ground acceleration (pga) was considered to be the most important measure of shaking intensity. There were lengthy debates about whether or not pga could exceed 1 g. We now have numerous well documented cases where the acceleration was substantially larger than 1 g. Interestingly, the damage to buildings adjacent to high acceleration records has typically been minor. It has long been noted that high peak ground velocity (pgv

greater than 1 m/s) is very damaging to buildings. This observation has often been viewed as counter intuitive; doesn't $f = ma$? Not really; $f = ma$ only applies to rigid bodies (or alternatively it applies when a is the acceleration of the center of mass).

I will now describe another way to understand why pgv is important for damaging buildings. Start with the equation of motion for a linear sdof

$$\ddot{u} + \ddot{x} + 2\omega_0 \varsigma \dot{x} + \omega_0^2 x = 0 \tag{6.99}$$

Where u is the ground motion (displacement), x is the motion of the mass relative to the base of the spring, ς is the fraction of critical damping, and ω_0 is the natural frequency of the undamped oscillator. When the damping becomes large, then the dissipation energy becomes large compared with the elastic energy in the spring. That is $\omega_0^2 x$ becomes smaller than other terms and (6.92) can be approximated (very roughly) as

$$\ddot{u} + \ddot{x} + 2\omega_0 \varsigma \dot{x} \approx 0 \tag{6.100}$$

Integrating with respect to time gives

$$\dot{u} + \dot{x} + 2\omega_0 \varsigma x \approx 0 \tag{6.101}$$

Now we want to find the maximum of x over time (the maximum drift). When x is a maximum, then $\dot{x} \approx 0$, and so

$$\dot{u}_{\max} + 2\omega_0 \varsigma x_{\max} \approx 0 \tag{6.102}$$

Or

$$x_{\max} \approx \frac{1}{2\omega_0 \varsigma} \dot{u}_{\max} \tag{6.103}$$

That is, we **expect the maximum drift of the plastic sdof to approximately scale with peak ground velocity**. This is consistent with my remarks in Figure 6.51 about the characteristics of the response of a linear sdof.

It amazes me that there still seems to be considerable attention given to pga and its related parameter, high-frequency response spectral acceleration. These parameters are widely thought to be important for nuclear power stations. Nuclear power stations are usually comprised of massive reinforced concrete shells and diaphragms that are exceedingly stiff. Although these structures are designed for much larger loads than what happens in earthquake shaking, the mechanical equipment that controls the station (e.g., pipes, pumps, motors, etc.) could be vulnerable. Personally, I think that it is best to invest in quality bushings and shock mounts that isolate critical equipment from high frequency shaking. That is, I am very skeptical that high frequency shaking is a significant problem for the design of nuclear power stations.

Figures 6.59 and 6.60 show the median behavior of near-source (rupture distance less than 5 km) pga and pgv as a function of earthquake magnitude. A variety of different studies are shown. It seems pretty clear that near-source pga saturates fully as a function of magnitude. In contrast, pgv shows a change in the magnitude scaling at about M 6 ½, but it does not appear to fully saturate.

As is often the case, the things that matter the most are also the things that we know the least about. Since the time of the creation of the ground motion prediction equations (gmpe's) shown in Figures 6.59 and 6.60, there have been a number of new large earthquakes around the World. Table 6.2 shows pga, pgv, and pgd from near-source records from recent large earthquakes. Clearly, future earthquakes will provide a new wealth of records with impressively large pgv and pgd. The column labeled as tilt is the size of a step in tilt during strong shaking that makes the double integration of acceleration a stable procedure (see Chapter 2).

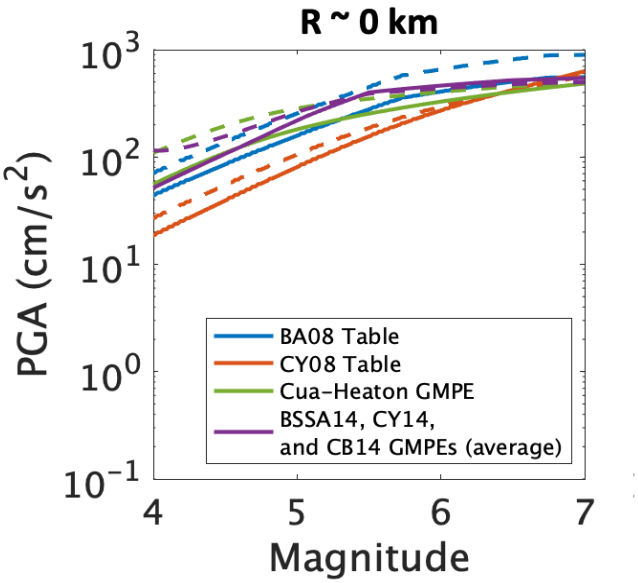


Figure 6.59. A variety of ground motion prediction equations (gmpe's) that show the best fit of near-source pga data as a function of magnitude. The Cua-Heaton and relationships with a 14 are the most recently published.

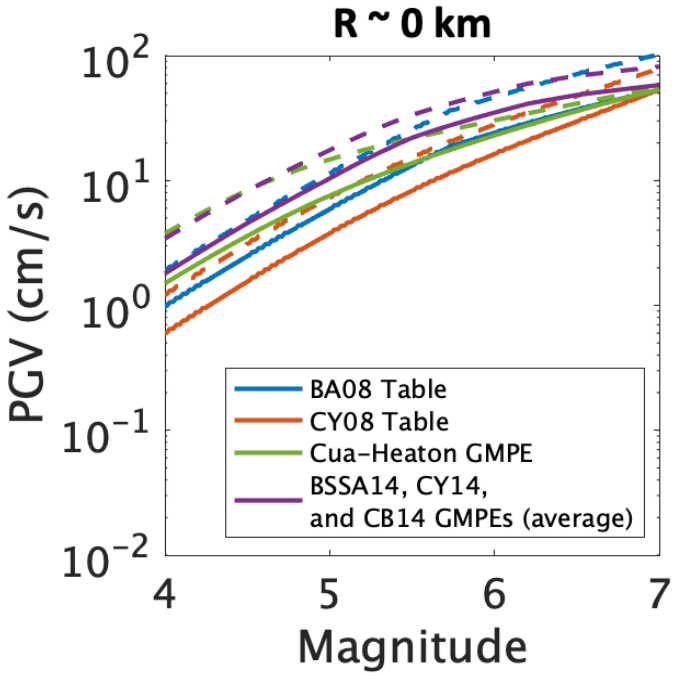


Figure 6.60. A variety of ground motion prediction equations (gmpe's) that show the best fit of near-source pgv data as a function of magnitude.

Design Response Spectra

5 % damped response spectra are typically very complex functions of period (or frequency). They are thought to provide estimates of the base shear stress for the fundamental mode. The fact that the spectra vary so strongly makes it difficult to determine a shear strength that is appropriate for the design of buildings. In reality, the narrow-band peaks really only apply to linear oscillators; nonlinear oscillators shift their resonance frequency as the amplitude increases. Furthermore, it's low damping that makes for resonant peaks in the first place.

In order to actually use response spectra for the design of buildings, seismic building codes use response spectra that are smoothed version of actual spectra. These spectra are not only smoothed, but they are decreased in amplitude because designers believe that buildings can withstand drift displacements that far exceed the displacement at which the building will begin to yield. Designers used this principle to introduce the **R-Factor** (or reduction factor) into building codes. They argued that buildings could survive drifts that were up to ten times larger than the drift at which the building begins to yield. Since the designs were based on linear modal analysis, they argued that response spectra that should be

assumed for design should not only be a simple smooth function, it should also be much less than observed response spectra that assume linear sdof's.

Figure 6.61 shows the response spectra of ground motions that were just large enough to cause $P-\Delta$ collapse of a simulation of a six-story steel mrf. These spectra are shown for both 5 % damping (red) and also 71 % damping (black), which is the same as a low-pass Butterworth filter. This figure is from Song's dissertation and it shows that use of 71 % damping produces spectra that are much more similar to the standard design spectrum (blue). When comparing the 71 % damped spectra to the standard design spectrum it's important to remember that the building's effective period should be 1.5 times that of the period used for the standard design spectrum.

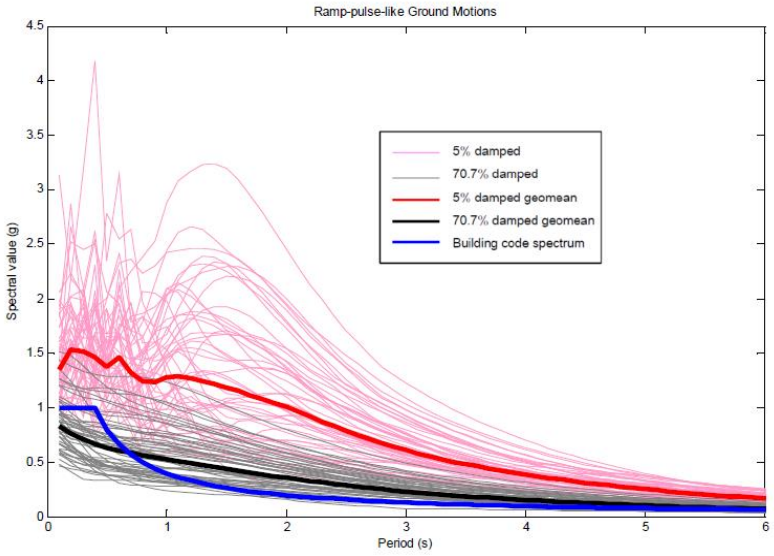


Figure 6.61 These are the response spectra of the 50 scaled records that were just large enough to cause $P-\Delta$ collapse in Song's study (see Fig. 6.54). The red curves assume the industry standard of 5 % damping, whereas the black curves assume 71 % damping. The heavy lines are the geometric means of the individual curves. The solid blue line is the spectral acceleration specified for the design of California buildings in high hazard regions.

Role of Peak Ground Displacement

There is yet another way to understand the vulnerability of buildings to $P-\Delta$ collapse. That is, there are two conditions that must occur for collapse. First, the building must experience inter-story plastic yielding, and second, the inter-story displacements must be large enough that the building is $P-\Delta$ unstable. In

particular, we can anticipate that 1) pgv must be large enough to cause elastic waves with large enough strains to cause yielding and 2) that simultaneously pgd must be large enough to make the building unstable.

As an example, I will return to the models of six- and twenty-story steel frame buildings. Pushover curves for variations of these buildings (e.g., US code and Japanese code) are shown in Figure 6.62. Notice that the Japanese code buildings have considerably larger yield strength than the comparable US code buildings. Also notice that the largest roof drift of both 20-story buildings is a little over 2 m. That is, these buildings are **gravitationally unstable if the roof moves more than 2 m** with respect to the base.

Anna Olsen did a full nonlinear finite element analysis on these buildings using 45,000 synthetic records of ground shaking that assumed a variety of rupture models for earthquakes as large as M 7 ¼ (Olsen, Heaton, and Hall, Spectra). These simulated ground motions were the result of a large collaborative project by the Southern California Earthquake Center (SCEC). The SCEC motions were low-pass filtered at 0.5 Hz (these were band-limited finite element models). Actual records with similar low-frequency amplitudes have pgv's that are about 50 % larger than those in this study.

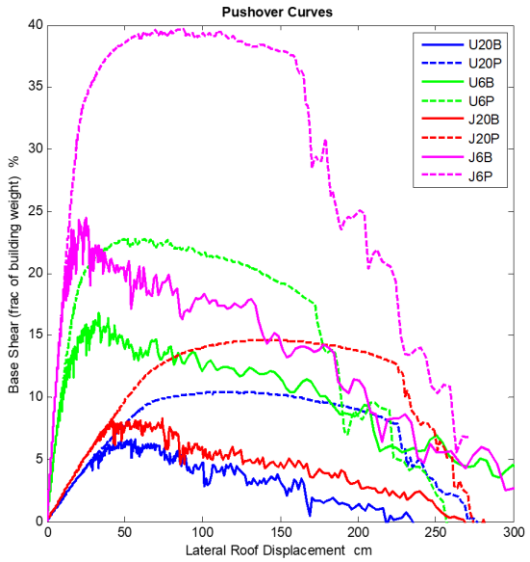


Figure 6.62. Pushover curves of the six- and twenty-story steel mrf buildings used by Olsen and Heaton to determine the nature of ground motions that cause damage and collapse.

Each of the 45,000 nonlinear simulations of the 20-story, US-buildings is represented by a dot in Figure 6.63. The location of each dot shows the filtered pgv and the broad-band pdg for each record. If the beams in the simulations have a permanent plastic strain of less than 8×10^{-3} then the building is represented by a grey

dot and it is called repairable. If the building experiences a side-sway collapse, then the dot is red. Finally, if the building is permanently bent more than $8x10^{-3}$, but it doesn't collapse, then the dot is black and the building is called unrepairable.

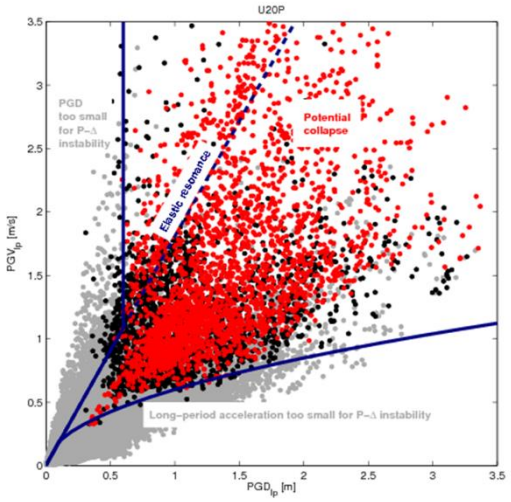


Figure 6.63 Data from simulations of the U20P building model response to long-period ground motions The building responses are: "standing" or "collapse" (left column) and "repairable" or "unrepairable" (right)responses are: "standing" or "collapse" (left column) and "repairable" or "unrepairable" (right). 32,000 simulated near-source ground motions are represented.

Notice that the red dots are generally in the upper right of the plot. Most collapses happen when pgv>60 cm/s **and** pgd > 60 cm. Remember that when an upward traveling shear wave reflects off the roof, the amplitude of the roof displacement is twice that of the ground displacement. That is, a pgd of 60 cm should produce a roof drift of about 120 cm, which is getting near the limit for $P-\Delta$ stability. The diagonal dotted line called elastic resonance is $pgv = \omega_0 pgd$, where $\omega_0 = 2\pi/T_0 = 1.8 \text{ radians}/s$. Notice that most of the collapses occur to the right of this line. That is, pgd is a very important parameter when it comes to collapse of tall buildings.

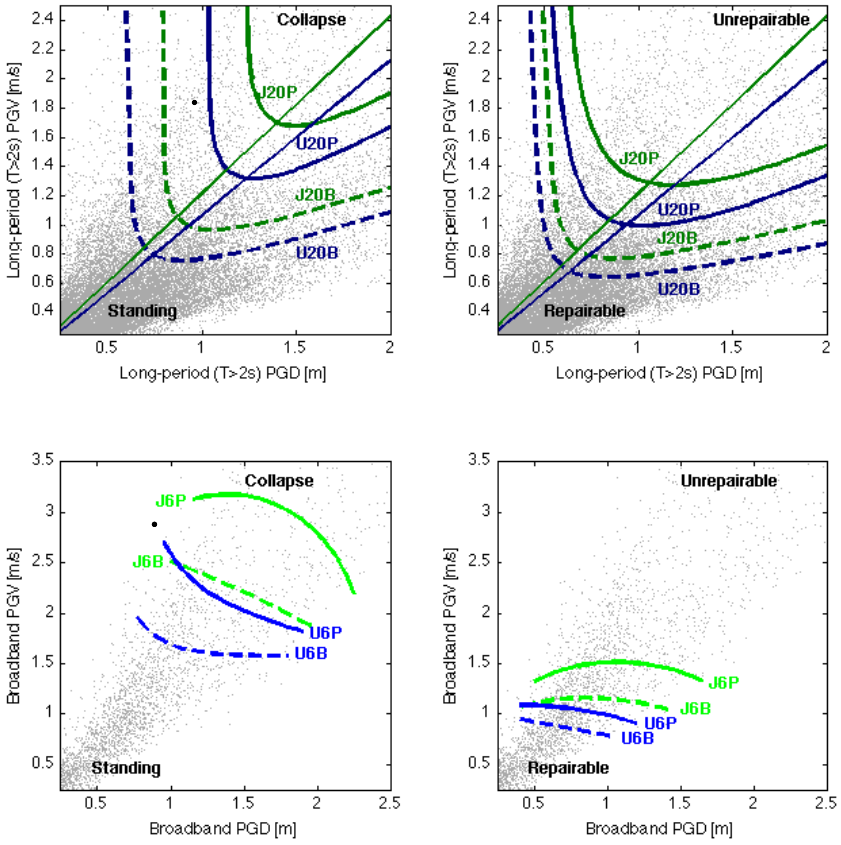


Figure 6.64. Contours where the probability of "collapse" (left) or "unrepairable" (right) is 30%. The gray points are the PGD and PGV values of the long-period (top) and broadband (bottom) ground motions. The equivalent PGV_{bb} is approximately $1.5 \cdot PGV_{lp}$ with a standard deviation of 0.24 m/s. The diagonal lines in the top plots are $PGV_{lp} = \frac{2\pi}{1.7 \cdot T} PGD_{lp}$, where T is the fundamental elastic period of the J20P or U20P building model.

Anna Olsen ran this same analysis for the Japanese code buildings and also for buildings with brittle welds. Figure 6.64. shows lines of pgv vs pgd that separate regions of collapse from regions of no collapse. The diagonal resonance line for the Japanese buildings are slightly steeper because the Japanese code results in a stiffer building ($T_0 = 3.0$ s for the Japanese design as opposed to $T_0 = 3.5$ s for the US). Notice that brittle welds greatly decrease the collapse resistance of both US and

Japanese designs. Also notice that the Japanese code results in a substantially better building than the US design.

I really like the clarity provided by Figure 6.64. It allows us to see what is the better design. Throughout much of my professional life, there was a debate about the merits of the Japanese building code vs the US building code (the Uniform Building Code, UBC). Many of my colleagues argued that while the Japanese code seems to produce a stronger building, it comes with the penalty that it is so stiff that it has higher forces than comparable US code buildings. Figure 6.64 seems to clearly settle that debate; you are more secure in Japanese code buildings. Unfortunately, most Japanese steel frame buildings built before 1995 suffer from the same type of brittle welds for steel buildings constructed prior to the 1994 Northridge and 1995 Kobe earthquakes.

The bottom panels of Fig. 6.64 a comparable analysis, but the buildings are all 6-story steel mrf's. Because the 6-story buildings are shorter period, the pgv values are increased by 50% to account for the fact that actual pgv from broadband records are larger than the pgv's from low-pass filtered records. Although both pgv and pgd are important for collapse of the 20-story buildings, pgv is the key parameter for the 6-story building. Again, the Japanese code building performs better than the US code building.

It's one thing to compare simulations of buildings that comply with different building codes and quite another thing to compare the actual performance of buildings in real earthquakes. I have visited Japan many times and I am always impressed when I see Japanese buildings that are under construction. They appear quite different to US buildings. In particular, Japanese steel construction often uses rectangular box sections for columns. This means that all sides of a column are flanges that can be used to construct a moment-resisting connection. It is common to see Japanese buildings in which all beam-column connections are moment resisting. In contrast, US construction typically uses I-sections for columns. That means that moment resisting connections can only be constructed in one direction. The net result is that there are far fewer moment resisting connections in a US building than in a comparable Japanese building.

Part of the story of the difference between US and Japanese buildings has to do with the building code. You might think that these codes are written similarly, but with different constants in the code. In fact, these codes are a set of complex rules and it's difficult to anticipate the characteristics of the resulting building until you go through the process of checking a design to ensure that it meets all of the code requirements. Personally, I find the details of building codes to be impenetrable. Fortunately, I have collaborated with colleagues (Prof. John Hall, Prof. Swaminathan Krishnan, Dr. Kenny Buyco) who have the patience and skill required to produce models that meet building codes. In the US, most structural engineering firms that design earthquake-resistant buildings aspire to produce designs that meet code requirements, but they also try to minimize construction costs by specifying designs that do not exceed code requirements. Importantly, US cities have not been

subjected to large near-source shaking (large pgv, pgd) since the 1906 San Francisco earthquake. In contrast, Japan experiences earthquakes at a rate that is roughly ten times higher than the west coast of the US. This means that Japanese buildings have been tested in the real world more frequently than US buildings. As a result, Japanese engineers are usually more cautious than US engineers.

In my view, the most important difference is that Japanese buildings are typically developed by companies that coordinate all aspects of building construction; architecture, geotechnical engineering, structural design, construction, and long-term maintenance is often handled by the same company. For example, Shimizu and Kajima are large construction companies that take a building from design to operation. These companies are proud of their buildings, and they openly show the buildings that are their product. If one of their buildings was destroyed by an earthquake, then this would be viewed as an embarrassment that would damage the reputation of the company. It is my view that many Japanese buildings exceed the required code to ensure the long-term viability of these construction companies.

Near-Source Ground Motions and Frame Buildings

Table 6.2 is copied from Buyco, Roh, and Heaton. This paper discusses how standard signal processing of accelerograms can decrease the likelihood that the record will cause simulated collapse of long-period building designs. The table lists records from many large earthquakes as well as estimates of the true pgv and pgd. Notice that there are numerous examples of recorded pgv > 1 m/s and pgd > 1 m.

Earthquake	M	Station	R_{JB}	PGA (g)	PGV (cm/s)	PGD (cm)	Tilt (°)
		CULC	15.6	0.27	29	75	0.62
		KEKS	3.0	1.97	269	867	1.54
2016 Kaikōura	7.8	KIKS	0.7	0.51	160	304	1.83
		WDFS	8.5	2.51	210	816	1.23
		WIGC	18.0*	0.75	64	52	2.73
		WTMC	0.7	1.12	117	284	0.01
		93048	0.6	0.79	264	186	0.58
		93051	0.5	0.84	178	105	0.48
		KMM001	5.0*	0.22	39	45	0.58
		KMM004	3.9*	0.35	82	74	0.12
2016 Kumamoto	7.0	KMM005	5.6	0.54	69	115	0.55
		KMM007	3.5*	0.43	44	40	0.27
		KMM009	2.2*	0.79	38	41	0.19
		KMMH16	0.5	1.18	142	228	0.08
		OIT009	7.8*	0.73	78	102	0.11
2015 Gorkha	7.8	KATNP	0.1	0.16	112	246	0.02
		AXT	9.8	0.29	31	105	0.04
2008 Wenchuan	7.9	MZQ	0.8	0.82	136	213	0.07
		SFB	4.8	0.58	81	318	2.04
2002 Denali	7.9	PS10**	3.0	0.33	137	302	–
		TCU052	1.8	0.45	225	740	–
		TCU065	2.5	0.79	135	198	–
1999 Chi-Chi	7.6	TCU067	1.1	0.50	100	191	–
		TCU068	3.0	0.51	298	885	–
		TCU084	11.4	1.00	118	251	–
1992 Landers	7.3	LUC**	2.0	0.76	146	263	–

* Epicentral distance is reported because R_{JB} is not available.
** Two horizontal directions are oriented parallel and normal to the ruptured fault.

Table 6.2. This table is copied from Buyco, Roh, and Heaton (2020, Spectra) and it lists the pga, pgv and pgd for near-source observations of earthquakes larger than M 7.

Figure 6.65 is from a study by Aagaard, Heaton, and Hall and it shows the median pgd as a function of distance from the surface projection of the rupture area of simulate M 7 earthquakes. This figure shows that a good rule of thumb is that the area directly above the rupture has a median pgd that is about 2/3 the average slip on the fault. The vertical scale is pgd per unit of average slip on the fault. Average slip scales as $10^{\frac{1}{2}M}$. That is, average slip increases by a factor of 3 for every unit increase in M. While table 6.2 has some very large pgd values, there are numerous examples of historic earthquake fault scarps that imply much larger pgd's.

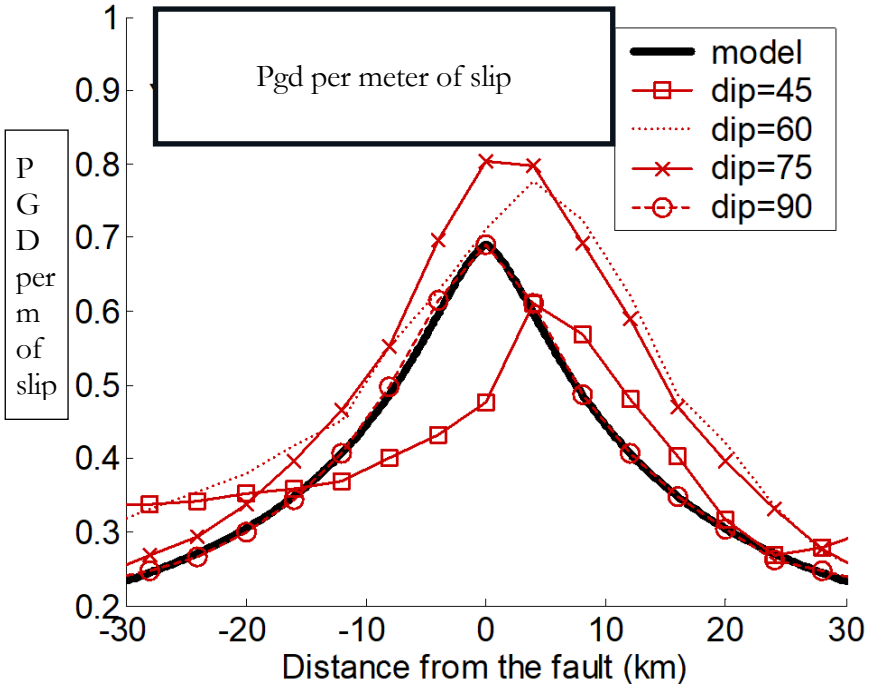


Figure 6.65. Summary of the amplitude of near-source pgd as a function of distance from rupture on simulations of long ruptures (from Aagaard, Heaton, and Hall). The x-axis is distance from the surface projection of the up-dip limit of rupture. The y-axis is the pgd for a spatially averaged slip of 1 m. pgd scales linearly with the slip.

Simulation of a repeat of the 1906 San Francisco earthquake

Figures 6.66 through 6.70 show simulations of M 7.8 earthquakes on the San Andreas fault in the San Francisco Bay region. Fault slips are inferred from real geodetic data from the M 7.8 1906 San Francisco earthquake. In addition, there are extensive observations of the offset of roads and fences that spanned the fault trace. These indicate that the average fault slip was about 7 m on the rupture north of the Golden Gate. In contrast, the section of the rupture south of the Golden Gate had an average slip of about 3 ½ m. The simulated ground motions are from finite-element models of a slip pulse (see Chapter 8) propagating at 85 % of the average S-wave velocity in the crust. This produces strong pulse-like ground motions for locations within 10 km of the rupture. The maximum near-source displacement are on the order of 2/3 the average slip, which is consistent with Fig. 6.65.

These figures show that the expected motions from such an event exceed the capacity of flexible frame buildings and base-isolated buildings that are constructed in California. The figures also show that pre-1994 steel frame buildings (brittle welds) are expected to collapse in the near-source region of such an event.

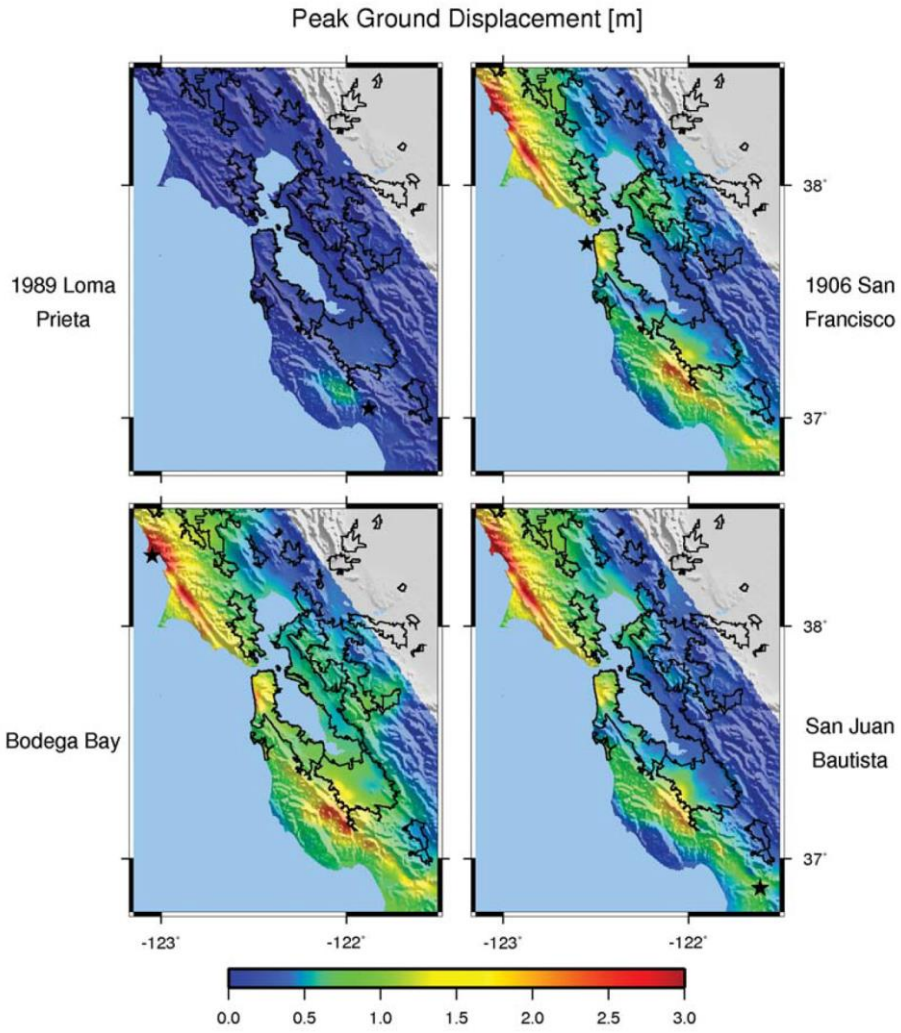


Figure 6.66 Peak Ground Displacement from simulations of San Francisco Bay area earthquakes. The upper left panel is based on a finite fault model of the 1989 M 6.9 Loma Prieta earthquake. The other panels show simulations of a M 7.8 earthquake on the San Andreas fault. The fault slip distribution is the derived from geodetic data from the M 7.8 1906 San Francisco earthquake, which indicates that the average fault slip was 7 m for the segments north of the Golden Gate and 3.5 m to the south. The scenarios are identical except for the location of the epicenter. That is, directivity effects are the only difference in these simulated motions (from Aagaard, 2007).

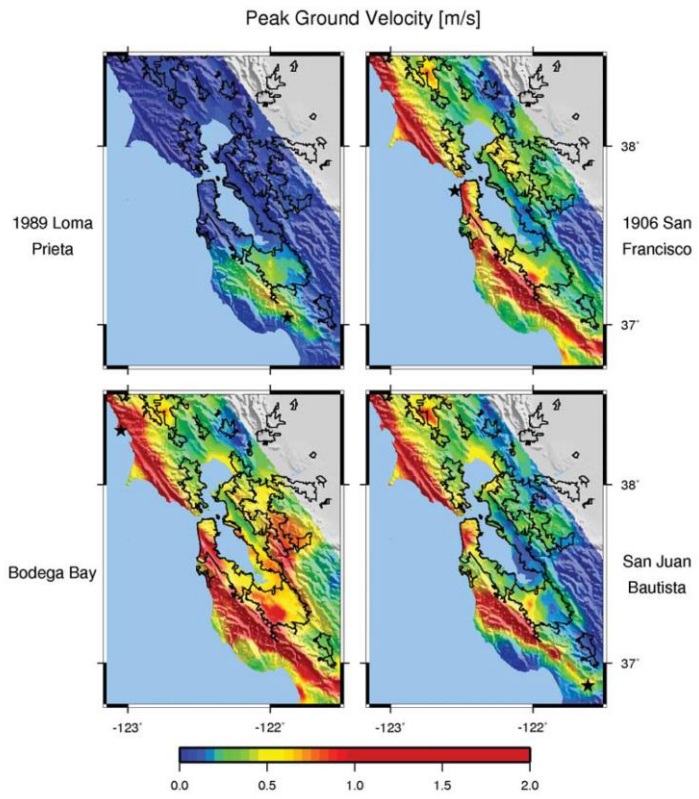


Figure 6.67. Peak ground velocities for simulated earthquakes in the San Francisco Bay region. The figure is the same as Fig. 6.66, except that pgv is shown.

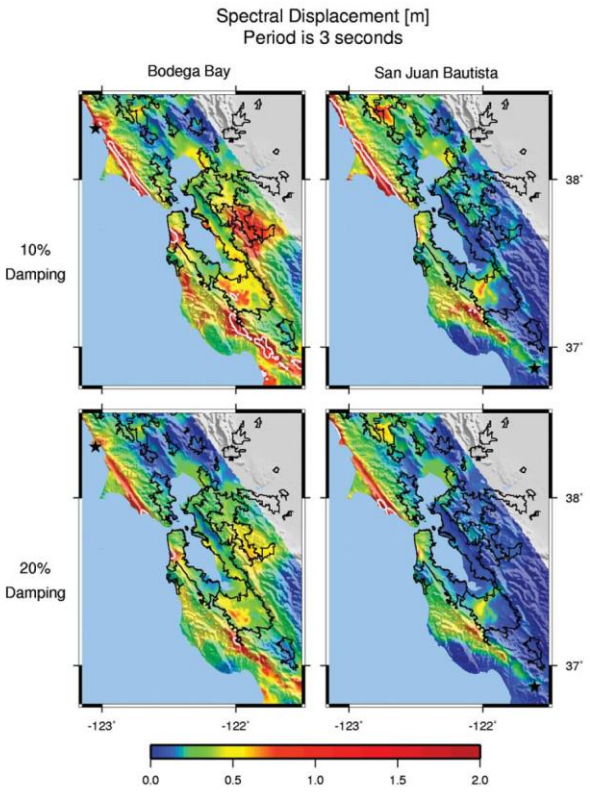


Figure 6.68. Response spectral displacement at 3 s for simulated ground motions shown in Fig. 6.66. The panels on the left assume an epicenter in Bodega Bay and the panels on the right assume an epicenter in San Jauan Bautista. The top panels and bottom panels assume 10 % and 20 % damping, respectively. Typical building isolators are expected to fail for spectral displacements of more than 50 cm (the yellow contour). Figure is from (Olsen, Heaton, and Aagaard, 3007).

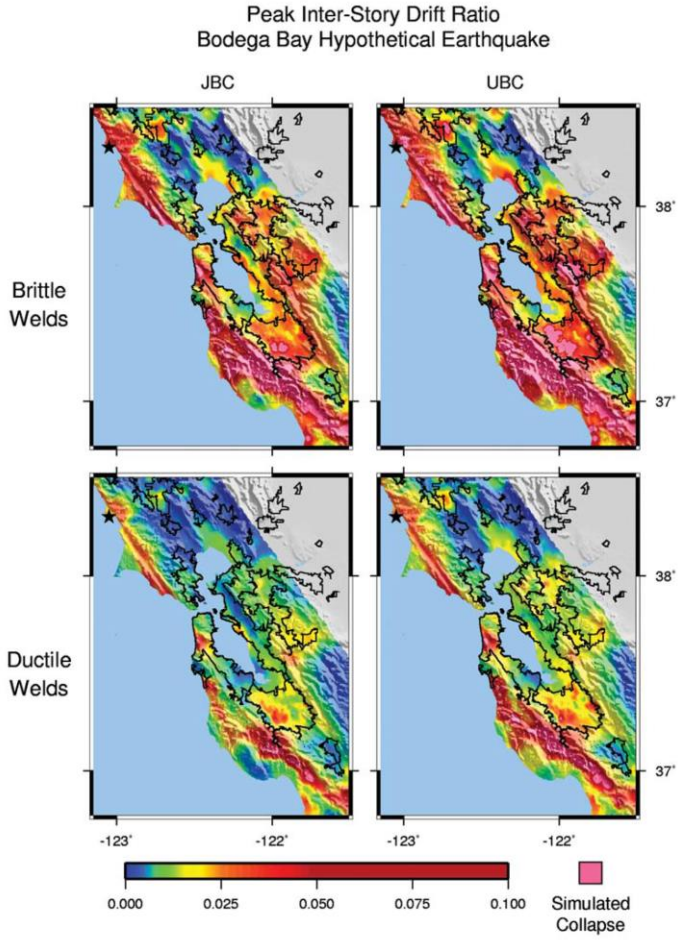


Figure 6.69. Peak inter-story drifts on nonlinear simulations of 20-story steel frame buildings that meet the UBC (1994) UA code (right panels) and the 1994 Japan building code (right panels). The top panels assume brittle welds and the bottom panels assume ductile welds that never fail. The simulated motions assume an epicenter to the north of the bay area. Building simulations are performed for all grid points in the finite element simulation (from Olsen, Heaton, and Aagaard, 2007).

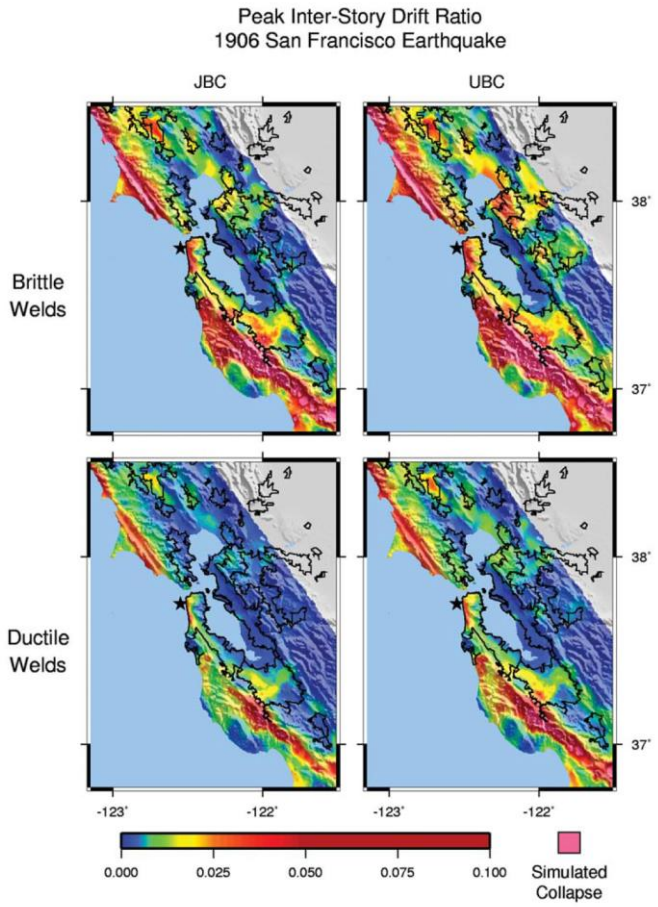


Figure 6.70. The same as Fig. 6.69, except that the epicenter is assumed to be in the Golden Gate. Directivity effects are less than those from a northern epicenter (Bodega Bay, see Fig.6.69)

US Building Code through the years

"Is my building safe in an earthquake?" I am frequently asked this question and I often reply with another question; "when was it built?" After Santa Barbara suffered significant earthquake damage in a M 5.9 in 1925, the first building codes were developed. Steel reinforcement of concrete structures was introduced about the beginning of the 20th century; the reinforced concrete U.S. Mint building performed better than other nearby buildings in the 1906 San Francisco earthquake. It is natural that the first codes focused on un-reinforced brick structures, and the first codes emphasized strength.

The simplest approach was to specify that buildings were approximately rigid and that the loads scaled with the pga. However, there was too little data to characterize the pga in damaging earthquakes. It was more practical to estimate the strength of buildings that had survived strong shaking. In this case the strength was defined to be the ratio of the maximum horizontal load, v, to the weight of the building, w. The 1946 Uniform Building Code (UBC) specified that v/w should not be less than 0.12, independent of the height of the building. Modal analysis was introduced in the next decade and the building's fundamental period was used to specify the minimum base shear. In particular, longer period buildings were allowed to have lower v/w. The El Centro strong motion recording of the M 7.1 1940 Imperial Valley earthquake was the first and only near-source recording of a major earthquake and it was considered to be representative of the strongest shaking possible. Unfortunately, we have since learned that this El Centro record was atypically small compared to other near-source records.

Figure 6.71 is from John Kenny Buyco's PhD thesis and it shows how v/w has evolved over time.

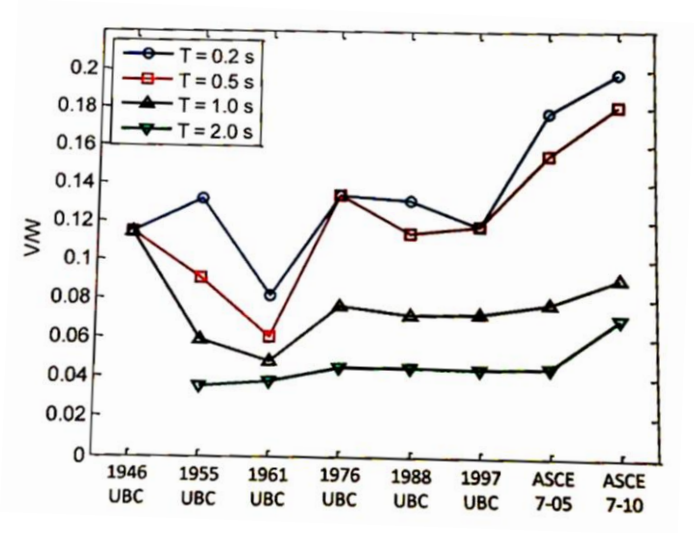


Fig. 6.71 is a simple representation of building code strength requirements. From 1946 through 2005, California cities used the Uniform Building Code that was developed by the Seismology Committee of the Structural Engineers Association of Southern California. In an effort to unify building codes throughout the US, FEMA coordinated a new code in 2007 (sometimes called the International Building Code, IBC) that was published by the American Society of Civil Engineers (ASCE).

Of course, building codes are complex documents that have evolved over the decades and Figure 6.71 does not fully describe building design through the years. For example, building designers must consider horizontal wind loads. In addition, a new type of building code known as Performance Based Earthquake Engineering (PBEE) has mostly replaced prescriptive codes since about 2010. Currently, most California cities allow designers to choose between the current prescriptive code (ASCE 7-10) and Performance Based Engineering. PBEE is intended to create structural designs that give acceptable performance in nonlinear analysis for suites of ground motions that are derived from site specific probabilistic seismic hazard analysis (PSHA). For example, a structural design begins with the creation of a 5 % damped design spectrum that is a composite model of hazard that is based on a statistical model of seismicity (what is the likely distribution of earthquakes on known faults?) and ground motion prediction equations. The PSHA is typically performed by a geotechnical consulting company that writes a report that is used by the structural engineers that design a building. The suite of design ground motions and the associated structural design are reviewed by a design review committee that is comprised of experts in engineering seismology and structural engineering.

When PBEE was being developed, it was often argued that it would provide the analysis that would motivate developers to design buildings to a higher performance standard than that required in the existing codes. For example, recent building designs for downtown Los Angeles state that the buildings should survive most examples of the strongest shaking to be expected in 2,500 years. This makes it appear that PBEE designs are better than older standards. I admit that I am agitated when a colleague describes the strength of a building in years. I want to see the nonlinear pushover analysis for individual buildings. Unfortunately, PBEE design procedures require that the technical reports that describe the actual capacities of buildings are proprietary. Since most recent buildings have been designed using PBEE, it is very difficult to know how the capacity of buildings has changed over time. Furthermore, many academic structural engineers are paid members of the review panels for important buildings. This means that the academic institutions that are responsible for basic research in earthquake engineering have key faculty who have signed reports that conclude that current designs are appropriate for thousands of years. Once you have approved the design of an important building, it's very difficult to change your mind. However, the prime directive of academics is to make the discoveries that drive change.

Chapter 5 of John Kenneth Buco's Caltech PhD thesis shows how the collapse capacity of simple nine- and twenty-story steel moment resisting frame buildings have changed over decades of different building codes used for downtown Los Angeles. Buyco used the different codes to vary the designs and he then subjected the different building designs to incremental dynamic analysis for four different recorded ground motions that were among the largest available. Buyco's analysis is thorough and difficult to summarize. I

suggest that you read his excellent thesis if you want to understand what he found. Figures 6.72 and 6.73 are a summary of his finding.

Buyco's analysis clearly shows that steel frames with brittle welds are significantly more vulnerable to collapse. If all moment resisting connections performed as intended, then there appears to be similar performance by the different codes through the decades. Even the performance-based designs seem to be comparable to older designs. I have often wondered why the new PBEE designs have chosen 2,500 years as a basis for design. I suspect that this choice ensured that the transition from rule-based codes to performance-based codes was designed to ensure that construction costs did not jump when the new design procedures were adopted.

It should be obvious that I oppose the adoption of PBEE. My objections are: 1) Current understanding of earthquake rupture physics is inadequate to calculate the size of infrequent ground shaking (see Chapter 8), 2) PBEE design procedures are so complex that almost no one really understands what the resulting building's capacity is, and 3) stating that we know how to design for 2,500 year motions gives the impression that earthquake engineering is a solved problem. That is, research universities should prioritize other societal problems (e.g., climate, energy, conflagrations, pandemics, etc.) since earthquake engineering is well enough understood to reduce it to a statistics problem.

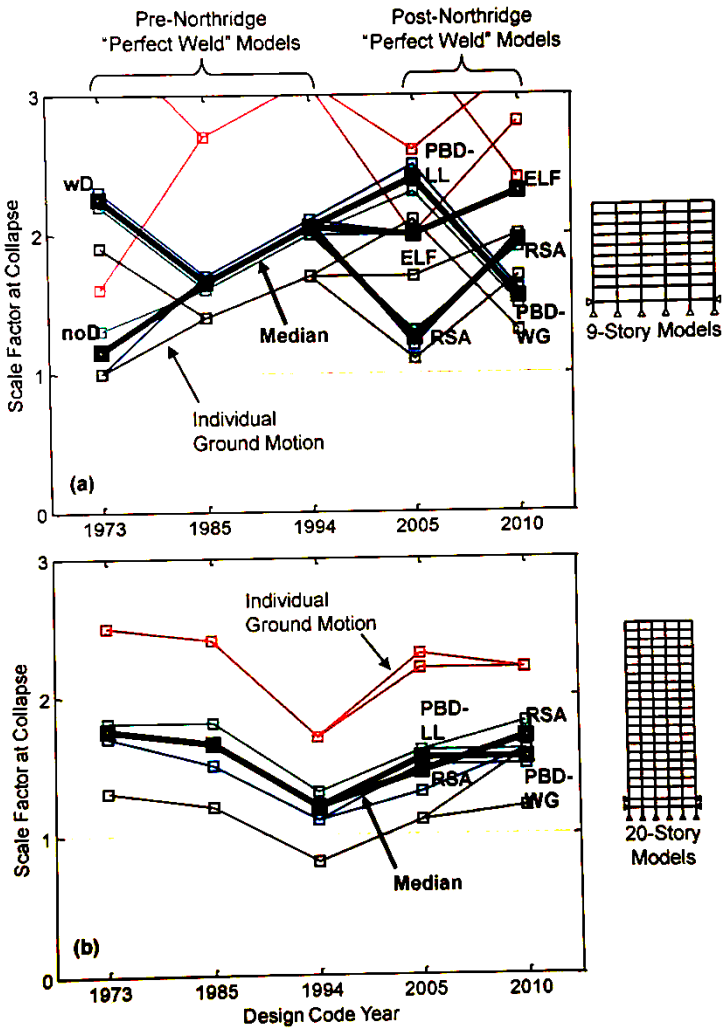


Figure 6.72 Collapse resistance of simple steel MRF buildings designed to different codes in downtown Los Angeles. The vertical scale is the amount that an individual strong motion record was increased by to cause P-delta collapse. Four different records were used for this incremental dynamic analysis. The analysis assumes that welds never fracture. PBD-LL and PBD-WG signifies Performance-Based designs for the LA-Live and Wilshire-Grand buildings respectively.

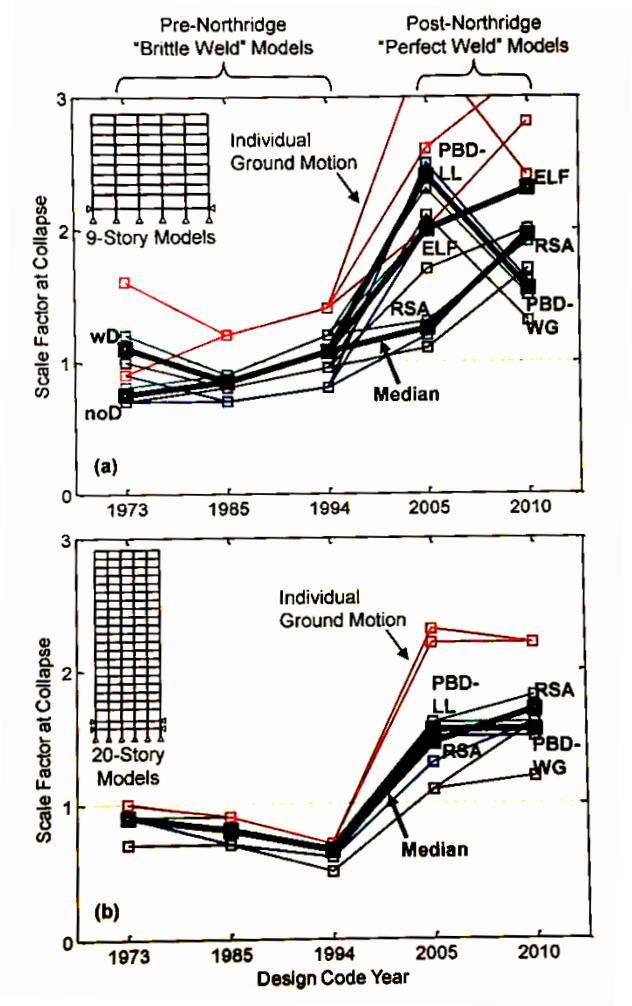


Figure 6.73. Collapse resistance of simple steel MRF buildings designed to different codes in downtown Los Angeles. This is the same as Figure 6.72, except that welded moment resistant connections are assumed to be brittle for designs prior to 1995.

Chile – a different approach

Following the realization that flexible buildings had lower forces than similar sized stiff buildings, U.S. codes bifurcated. That is, a designer could choose either a strong building or, alternatively, a flexible building. If a strong design was chosen, then it resulted in a stiff building with a lower fundamental period. Since most design response spectra have their peak at periods nears 1 s., strong buildings are required to be very strong. If a designer chooses a flexible design, then this comes with advantage that the building is longer period and the design spectral acceleration is decreased. When it comes to tall US buildings, almost everyone

chooses designs that are flexible. As I have argued throughout this chapter, making a tall building flexible makes it vulnerable to P-Δ collapse in near-source ground motions.

Chile is a country that has a long history of major subduction earthquakes. For example, the M 9 ½ 1960 Chilean earthquake is recognized as the largest earthquake in the past several centuries (see Chapter 8). Unfortunately, there were no strong motion recordings of the 1960 earthquake. Despite the paucity of 20th century Chilean strong motion records, it was clear that Chile experienced large, long-period ground motions. Chilean structural engineers responded by developing a robust building code that emphasizes high shear strength. Figure 6.74 shows the skyline of Valparaiso, which sits 20 km above the Chilean subduction zone.

Although the multi-story concrete buildings may look similar to US designs, they are actually very different. In particular, extensive reinforced concrete shear walls are typical for Chilean buildings. For example, Figure 6.75 shows the floor plan for a typical Chilean high rise. Notice that almost all of the walls are shear walls.

Figure 6.74. Skyline of Valparaiso, Chile, shows many multi-story reinforced concrete shear wall buildings that performed well in the 2010 M 8.8 Maule earthquake that ruptured 20 km directly beneath the city.

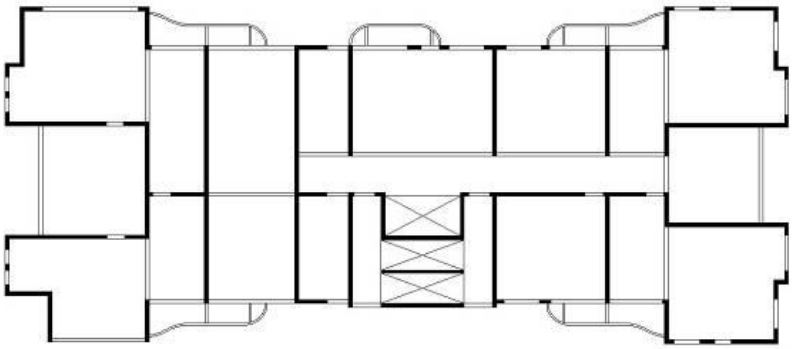


Figure 6.75. Floor plan for a typical Chilean high-rise building. The infilled lines represent reinforced concrete shear wall, whereas the open lines represent non-structural walls. Notice that most wall are shear walls. This is very different from typical US high rises where most walls are not structural.

Figure 6.76 shows the horizontal ground acceleration in Concepcion during the 2010 M 8.8 Maule earthquake. Concepcion is also 20 km above the rupture. The long duration of shaking is caused by the very long rupture (see Fig. 8.41). The duration of the record is 90 seconds, which tricks the eye into believing that this is high-frequency shaking. However, this record is dominated by low-frequencies as can be seen in the ground displacments in Figure 6.77.

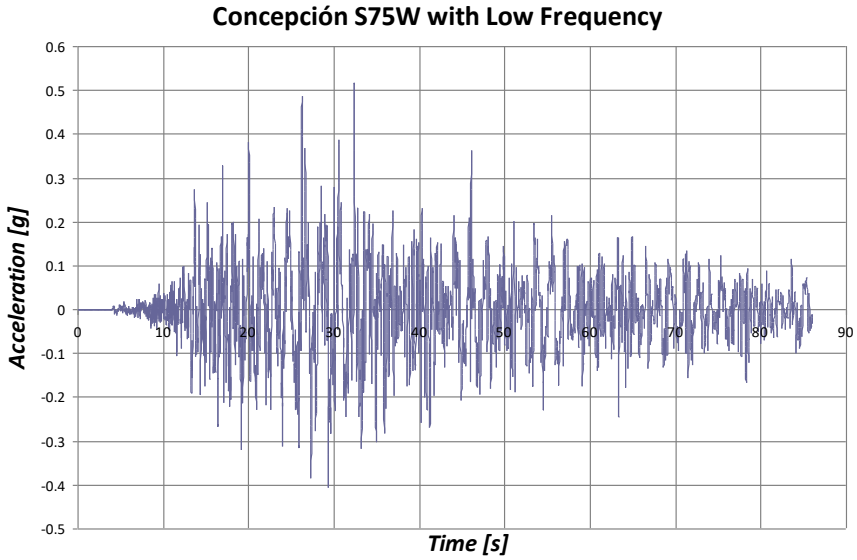


Figure 6.76. Gound acceleration recorded in Concepcion, Chile, during the M 8.8 2010 Maule earthquake. Notice the very long duration of long-period shaking.

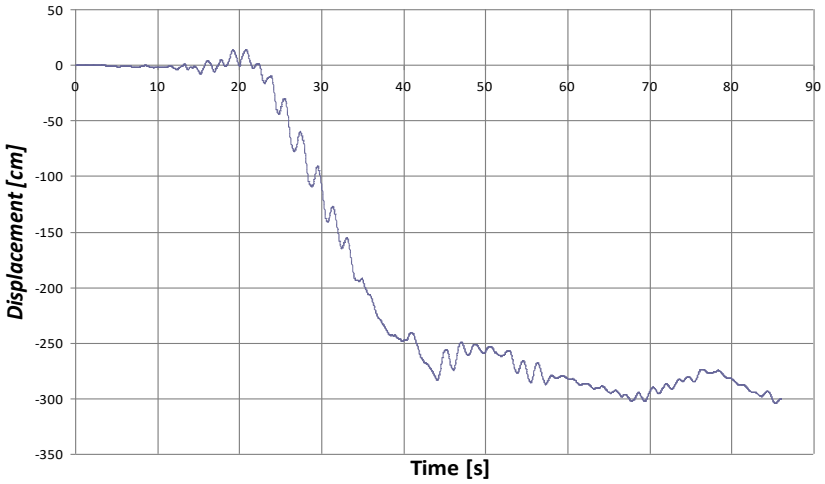


Figure 6.77. Ground displacement in Concepcion during the acceletation shown in Fig. 6.76. Continuously recording GPS satellite positioning is combined with the acceleration data to derive the displacment time history, Notice that the ground moved more than 2 ½ meters during the 20 second period that a slip pulse was propagating beneath the city (see Figure 8.41).

Although Chilean buildings performed well in the Maule earthquake, there were examples where the shear walls were severely damaged. Figure 6.78 shows a concrete shear wall in Santiago that was heavily damaged. As is often the case, Chilean real-estate developers have pushed for very tall buildings. The failure shown in Fig. 6.78 seems to be the consequence of pushing the height of shear wall buildings beyond practical limits.

Figure 6.78. Shear failure in a new 25-story shear wall building in Santiago during the 2010 Maule earthquake.

Figure 6.79 shows the collapse of a 15-story shear wall building in Concepcion. Post collapse inspection shows that the exterior concrete shar walls were

destroyed by longitudinal compression that occurred as the building rocked laterally. Amazingly, the building toppled sideways (like felling a tree) when the exterior walls lost compressional strength at the building's base. The building landed on top of a neighboring 5-story building. The 15-story building broke into two pieces, but most of theliving spaces remained in tact, saving the lives of residents in the toppled building.

Figure 6.79. Collapse of a shear wall building in the 2010 Maule earthquake. Notice that although the building collapsed, the high strength of the shear walls preserved the integrity of most of the interior rooms. Fortunately, the area on which the collapsed building landed was unoccupied.

Homework Chapter 6

1. Assume that the base of a shear-beam (neglect bending deformations) with rigidity μ and density ρ is subject to tangential displacement that given by $u_x(t)$. Assume that the top of the shear beam is a free surface.

$$u_x(t) = \begin{cases} 0, & t < 0 \\ t, & t > 0 \end{cases}$$

a) Write the motion of the free end of the beam as a function of time.
b) Write the shear stress at the forced end of the beam as a function of time.
c) How would this problem change if you were to allow bending deformations in the beam?

10/8/2024
Chapter 7. Earthquake Sources

Earthquakes are primarily caused by spontaneous dynamic slippage on a surface in the Earth (a fault). It turns out to be quite difficult to solve elasticity problems in which there is an arbitrary cut in the medium (a slipping fault). One common approach is to create a finite-element model of the volume that contains the fault slippage. Brad Aagaard's PhD thesis has an excellent description of this problem (Finite-element simulations of earthquakes, Aagaard, B, 1999, Finite-element simulations of earthquakes. EERL Report, 99-03. California Institute of Technology , Pasadena, CA., https://resolver.caltech.edu/CaltechEERL:1999.EERL-99-03).

Specialized elements that have split nodes create lateral offsets within the finite element grid. Except for these specialized split elements, all other elements are assumed to deform elastically. The slips at the fault elements can be determined in two ways, 1) the slip vector can be prescribed as a function of location and time (this is usually called a **kinematic source**), or 2) a prestress and a friction law is prescribed, and the slip vector is calculated (this is called a **dynamic source** model). Since both modeling approaches obey the laws of linear continuum mechanics, the motions that occur in their volumes are identical as long as the slip time histories are the same.

One of the difficulties with kinematic sources is that the modeler can choose slip vectors that are inconsistent with classical dynamics; for instance, one could choose a slip that reverses direction, or a slip that jumps instantaneously from one position to another. It is up to the modeler to decide what slip distribution is reasonable.

In contrast, dynamic models use the continuum equations to determine the slip history and therefore, they are often referred to as physics-based models. Of course, there is no free lunch; that is, the modeler must specify the prestress and the friction law. I will argue in Chapter 8 that neither of these quantities are known. Furthermore, dynamic modeling is considerably more difficult to interpret, and it is computationally expensive. Therefore, there are many applications that are more efficient using kinematic models. This is particularly true for source inversions, where the object is to find source slip histories that best simulate recorded seismic information. Seismograms are linear functions of slip, whereas they are highly nonlinear functions of prestress and friction. Linear inversions are typically fast, and more importantly, their solutions are unique.

Although finite element models are versatile, they can be cumbersome for problems with large spatial domains. There have been attempts to extend the

domain of finite-element solutions by using the motions on the boundary of the discrete domain as boundary conditions for analytic solutions, but that subject is beyond the scope of this class.

Seismologists simulated earthquakes decades before the advent of finite-element modeling. My favorite papers on this topic were both printed in the same 1964 issue of the Bulletin of the Seismological Society of America. Haskell (1964) described a method to simulate the far-field waves from dislocations (a truly profound paper), and Burridge and Knopoff (1964) described how different types of seismic sources could be simulated using equivalent forces. Reading these early papers provides the background to understand the meaning of the terms used in earthquake studies.

It is important to understand that researchers sought a methodology to modify known solutions to continuum mechanics so that they simulate earthquakes. An important example is shown in Figure 7.1.

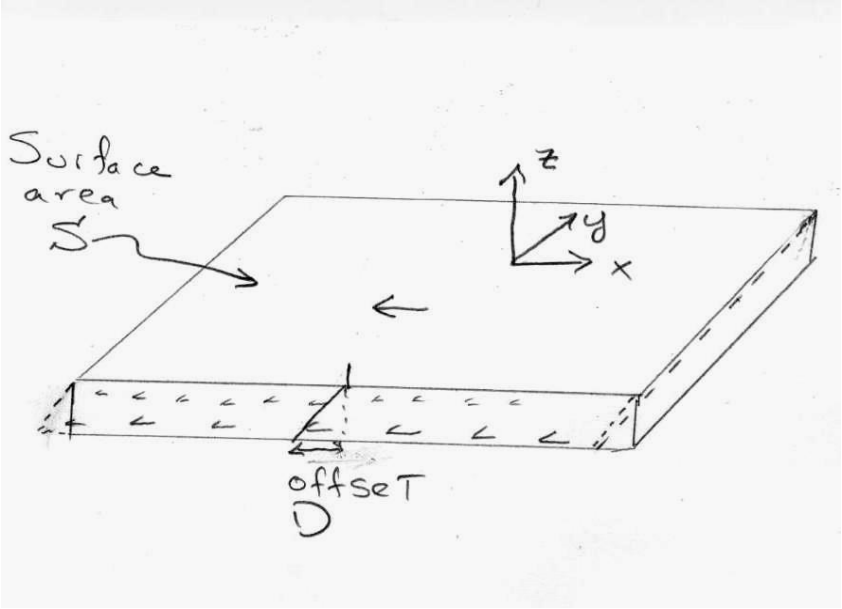


Figure 7.1 Mathematical idealization of an "earthquake fault." In this case we have an elastic material in which we induce a shear strain inside a tabular body. The opposing surfaces of the tabular body are offset by the induced shear strain. We then shrink the thickness of the tabular body to zero but increase the shear strain so that the net offset is constant; that is, we make the shear strain an impulse in the z-coordinate.

It is usually assumed that the Earth is linearly elastic everywhere, except in the immediate vicinity of the failure surface. It is very difficult to find analytic

equations to a continuum that has a surface with discontinuous displacement (fault slip). To find analytic solutions, it is convenient to draw a tabular box around the failure surface as is shown in Figure 7.1. In fact, it is usually far easier to solve problems in which we apply stresses in such a manner that the medium deforms "as if" it had a cut along some plane. Now if strains are small and if the material is isotropic, then the strain anywhere in the medium can be related to stress by

$$\varepsilon_{ij} = \frac{1+\nu}{E}\sigma_{ij} - \frac{\nu}{E}\delta_{ij}\sum_{k=1}^{3}\sigma_{kk} \tag{7.1}$$

where ν and E are Poisson's ratio and Young's modulus, respectively. We can also write (7.1) in terms of the Lame constants, where

$$\nu = \frac{\lambda}{2(\lambda+\mu)} \tag{7.2}$$

$$E = \frac{\mu(3\lambda+2\mu)}{\lambda+\mu} \tag{7.3}$$

A little algebra gives

$$\varepsilon_{ij} = \frac{1}{2\mu}\sigma_{ij} - \frac{\lambda}{2\mu(3\lambda+2\mu)}\delta_{ij}\sum_{k=1}^{3}\sigma_{kk} \tag{7.4}$$

Now if we were God, then we could apply a stress $\boldsymbol{\sigma}^0(x,y,z)$ within the tabular box. Now the stress at any point inside the box will be the sum of the stress that we applied at that point $\boldsymbol{\sigma}^0(x,y,z)$ plus the stresses $\hat{\boldsymbol{\sigma}}^0$ at (x,y,z) that are caused by application of stress $\boldsymbol{\sigma}^0(\hat{x}\neq x,\hat{y}\neq y,\hat{z}\neq z)$ at all other points than (x,y,z). That is, the stress anywhere in the medium can be decomposed as

$$\boldsymbol{\sigma}(x,y,z) = \boldsymbol{\sigma}^0(x,y,z) + \hat{\boldsymbol{\sigma}}^0(x,y,z) \tag{7.5}$$

The terms $\hat{\boldsymbol{\sigma}}^0(x,y,z)$ are actually very complex and, for each point in space, they involve integrating Green's functions for stress over the entire box. However, we will assume that our box becomes vanishingly thin. In fact, we will assume that

$$\boldsymbol{\sigma}^0(x,y,z) = \boldsymbol{\sigma}^0(x,y,0)\delta(z) \tag{7.6}$$

where $\delta(z)$ is an impulse function in z. Now as it turns out, the Green's function for stress at (x,y,z) caused by applying another stress at $(\hat{x},\hat{y},\hat{z})$ is a complex function with a scaling of $|\sigma| \sim r^{-3} = \left[(x-\hat{x})^2 + (y-\hat{y})^2 + (z-\hat{z})^2\right]^{-3/2}$. This means that if (7.6) is true, then within the box, the stresses are completely dominated by the impulsive stress applied at each point within the box (that is, stress decays as the inverse cube of the distance from the point at which the stress is applied) and

$$\boldsymbol{\sigma}(x,y,z \approx 0) \approx \boldsymbol{\sigma}^0(x,y,0)\delta(z) \tag{7.7}$$

Now if we are interested in the relative displacement vector $\mathbf{D}^z$ of the two opposite faces (perpendicular to the z direction) of the tabular box, then we can integrate the strain from one side of the box to the other. The strain in the box is just

$$\varepsilon^0_{ij} \approx \frac{1}{2\mu}\sigma^0_{ij}(x,y)\delta(z) - \frac{\lambda}{2\mu(3\lambda+2\mu)}\delta_{ij}\sum_{k=1}^{3}\sigma^0_{kk}(x,y)\delta(z) \tag{7.8}$$

Since we are interested in the jump in displacement $\mathbf{D}(x,y)$ that occurs from the top to the bottom of the box, we can write that

$$\begin{aligned}
D_x(x,y) &= \int_{-\varepsilon}^{+\varepsilon} 2\varepsilon_{xz}(xy)\,dz = \frac{\sigma^0_{xz}}{\mu} \\
D_y(x,y) &= \int_{-\varepsilon}^{+\varepsilon} 2\varepsilon_{yz}(xy)\,dz = \frac{\sigma^0_{xy}}{\mu} \\
D_z(x,y) &= \int_{-\varepsilon}^{+\varepsilon} \varepsilon_{zz}(xy)\,dz = \frac{1}{2\mu}\sigma^0_{zz}(x,y) - \frac{\lambda}{2\mu(3\lambda+2\mu)}\sum_{k=1}^{3}\sigma^0_{kk}(x,y) \\
&= \frac{\sigma^0_{zz}(x,y) - \nu\sigma^0_{xx}(x,y) - \nu\sigma^0_{yy}(x,y)}{E}
\end{aligned} \tag{7.9}$$

Here we have used the fact that the spatial integral of a delta function is a step function.

If we assert that an earthquake is the horizontal sliding along a fault surface, we find that we can produce a jump in shear displacement across our thin tabular box by applying a shear stress equal to

$$\sigma_{xz}^{0}(x,y,z) = \mu D_x(x,y)\delta(z)$$
$$\sigma_{yz}^{0}(x,y,z) = \mu D_y(x,y)\delta(z) \tag{7.10}$$

We can also produce a jump in the displacement normal to the tabular box (e.g. opening of a crack) by applying extensional stresses inside the box.

I have just described a clever method for deriving the elastic response of a medium to inelastic deformation across some surface. Instead of solving for the continuum response of that inelastic slipping region, we replace it with a region that is assumed to be linearly elastic. We introduce singular stresses on that surface (delta functions in space) that stretch the elastic surface so strongly that there is a net displacement across the stretched region. Of course, this is a physical fiction; we are using infinitesimal linear elasticity to calculate the response of infinitely large strains within the box. However, we don't really care about what is exactly happening inside the box. We just want there to be an offset of one side of the box relative to the other.

If we can calculate the response of the rest of the medium to the introduction of the applied stresses in the source region, then we can calculate the response of the rest of the medium (outside the box) to a jump in the boundaries of the box. This source modeling trick is sometimes referred to as a **dislocation source**. This trick of applying singular stresses to produce slip is also often used in finite-element modeling. That is, forces are applied to grid points in such a way that they produce jumps in displacement between grid nodes. This can save effort by eliminating the need for split nodes.

It is customary to parameterize the "size" of the dislocation source by integrating the applied stress tensor $\boldsymbol{\sigma^0}(x,y,z)$ over space. This integrated stress is called the **Moment Tensor**; that is stress integrated over a volume has units of stress times volume, or units of force times length. These are the units of torque. That is, sources are often described by a torque (hence the origin of the word, moment). **DO NOT MISTAKE SEISMIC MOMENT FOR AN ENERGY. Although torque and energy seemingly have the same units, torque is a force and energy is … well it's energy.**

$$\mathbf{M_0} \equiv \int_0^l \int_0^w \int_0^h \boldsymbol{\sigma^0}(x,y,z)\,dx\,dy\,dz \tag{7.11}$$

If the source is a simple shear dislocation across a boundary then we can substitute (7.10) into (7.11) to obtain

$$\mathbf{M_0} \equiv \int_0^l \int_0^w \int_0^h \mu \mathbf{D}(x,y)\delta(z)dxdydz = \int_0^l \int_0^w \mu \mathbf{D}(x,y)dxdy \tag{7.12}$$

The size of the moment tensor is often described with the simple scalar seismic moment, defined by

$$M_0 \equiv \mu S \bar{D} \tag{7.13}$$

where S is the rupture area and $\bar{D}$ is the average slip on the rupture surface. As we will see in a second, the seismic moment is the size scaling factor for long wavelength waves. The key elements in the scaling factor are the average slip, and the rupture area. This makes physical sense since our medium is assumed to be linear. Double the slip and we expect to double the ground motions. Also, for long wavelength waves, doubling the rupture area should also double the ground motions.

Confusions about seismic moment

While we expect ground motions to scale linearly with rupture area and average slip, there is great confusion about the meaning of including rigidity in our size scaling factor called seismic moment. In particular, the μ is the value of rigidity inside our infinitesimally thin box. It has no real physical meaning, since we only use it to determine the size of the infinite elastic strain that is calculated using infinitesimal elasticity inside of our thin source box (a true oxymoron). Unfortunately, it has become common to hear the phrase "moment release." However, earthquakes do not release moment. When people talk about how much moment is released, what they really mean to convey is the slip D integrated over the rupture area S. This quantity has been called **Potency** P by Ben Menahem and Singh (1981, *Seismic Waves and Sources*, Springer-Verlag, New York, 1108 pp) and it has units of strain integrated over a volume, which is units of volume.

$$P \equiv \int_0^l \int_0^w \int_0^h \mathbf{D}(x,y)\delta(z)dxdydz = \int_0^l \int_0^w \mathbf{D}(x,y)dxdy = S\bar{D} \tag{7.14}$$

We can also define a **Potency History** $P(t)$ as follows.

$$\mathrm{P}(t) = \int_0^L \int_0^W D(x,y;t)dxdy \tag{7.15}$$

Point Sources

Although it is clear that earthquake sources occur primarily as slip on a surface (a fault), it is often convenient and adequate to approximate the source as having infinitesimal rupture area, but finite Potency or Seismic Moment. The solution for this problem is much simpler than that of a source with a finite rupture area. Furthermore, the point-source solution can be used as a Green's function to construct the solution to more complex sources since the medium is considered to be linear away from the fault surface.

This type of source was first introduced by Love in his classic book, A Treatise on the Mathematical Theory of Elasticity in 1927. He used the term, **nucleus of strain**, for such a point source. By that, he meant a singular strain could be inserted at a point in a medium. In order to produce finite displacements at points away from the source, it is necessary for the source to have units of volume (potency). If the material is considered to be linearly elastic, then we can also introduce a point source defined by the stress tensor at the source point, which is called the **Moment Tensor**. Of course, this is even more fictitious than before since we are using infinitesimal strain theory to describe triply singular stress and strain. However, at distances large compared to the rupture lengths of a particular earthquake, and at frequencies that are low compared to times required by seismic waves to travel across the source region, a point approximation of the source is adequate for simulating earthquakes in elastic media.

It is helpful at this point to realize that the Moment tensor is simply a spatial integral of a stress tensor that is applied to the medium. Like any stress tensor, it can be decomposed into an isotropic stress $\boldsymbol{\sigma}_{\mathrm{H}}$ (sometimes called hydrostatic stress) and a deviatoric stress $\boldsymbol{\sigma}'$.

$$\boldsymbol{\sigma}' = \boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{H}} \tag{7.16}$$

where

$$\sigma_{\mathrm{H}\,ij} = \tfrac{1}{3}\sigma_{kk}\delta_{ij} \tag{7.17}$$

In stress tensors, the hydrostatic stress gives a measure of the pressure at a point and the deviatoric stress is used to describe the shearing in the medium. In a similar way, we can decompose the moment tensor into an isotropic component and a deviatoric component. Now, if an earthquake is composed of unidirectional slip on a plane, then the shear strain and stress necessary to simulate that slip is very simple in the coordinate frame of the fault (see equation (7.10)). It is given by

$$\mathbf{M}_o = \begin{bmatrix} 0 & \mu S\bar{D} & 0 \\ \mu S\bar{D} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{7.18}$$

Transforming this stress to the principal coordinate frame, which is oriented at 45^o to the fault coordinate frame, the stress is written as

$$\mathbf{M}_o = \begin{bmatrix} \mu S\bar{D} & 0 & 0 \\ 0 & -\mu S\bar{D} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{7.19}$$

A point source of this type is commonly called a **double-couple source**. Unfortunately, this term has also created tremendous confusion. That is, the word **non-double-couple source** is common in geophysical literature, and it is often taken to mean sources that are not well described as slip on a plane (e.g., an explosion). However, the word non-double-couple source technically means a source that is not fit with a simple shear at a point as given in (7.18) or (7.19). Earthquakes can be non-double-couple in the following ways. 1) if the earthquake has one fault plane, but the slip vector changes direction from one location to another, 2) The earthquake consists of more than one fault plane where the planes are neither parallel nor perpendicular 3) The fault is too large to be approximated as a point. In this sense, almost all earthquakes are non-double-couples. It is usually best to think clearly about what is really meant when someone tells you that the source was non-double-couple.

Reciprocity for a point force

At this point, I would like to discuss the property of reciprocity. Roughly speaking it says that the source and receiver can be interchanged for any linear continuum. Let us begin with a static deformation problem. Suppose we have a linearly elastic medium that is surrounded by a boundary that is either traction free or rigid (this is known as a **homogeneous boundary condition** and it implies that no work is done at the boundary). Now suppose that we apply point forces at two locations, $\mathbf{x}$ and ξ, as is shown in Figure 7.2. While the direction of the force is arbitrary, let us assume that the force at $\mathbf{x}$ is applied in the i^{th} direction and that an equal amplitude force is in the j^{th} direction at $\boldsymbol{\xi}$. That is assume that

$$\begin{aligned} \mathbf{F}^{\xi} &= F\mathbf{e}_i \\ \mathbf{F}^{x} &= F\mathbf{e}_j \end{aligned} \tag{7.20}$$

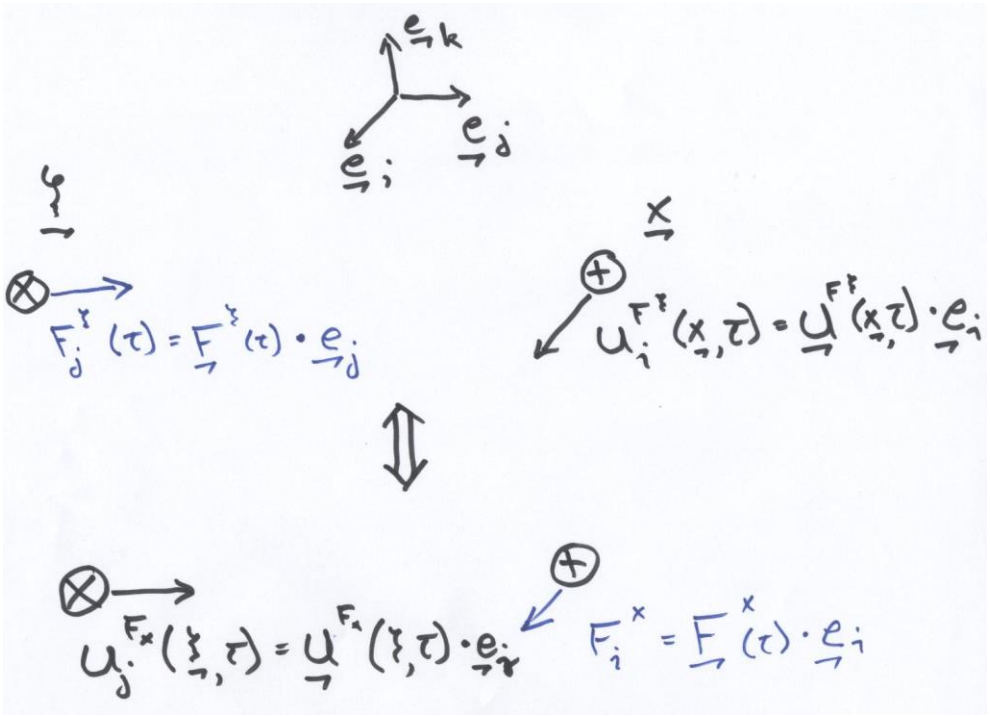


Figure 7.2. The displacement time history observed in the i[th] direction at $\boldsymbol{x}$ due to a point force in the j[th] direction at $\boldsymbol{\xi}$ is identical to the displacement time history in the j[th] direction observed at $\boldsymbol{\xi}$ due to a point force in the i[th] direction at x. The amplitude of the two point forces are assumed to be identical. In this example i=2 and j=3.

Assume that the displacement due to the application of this force is $\mathbf{U}^{F^x}$ and $\mathbf{U}^{F^{\xi}}$. We first apply $\mathbf{F}^{\xi}$ at $\boldsymbol{\xi}$ and since the medium is linearly elastic and no work is done at the boundary, the work done by this force is just the force dotted with the displacement at $\boldsymbol{\xi}$, or

$$\begin{aligned} W\left(F_j^{\xi}\right) &= \frac{1}{2}\mathbf{F}^{\xi} \bullet \mathbf{U}^{F^{\xi}} \\ &= \frac{1}{2}F_j^{\xi}U_j^{F^{\xi}}\left(\boldsymbol{\xi}\right) \end{aligned} \tag{7.21}$$

We now apply the second force $\mathbf{F}^x$ at $\mathbf{x}$. The work done by application of the second force is now the change in displacement $\mathbf{U}^{F^x}\left(\mathbf{x}\right)$ at $\mathbf{x}$ dotted with $\mathbf{F}^x$, but now plus the additional work done by $\mathbf{F}^{\xi}$ through the displacement $\mathbf{U}^{F^x}\left(\xi\right)$ at $\boldsymbol{\xi}$. That is,

$$W\left(\mathbf{F}^{\xi} \text{ then } \mathbf{F}^{x}\right)=\frac{1}{2}\mathbf{F}^{\xi}\bullet\mathbf{U}^{F^{\xi}}\left(\xi\right)+\frac{1}{2}\mathbf{F}^{x}\bullet\mathbf{U}^{F^{x}}\left(\mathbf{x}\right)+\mathbf{F}^{\xi}\bullet\mathbf{U}^{F^{x}}\left(\xi\right)$$
$$=\frac{1}{2}\left(F_j^{\xi}U_j^{F^{\xi}}\left(\xi\right)+F_i^{x}U_i^{F^{x}}\left(\mathbf{x}\right)\right)+F_j^{\xi}U_j^{F^{x}}\left(\xi\right) \tag{7.22}$$

Now if, instead, we first applied $\mathbf{F}^{x}$ at $\mathbf{x}$, and then we apply $\mathbf{F}^{\xi}$ at $\boldsymbol{\xi}$, then we can calculate the total work as

$$W\left(\mathbf{F}^{x} \text{ then } \mathbf{F}^{\xi}\right)=\frac{1}{2}\mathbf{F}^{x}\bullet\mathbf{U}^{F^{x}}\left(\xi\right)+\frac{1}{2}\mathbf{F}^{\xi}\bullet\mathbf{U}^{F^{\xi}}\left(\mathbf{x}\right)+\mathbf{F}^{x}\bullet\mathbf{U}^{F^{\xi}}\left(\mathbf{x}\right)$$
$$=\frac{1}{2}\left(F_j^{\xi}U_j^{F^{\xi}}\left(\xi\right)+F_i^{x}U_i^{F^{x}}\left(\mathbf{x}\right)\right)+F_i^{x}U_i^{F^{\xi}}\left(\mathbf{x}\right) \tag{7.23}$$

Now since the total energy of the system is independent of the order in which we apply the forces,

$$W\left(\mathbf{F}^{x} \text{ then } \mathbf{F}^{\xi}\right)=W\left(\mathbf{F}^{\xi} \text{ then } \mathbf{F}^{x}\right) \tag{7.24}$$

We conclude that

$$\mathbf{F}^{\xi}\bullet\mathbf{U}^{F^{x}}\left(\xi\right)=\mathbf{F}^{x}\bullet\mathbf{U}^{F^{\xi}}\left(\mathbf{x}\right)$$
$$F_j^{\xi}U_j^{F^{x}}\left(\xi\right)=F_i^{x}U_i^{F^{\xi}}\left(\mathbf{x}\right) \tag{7.25}$$

Now since the forces at the two sites are equal in amplitude, this means that the displacements at the two sites must also be equal in amplitude. That is, the displacement in the ith direction at $\mathbf{x}$ due to a force in the jth direction at $\boldsymbol{\xi}$ is the same as the displacement in the jth direction at $\boldsymbol{\xi}$ due to a force in the ith direction at $\mathbf{x}$. This can be generalized to the dynamic displacements that occur in a linear medium. For derivations see Chapter 2 of Aki and Richards (Quantitative Seismology, second edition) or see section 15.6 of Fung (Foundations of Solid Mechanics).

Now if we write the displacements using Green's functions $G_{ij}\left(\mathbf{x},t;\boldsymbol{\xi}\right)$ for a point force in the jth direction at $\boldsymbol{\xi}$ that produces a displacement time history in the ith direction at **x.**

$$U_i^{F^{\xi}}\left(\mathbf{x},t\right)=G_{ij}\left(\mathbf{x},t;\xi\right)*F_j^{\xi}\left(t\right) \tag{7.26}$$

No summation convention! If we switch the locations of the source and observer, but we now assume that the point force is in the ith direction (the direction of the observed ground motion at x) and that the observed motion at $\boldsymbol{\xi}$ is in the jth direction (the direction of the applied force at $\boldsymbol{\xi}$) we obtain

$$U_j^{F^{x}}\left(\boldsymbol{\xi},t\right)=G_{ji}\left(\boldsymbol{\xi},t;\mathbf{x}\right)*F_i^{x}\left(t\right) \tag{7.27}$$

Since the system is linear, the virtual work obtained by the product of the force and the displacement must be the same if the same amplitude force is applied

along the displacement at the source and observer points. That is, similar to the static case, but now the work is the integral over time.

$$\int_{-\infty}^{\infty} F_j^{\xi}(t) U_j^{F^x}(\boldsymbol{\xi},t)\,dt = \int_{-\infty}^{\infty} F_i^{x}(t) U_i^{F^{\xi}}(\mathbf{x},t)\,dt \tag{7.28}$$

where

$$F_j^{\xi}(t) = F_i^{x}(t) \tag{7.29}$$

Since (7.28) is true for any time history for the force, it must be true that

$$U_j^{F^x}(\boldsymbol{\xi},t) = U_i^{F^{\xi}}(\mathbf{x},t) \tag{7.30}$$

Substituting (7.26) and (7.27) into(7.30), we get

$$G_{ji}(\boldsymbol{\xi},t;\mathbf{x}) * F_i^{x}(t) = G_{ij}(\mathbf{x},t;\boldsymbol{\xi}) * F_j^{\xi}(t) \tag{7.31}$$

From which we can conclude that

$$G_{ji}(\boldsymbol{\xi},t;\mathbf{x}) = G_{ij}(\mathbf{x},t;\boldsymbol{\xi}) \tag{7.32}$$

This is the usual statement of reciprocity. If we substitute (7.32) into (7.27), we get

$$U_j(\boldsymbol{\xi},t;\mathbf{x}) = G_{ij}(\mathbf{x},t;\boldsymbol{\xi}) * F_i^{x}(t) \tag{7.33}$$

Reciprocity for force couples

Suppose that we consider the displacement at x caused by a force couple (see figure 7.3).

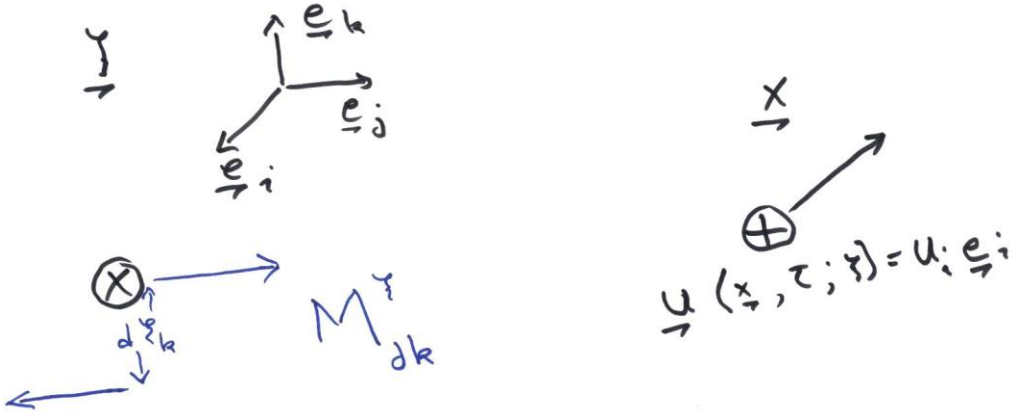


Figure 7.3. Displacement observed at x due to a force couple at $\boldsymbol{\xi}$. Specification of the orientation of the force couple requires two vectors, the direction of the force and the direction of the moment arm.

Now if the source is a force couple, $M_{jk}^{\xi}(t) \equiv \frac{\partial F_j^{\xi}(t)}{\partial \xi_k}$, and if our observer is measuring displacement at x, then

$$U_i^{M^{\xi}}(\mathbf{x},t) = M_{jk}^{\xi} * \frac{\partial G_{ij}(\mathbf{x},t;\xi)}{\partial x_k} \tag{7.34}$$

In this case, we cannot exchange the source and the receiver. To begin with, one cannot derive the virtual work done by a torque operating on a displacement. In addition, the orientation of the force couple requires two vectors, the orientation of the force and the orientation of the gradient. However, the displacement only describes one orientation. In the case of force couples, reciprocity must be stated in terms of displacement gradients as shown in Figure 7.4.

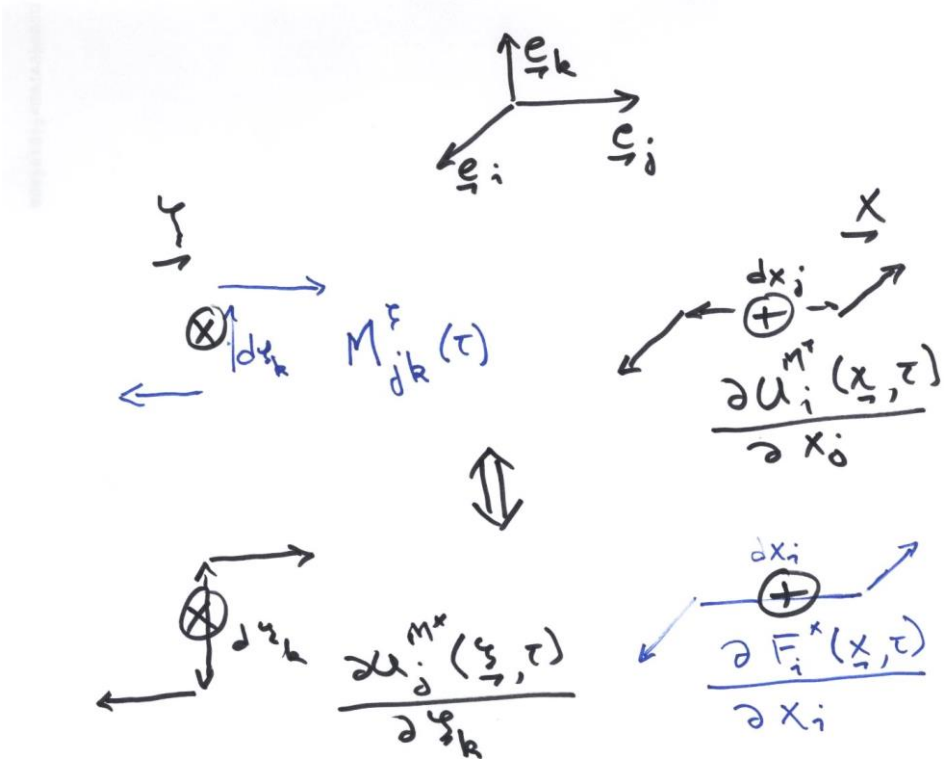


Figure 7.4. Reciprocity for a force couple at ξ involves observing a displacement gradient at x. The resulting time history of the displacement gradient is the same as the time history of a displacement gradient observed at ξ, but with the force couple at x, and with the orientation of the observation of the displacement gradient.

In this case the principle of virtual work is stated in terms of a torque times a rotation (similar to strain energies) at the source and receiver. That is,

$$\int_{-\infty}^{\infty} M_{jk}^{\xi}(t)\frac{\partial U_j^{M^x}(\boldsymbol{\xi},t)}{\partial \xi_k}dt = \int_{-\infty}^{\infty}\frac{\partial U_i^{M^\xi}(\mathbf{x},t)}{\partial x_l}M_{il}^{x}(t)dt \tag{7.35}$$

Where $M_{jk}^{\xi}(t) = M_{lm}^{x}(t)$, and as before this implies that

$$\frac{\partial U_j^{M^x}(\boldsymbol{\xi},t)}{\partial \xi_k} = \frac{\partial U_i^{M^\xi}(\mathbf{x},t)}{\partial x_l} \tag{7.36}$$

where

$$U_i^{M^\xi}(\mathbf{x},t) = \frac{\partial G_{ij}(\mathbf{x},t;\xi)}{\partial \xi_k} * M_{jk}^{\xi}(t) \tag{7.37}$$

And

$$U_j^{M^x}(\boldsymbol{\xi},t) = \frac{\partial G_{ji}(\boldsymbol{\xi},t;\mathbf{x})}{\partial x_l} * M_{il}^{x}(t) \tag{7.38}$$

Substituting (7.37) and (7.38) into (7.36) we obtain

$$\frac{\partial^2 G_{ji}(\boldsymbol{\xi},t;\mathbf{x})}{\partial \xi_k \partial x_l} * M_{il}^{x}(t) = M_{jk}^{\xi}(t) * \frac{\partial^2 G_{ij}(\mathbf{x},t;\xi)}{\partial x_l \partial \xi_k} \tag{7.39}$$

Since the moment time histories are all assume to be the same, reciprocity for force couples can be written as

$$\frac{\partial^2 G_{ji}(\boldsymbol{\xi},t;\mathbf{x})}{\partial \xi_k \partial x_l} = \frac{\partial^2 G_{ij}(\mathbf{x},t;\xi)}{\partial x_l \partial \xi_k} \tag{7.40}$$

This same result could have been obtained from the force statement of reciprocity. That is, if we differentiate (7.32) with respect to the kth direction at the point the force is applied, then

$$\frac{\partial G_{ij}(\mathbf{x},t;\xi)}{\partial \xi_k} = \frac{\partial G_{ji}(\xi,t;\mathbf{x})}{\partial x_k} \tag{7.41}$$

Similarly, if we differentiate (7.41) with respect to the lth direction at the point of observation, we obtain

$$\frac{\partial^2 G_{ij}(\mathbf{x},t;\xi)}{\partial x_l \partial \xi_k} = \frac{\partial^2 G_{ji}(\xi,t;\mathbf{x})}{\partial \xi_l \partial x_k} \tag{7.42}$$

Which is identical to the result obtained in (7.40). Fung shows that this result also applies to any linear visco-elastic material.

Point Double-Couple in an isotropic Whole Space

Consider a point double-couple source whose orientation is shown in Figure 7.5. Since we are using a point source in a homogeneous space, we expect to see radially spreading waves and we use spherical coordinates (r, ϕ, ζ) to describe the location of the observer. However, it is customary to describe the orientation of the fault with respect to a local coordinate system that is North, East, and up.

To describe the orientation of the rupture, we use Euler angles, and proceed in the following way. 1) Determine the **dip** angle δ of the fault. This is always between 0° (a flat plane) and 90° (a vertical plane). 2) Determine the **strike** angle θ of the fault. This is between 0° and 360° and it is the angle between geographic North and the surface projection of the fault plane. A 0° strike designates a fault that dips to the East, a 90° strike dips to the South, a 180° strike dips to the West, and a 270° strike dips to the North. 3) Determine the **rake** angle of the slip on the fault plane. A rake angle of 0° signifies a left-lateral slip (usually called **left-lateral strike-slip**), in which case the opposite side of the fault moves horizontally and to the left when looking across the fault. A rake of 90° signifies that the "**hanging wall**" moves vertically upward over the "**foot wall**," which is usually referred to as thrust type of motion and which produces crustal shortening. Hanging wall and foot wall are mining terms used for dipping seams of coal. When walking down the dipping surface of the mine, the hanging wall is the overhead surface (the ceiling), and the footwall is the floor. A rake of 180° signifies right-lateral strike-slip motion. A rake of 270° means the hanging wall moves downward with respect to the footwall. This is usually referred to as a **normal** fault and it results in stretching of the Earth's crust.

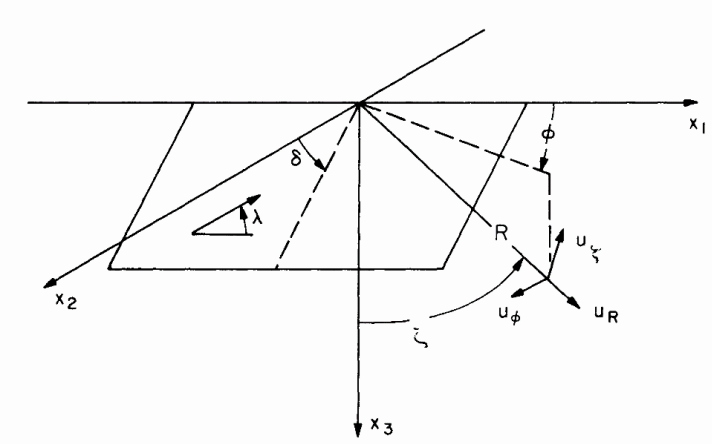


Figure 7.5. Shows the spherical coordinate system used together with the angles that describe the orientation of the fault and slip vector.

While the strike angle θ is defined with respect to geographic north, the observer's azimuth ϕ in Figure 7.2 is defined with respect to the strike line of the fault plane. We assume that the motion on a fault of area S (a very small area) is given by

$$D(t) = D\mathrm{H}(t) = \begin{Bmatrix} 0 & t<0 \\ D & t>0 \end{Bmatrix} \tag{7.43}$$

In this case the solution can be found by differentiating the point source solution (Equation 3.85) with respect to the appropriate source coordinate (this gives us a shear-couple of forces). This solution is then summed with the solution for a perpendicular force couple which gives a double-couple solution. The solution for displacement is

$$G_R = \frac{S\bar{D}\Re_R(\theta,\delta,\lambda,\phi,\zeta)}{4\pi}\begin{cases} 0 & ;t < r/\alpha \\ \dfrac{\beta^2}{\alpha^2}\left[\dfrac{\delta\left(t-r/\alpha\right)}{\alpha r} - \dfrac{1}{2r^2} + \dfrac{9\alpha^2 t^2}{2r^4}\right] & ;r/\alpha < t < r/\beta \\ \dfrac{\left(3\alpha^2-\beta^2\right)}{2\alpha^2 r^2} & ;t > r/\beta \end{cases} \tag{7.44}$$

and

$$G_\xi = \frac{S\bar{D}\Re_\xi(\theta,\delta,\lambda,\phi,\zeta)}{4\pi}\begin{cases} 0 & ;t < r/\alpha \\ \dfrac{\beta^2}{\alpha^2}\left[\dfrac{1}{r^2} + \dfrac{3\alpha^2 t^2}{r^4}\right] & ;r/\alpha < t < r/\beta \\ \dfrac{\delta\left(t-r/\beta\right)}{\beta r} + \dfrac{\beta^2}{\alpha^2 r^2} & ;t > r/\beta \end{cases} \tag{7.45}$$

Where ξ signifies either the ϕ or the ζ direction. The notation G_R and G_ξ are used to signify that this is the Green's function for a point dislocation in a whole space, and they signify displacements in the corresponding directions. Notice that everything is linearly scaled by $S\bar{D}$, which is the Potency. It is

common to write this solution using the Seismic Moment M_O , but these expressions always end up being written as M_O/μ.

I apologize for the complication that λ is used to signify both the first Lame constant and also the rake angle. The factors $\Re_R(\theta,\delta,\lambda,\phi,\zeta)$ and $\Re_\xi(\theta,\delta,\lambda,\phi,\zeta)$ signify the **Radiation pattern** and they are given by

$$\Re_R(\theta,\delta,\lambda,\phi,\zeta) = \cos\lambda\left(\sin\delta\sin^2\zeta\sin 2\phi - \cos\delta\sin 2\zeta\cos\phi\right) + \sin\lambda\left[\sin 2\delta\left(\cos^2\zeta - \sin^2\zeta\sin^2\phi\right) + \cos 2\delta\sin 2\zeta\sin\phi\right] \tag{7.46}$$

$$\Re_\zeta(\theta,\delta,\lambda,\phi,\zeta) = \cos\lambda\left(\tfrac{1}{2}\sin\delta\sin 2\zeta\sin 2\phi - \cos\delta\cos 2\zeta\cos\phi\right) - \sin\lambda\left[\tfrac{1}{2}\sin 2\delta\sin 2\zeta\left(1+\sin^2\phi\right) - \cos 2\delta\cos 2\zeta\sin\phi\right] \tag{7.47}$$

$$\Re_\phi(\theta,\delta,\lambda,\phi,\zeta) = \cos\lambda\left(\sin\delta\sin\zeta\cos 2\phi + \cos\delta\cos\zeta\sin\phi\right) + \sin\lambda\left(\cos 2\delta\cos\zeta\cos\zeta - \tfrac{1}{2}\sin 2\delta\sin\zeta\sin 2\phi\right) \tag{7.48}$$

Since the problem is linear, we can find the solution for any potency rate history by convolving solutions (7.44) and (7.45) with any potency rate function $\dot{P}(t) \equiv \partial(SD)/\partial t$.

While this solution may look complex, it is the simplest solution that approximates an earthquake and it is important to gain some physical intuition about the different parts of this solution. Let's start with the radial displacements given by equation (7.44). In Figure 7.6, I sketch the nature of the solution at several different distances. Notice that there is an impulse function in the radial component that arrives at the P-wave time, and its amplitude decays with distance as $1/r$. This is the **far-field P-wave** and its time history is the time derivative of the Potency on the fault. Its amplitude is also scaled by the inverse of the P-wave velocity α. That is, the far-field waves are the transient part of the solution, and

they carry the **radiated energy** in the wavefield. These radiated waves become smaller as the stiffness of the medium increases or as the density decreases. The amplitude of the far-field P-wave is different at different take-off angels and this amplitude is determined by $\Re_R$. Since the power flux through a surface depends on surface integral of $\dot{u}^2 \sim 1/r^2$, the *power flux of far-field waves through any sphere centered on the source is independent of the radius sphere*. Hence, in the absence of anelastic attenuation, the energy of the radiated wave is constant in time.

Now notice that after the arrival of the far-field P-wave, there is a term that is constant in time up until the S-wave arrives. This term is part of **the near-field P-wave** and its amplitude decays with distance as $1/r^2$. There is also a second near-field term that begins after the P-wave arrives. It grows in time as t^2 and decays with distance as $1/r^4$. While this tern may look like a very near-field term, notice that by the time of the S-wave, its amplitude has grown to order $\left(t_S - t_P\right)^2 = r^2\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)$, so it is actually of the same order as the other near-field terms. This very-near-field term (which behaves like the second time integral of the Potency time history) only exists up until the time of the S-wave arrival. After the S-wave arrival, the radial displacement is simply a constant in time; this is the static offset and it decays in distance as $1/r^2$.

The tangential components of the motion (equation (7.45)) are also of interest. Notice that there is a near-field P-wave static term that starts at the P-wave arrival time (yes there truly are P-waves on the tangential component). This is also accompanied by a very-near-field term that exists only between the P- and S-wave arrival times. The far-field S-wave then arrives and the solution then becomes the static displacement in the tangential directions. **Notice that the far-field S-wave is larger than the far-field P-wave by a ratio** of α^3/β^3 (this is a factor of 5.2 for a Poisson solid). Figure 7.6 shows the nature of the solution at different distances.

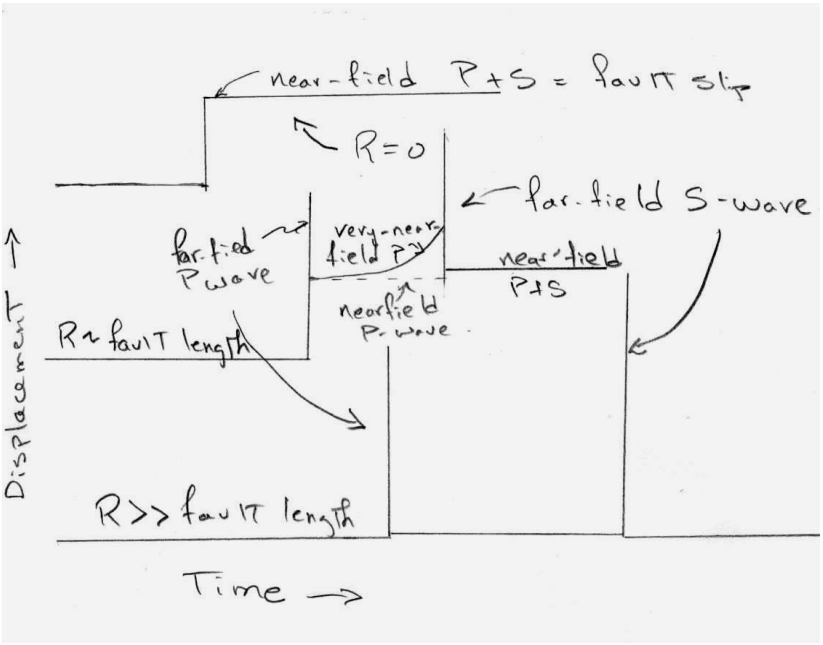


Figure 7.6 Cartoon showing the characteristics of the displacements generated by a point double couple source in a whole space. Far-field P- and S-waves have a time history that is the first time derivative of the potency time history and they decay as $1/r$. Near-field terms have the same time behavior as the potency history and they decay as $1/r^2$. Very-near-field terms only occur between the P- and S- arrival times, they have a time behavior that is the 2nd time integral of potency, and they decay as $1/r^4$.

Radiated energy from a point source

Normally, we consider the radiated energy to be the part of the energy that is associated with the waves long after the earthquake is over (assuming that there is no anelastic attenuation). In this case, we consider only the far-field terms and the radiated energy is the integral over time of power $p(r,t)$ through an enclosing surface at large distance *r*. That is,

$$E_R \sim r^2 \int_{starttime}^{endtime} p(r,t)dt \tag{7.49}$$

The power through any point on the surface is particle velocity times the stress associated with the traveling wave, or

$$p(r,t) = \dot{u}(r,t)\sigma(r,t) \tag{7.50}$$

In a traveling far-field S-wave (which contains most of the radiated energy), the stress scales as

$$\sigma \sim \rho\beta\dot{u} \sim \rho\beta\ddot{\mathrm{P}}(t)\Big/_{\beta r} = \rho\ddot{\mathrm{P}}(t)\Big/_{r} \tag{7.51}$$

Therefore, the power scales as

$$p(r,t) \sim \rho\beta\dot{u}^2 \sim \rho\beta\left(\ddot{\mathrm{P}}\ (t)\Big/_{\beta r}\right)^2 = \rho\ddot{\mathrm{P}}^2(t)\Big/_{\beta r^2} \tag{7.52}$$

Therefore, the radiated energy scales as

$$E_R \sim \rho \int_{starttime}^{endtime} \ddot{\mathrm{P}}^2(t)\Big/_{\beta} dt \tag{7.53}$$

This interesting result says that the radiated energy scales as the **2nd time derivative of the Potency history**. That is, *there is no radiated energy when the source is rupturing at a steady rate.*

Radiation Pattern

In addition to the time behavior and distance decay, there are a number of geometric factors $\left(\Re_r, \Re_\phi, \Re_\varsigma\right)$ given by equations (7.46) to (7.48) that determine the amplitude and direction of the particle motions of the solution. These factors are particularly important to engineering seismology, since they determine where the S-wave is going to be large. Unfortunately, these factors are complex enough that it is very hard to gain intuition simply by looking at equations (7.46) to (7.48). I like to use a simple tool to remind myself of the radiation pattern. This tool is called a **focal sphere** and an example of one is shown in Figure 7.7. I strongly suggest that you construct one for yourself (a softball and a magic marker work well), and that you learn how to use it.

Figure 7.7. A Focal Sphere. This is an idealization of the far-field waves from a point shear at the center of the sphere. The two planes that cut the sphere are the conjugate fault planes and the sense of slip on these two planes are shown by the arrows at the intersection of the planes.

The surface of the sphere represents an outgoing far-field wavefront (either a P- or an S-wave) from a point double couple. The sphere is cut by two perpendicular planes that represent the two conjugate slip planes. While real earthquakes occur on only one of these planes, when the source is approximated by a point double-couple, then the two conjugate planes produce identical deformations everywhere in the medium. Remember that the source is basically a point shear strain and that both stress and strain are symmetric tensors.

Notice that the conjugate slip planes divide the sphere into four quadrants, two of which are light colored and two are dark. This pattern is commonly referred to as a **beach ball** and seismologists commonly use it to show the orientation of the earthquake rupture (the strike, dip, and rake). The dark quadrants are used to show locations on the focal sphere where the P-wave is compressional (or radially outward from the source) and the light quadrants show where the P-wave is dilatational (or radially inward towards the source). While there are some exceptions, it is traditional to display the lower hemisphere of the focal sphere as it would be projected on a horizontal plane and viewed from directly above the source.

The far-field S-wave radiation pattern is denoted by the arrows on the focal sphere. That is, the arrows give the directions of the particle motion for the far-field S-waves. The points of intersection of the two planes turn out to be nodes for all motions (far-field P- S-, near-field motions, everything). These "super nodes" turn out to be 1) directly above a vertical strike slip rupture, 2) directly along strike for a dipping thrust or normal fault. At 90^o from these super-nodal points, and along both fault planes, the far-field S-wave attains its maximum value. Somewhat surprisingly, the direction of this maximum far-field S-wave is perpendicular to the fault plane. **I could draw many sketches trying to show what this radiation pattern looks like, but it is really better for you to make yourself a focal sphere and learn how to use it.**

Figure 7.8 shows some very interesting accelerograms recorded on SMA-1's for a M4.9 earthquake in the Imperial Valley, CA. The Imperial Valley is a deep basin filled with the sediments of the Colorado River delta. Seismic velocities in this basin increase continuously with depth; the overburden pressure compacts the sediments. This velocity gradient means that seismic rays are nearly vertically incident near the Earth's surface. Hence the P-waves are exclusively observed on the vertical component of the motion and the S-waves are observed on the horizontal components. Well, that is except for near-field P-waves that can be seen as the long-period ramp that precedes the S-wave arrival on several of the horizontal records. Notice that the far-field P-waves are much higher frequency than the S-waves. This is surprising since equations (7.44) and (7.45) indicate that the P- and S-wave Green's functions have the same behavior for the far-field waves; that is, they are both delta functions for a step function in potency. As it turns out, if the P-waves in these records are low-pass filtered then they do look similar to the S-waves (except that they are smaller, as they ought to be). So one explanation for this difference in frequency content of the P- and the S-waves is that both start the same at near the earthquake, but by the time they get to Earth's surface stronger anelastic attenuation of the S-wave causes the S-wave to be more attenuated than the P-wave. In Chapter 8, I suggest an alternative explanation; perhaps there are high-frequency dilatations happening on the rupture surface (acoustic fluidization).

While differences in anelastic wave attenuation may help to explain the different frequency contents of P- and S-waves, there is more to this problem. Figure 7.9 shows a polar plot of the amplitude of the far-field waves. That is the amplitude of the far-field waves are first corrected to be at the same distance, and then their amplitude is plotted as the radial distance from the origin for the appropriate direction to the observer. The expected amplitudes are the clover leaf pattern. What we see is that the amplitude of the low-pass filtered motion behaves as it should (according to our solution). However, the high-frequency waves do not have any apparent correlation with the expected radiation pattern. It is my experience that radiation pattern is clearly observable when the frequency is

approximately $1/(\text{duration of the earthquake})$ but that radiation pattern is generally not observable for frequencies higher than $10/(\text{duration of the earthquake})$.

While there are several speculations about why the radiation pattern "disappears" at high frequencies, I am not aware of a compelling study that explains this observation. Some have suggested that the breakdown in radiation pattern is due to short scale complexity in the source's geometry (a non-planar fault), others have suggested that there may be something happening other than pure shear sliding (dilatation?), and others have suggested that it has nothing to do with the source, but is caused by scattered waves. We will come back to these high-frequency waves in Chapter 8. At this point though, I think that it is sufficient to say that low-frequency waves can be explained by shear faulting on a plane in an elastic medium. High-frequency waves are usually best modeled as sources of random noise without any systematic azimuthal variation in amplitudes.

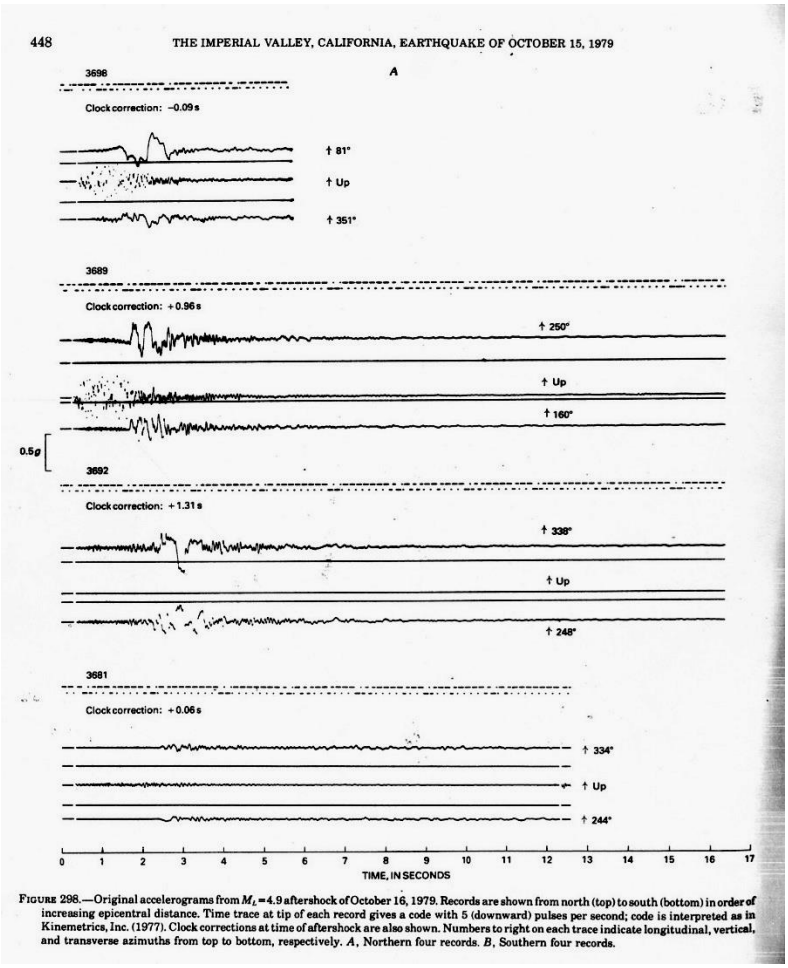


FIGURE 298.—Original accelerograms from $M_L = 4.9$ aftershock of October 16, 1979. Records are shown from north (top) to south (bottom) in order of increasing epicentral distance. Time trace at tip of each record gives a code with 5 (downward) pulses per second; code is interpreted as in Kinemetrics, Inc. (1977). Clock corrections at time of aftershock are also shown. Numbers to right on each trace indicate longitudinal, vertical, and transverse azimuths from top to bottom, respectively. *A*, Northern four records. *B*, Southern four records.

Figure 7.8. These are four near-source accelerograms recorded from a M 4.9 earthquake in the Imperial Valley, which is a large basin with deep horizontally layered sediments. Notice that the P-wave, which is observed on the vertical components, is much higher in frequency than the S-wave, which is seen on the horizontal components. Also notice the very-near-field waves seen on the horizontal accelerations before the S-wave arrives.

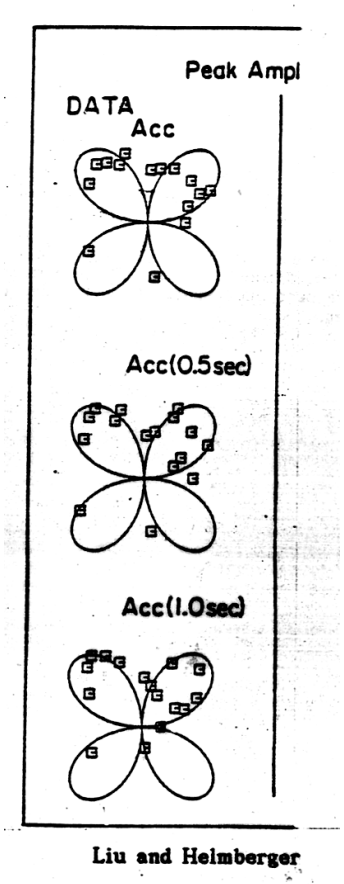


Figure 7.9. Polar plot of P-wave amplitudes (radial distance is amplitude at different azimuths) for a vertical strike slip earthquake in the Imperial Valley. The amplitudes are all corrected to the same distance and the "clover-leaf" pattern is the expected amplitude variation of far-field P-waves as a function of station azimuth. The raw acceleration records are dominated by high-frequency motions (> 5 Hz) and they show no apparent systematic variation with azimuth. However, when the records are low-pass filtered at 1 Hz, then there is a reasonable correspondence between the observed and predicted amplitudes.

Finite Sources

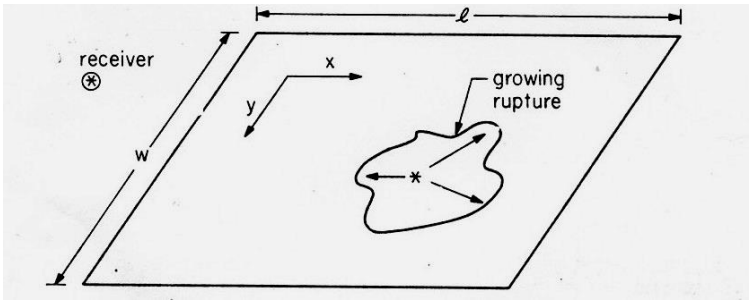


Figure 7.10. Simplified geometry of a finite fault model.

Of course, there are no true point sources in nature. Point sources can be useful approximations when the observer is at large distances compared with the source dimension and when the period of interest is long compared to wave travel times across the source region. However, it turns out that for many interesting problems, this approximation is not valid. While general seismic sources may involve inelastic deformations within a volume, it seems that many earthquakes can be simulated as slip on a plane. In this case we can write the solution at a point **X** (bold capital x is the vector location of the observer) by integrating the Green's function over the rupture surface. That is,

$$\mathbf{u}(\mathbf{X},t) = \int_0^L \int_0^W \dot{\mathbf{D}}(x,y,t) * \mathbf{G}(x,y;\mathbf{X};t)\, dydx \tag{7.54}$$

Where x and y are coordinates on the rupture surface, and $\mathbf{G}(x,y;\mathbf{X};t)$ is the vector Green's function for a point dislocation (a vector in the x,y plane) at (x,y) as observed at **X**. The * operator signifies a kind of generalized convolution in which each of the components of **D** and **G** are convolved. Since the source is finite, it is no longer convenient to choose spherical coordinates. In fact, since the source is located on a plane, it may be convenient to choose Cartesian coordinates. However, it is important to realize that the expression for the Green's function may become a very complex expression. Of course, it is a fairly straightforward procedure to solve (7.54) numerically once we know what **G** looks like, which is a difficult problem. However, if we are interested in the characteristics of the far-field waves, then we can approximate the Green's function as a sum of rays. Each ray represents either a P- or S-wave that travels to the receiver by a combination of refractions and reflections off of layer boundaries (see Chapter 4). That is, we can approximate the Green's function as

$$\mathbf{G}(x,y,t) = \sum_{i=1}^{\infty} \mathbf{G}_i(x,y,t) \tag{7.55}$$

Where we have dropped the coordinate of the fixed receiver, **X**, for convenience. For the case of a homogeneous space, there are only two rays to consider, direct P and direct S. In the case of a homogeneous half-space, there are 6 rays to consider P, pP, sP, S, pS, and sS. pP is a ray that travels upwards and then reflects at the free surface to become a down-going P-wave. The pS is a ray that travels upward from the source as a P-wave and is then converted into an SV-wave when it reflects off of the free surface. Likewise, sP leaves the source as an S-wave and then is converted into a P-wave when it reflects off the free surface. When the medium has more complexity than a half-space (e.g., a layer over a half space) then there are an infinite number of possible rays, although many of them are very small since they involve multiple reflections, each of which may involve a significant loss of energy.

If we are observing the source at a distance, then make the following approximation.

$$\mathbf{G}_i(x,y,t) \approx \mathbf{G}_i\left[x_0, y_0, t - T_i(x,y)\right] \tag{7.56}$$

Where (x_0, y_0) is some point on the fault plane (arbitrarily choose one), and $T_i(x,y)$ is the difference in the ith ray's arrival times for two point sources located at (x,y) and (x_0, y_0). We can rewrite (7.56) as

$$\mathbf{G}_i(x,y,t) \approx \mathbf{G}_i(x_0, y_0, t) * \delta\left[t - T_i(x,y)\right] \tag{7.57}$$

Where δ is a Dirac delta function. Substituting (7.56) into (7.55) into (7.54), we obtain

$$\begin{aligned}
\mathbf{u}(t) &\approx \int_0^L \int_0^W \sum_{i=1}^{\infty} \mathbf{G}_i(x_0, y_0, t) * \delta\left[t - T_i(x,y)\right] * \dot{\mathbf{D}}(x,y,t)\,dy\,dx \\
&= \sum_{i=1}^{\infty} \mathbf{G}_i(x_0, y_0, t) * \int_0^L \int_0^W \dot{\mathbf{D}}\left[x, y, t - T_i(x,y)\right] dy\,dx \\
&= \sum_{i=1}^{\infty} \mathbf{G}_i(x_0, y_0, t) * \mathbf{F}_i(t)
\end{aligned} \tag{7.58}$$

Where

$$\mathbf{F}_i(t) \equiv \int_0^L \int_0^W \dot{\mathbf{D}}\left[x, y, t - T_i(x,y)\right] dy\,dx \tag{7.59}$$

$\mathbf{F}_i(t)$ is called the **far-field time function** for the ith ray, and it is a function of time that is controlled by the spatial and temporal pattern of slip on the rupture

surface. If we ignore the travel time difference in (7.59), then we obtain the **moment rate function,** which is defined as

$$\dot{M}_0(t) \equiv \mu \int_0^L \int_0^W \dot{D}(x,y,t)\,dy\,dx \tag{7.60}$$

In (7.59) we kept track of the fact that slip is a vector quantity whose orientation (rake angle) can change with position on the fault. It is customary to define the moment rate function as a scalar quantity that depends on the integral of the amplitude of the slip vector over the rupture surface.

If the seismogram at a station is dominated by a particular ray (for instance the teleseismic P-wave from a deep earthquake), then the ground displacement at a station is largely determined by the far-field time function and anelastic attenuation that occurs during wave propagation. As an example, consider a simple Earth model in which P- and S-waves travel through a homogeneous half-space. Of course, the actual Earth is a spheroid with a complex structure. However, when P- or S-waves are observed at distances between 30° to 90° from the source, they appear to propagate without dispersion and the Green's functions for the P- and S-wave portions of the seismogram can be written as

$$\mathbf{G}_P(x,y) \approx \left\{ A_p \delta\left[t - T_P(x,y)\right] + A_{pP}\delta\left[t - T_{pP}(x,y)\right] + A_{sP}\delta\left[t - T_{sP}(x,y)\right] \right\} \mathbf{e}_P * Q_P(t) \tag{7.61}$$

$$\mathbf{G}_S(x,y) \approx \left\{ A_S \delta\left[t - T_S(x,y)\right] + A_{pS}\delta\left[t - T_{pS}(x,y)\right] + A_{sS}\delta\left[t - T_{sS}(x,y)\right] \right\} \mathbf{e}_S * Q_S(t) \tag{7.62}$$

Where the phases *P, pP, sP,* etc.are shown in Figure 7.9. $\mathbf{e}_p$ and $\mathbf{e}_S$ are unit vectors in the direction of the P- and S-wave particle motions at the receiver. A_P, A_{pP}, A_{sP}, etc. are amplitude factors that depend on the distance of the source (geometric spreading of the wavefront) and on the radiation pattern for the ray as it leaves the source [see equations (7.44) through (7.48)]. Finally $Q_P(t)$ and $Q_S(t)$ are Q operators for the P- and S-waves, respectively. That is, they are a kind of low-pass filter. Substituting (7.61) and (7.62) into (7.58), we find that the teleseismic P- and S-waves are approximately given by

$$\mathbf{u}_P(x,y) \approx \left\{ \begin{array}{l} A_p F_P\left[t - T_P(x_0,y_0)\right] + A_{pP} F_{pP}\left[t - T_{pP}(x_0,y_0)\right] \\ + A_{sP} F_{sP}\left[t - T_{sP}(x_0,y_0)\right] \end{array} \right\} \mathbf{e}_P * Q_P(t) \tag{7.63}$$

$$\mathbf{u}_S(x,y) \approx \left\{ \begin{matrix} A_S F_S\left[t - T_S(x_0,y_0)\right] + A_{pS} F_{pS}\left[t - T_{pS}(x_0,y_0)\right] \\ + A_{sS} F_{sS}\left[t - T_{sS}(x_0,y_0)\right] \end{matrix} \right\} \mathbf{e}_S * Q_S(t) \tag{7.64}$$

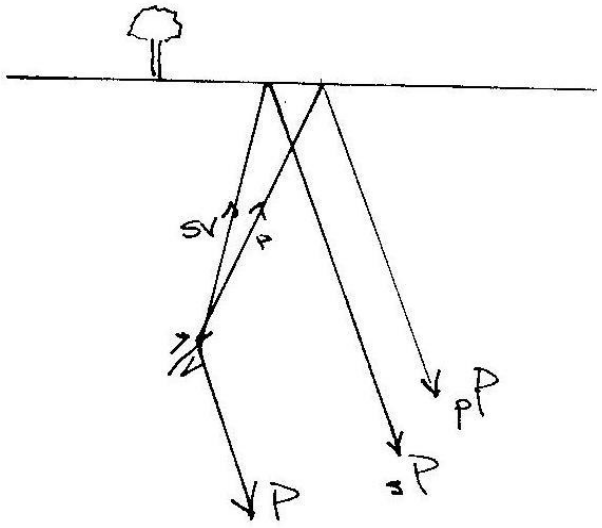


Figure 7.11 Idealization of the rays that comprise the P-wave group observed at teleseismic distances.

I show some examples in Figure 7.12 of how teleseismic P-waves are generated from (7.63). I assume a vertical strike-slip fault with a circular rupture whose diameter is 12 km and whose center is located at a depth of 7 km. The observer is located at a distance of 58° directly along the strike. The rupture velocity is assumed to be 2.8 km/sec, and the fault slip is assumed to be uniform everywhere with a time history that is an isosceles triangle of 1-second duration. In Figure 7.10, I show the far-field time functions and the final response of individual rays as well as the total final vertical long-period P-waveforms for models that differ only in the location of their hypocenters. It is assumed that the motions are recorded by a long-period WWSSN seismometer (15-s seismometer, 100-s galvanometer).

An inspection of the synthetic seismograms in Figure 7.12 yields insight into the role of rupture on a finite fault. The difference in the durations of the far-field time functions for the individual rays is due to a **Doppler effect that is called directivity** by seismologists.

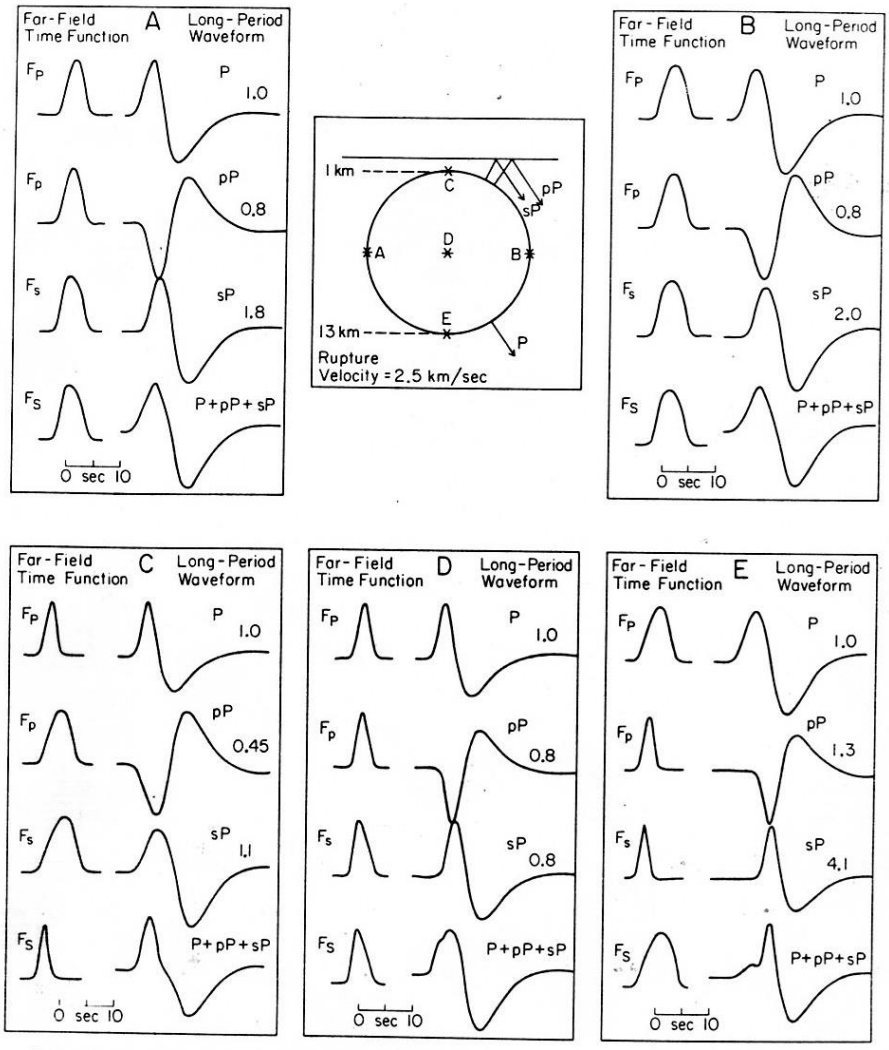


Figure 7.12. Illustration showing the effects of directivity on teleseismic P-waves from five earthquakes that are identical except for the location of their hypocenter. The ruptures are all on a circular, vertically-dipping strike-slip fault. F_P, F_p, F_s, and F_S, are the far-field time functions that are appropriate for the corresponding ray (see equation (7.59)). The seismograms on the right are those same time functions after convolving with a Q filter and an lon-period instrument response. The final seismograms are shown in the lower right. Notice that the effect of directivity is more important for hypocentral locations C (top of the fault) and E (bottom of the fault), since the P-waves leave the source region at near-vertical angles. The directivity effect is particularly evident when comparing F_S and F_s, since the rupture velocity is close to the shear wave velocity.

To provide insight into the solution to (7.54), I will begin with simple sources and then generalize to more complex sources. Let us first consider a **line source**. Of course, there are no true line sources in nature, but they demonstrate some important concepts in a simple way. Furthermore, they can be used to approximate ruptures whose along-strike length, *L*, is large compared to the down-dip width, *w*. For example, the 1906 San Francisco earthquake (M 7.8) has a rupture length of several hundred km and a width of less than 20 km.

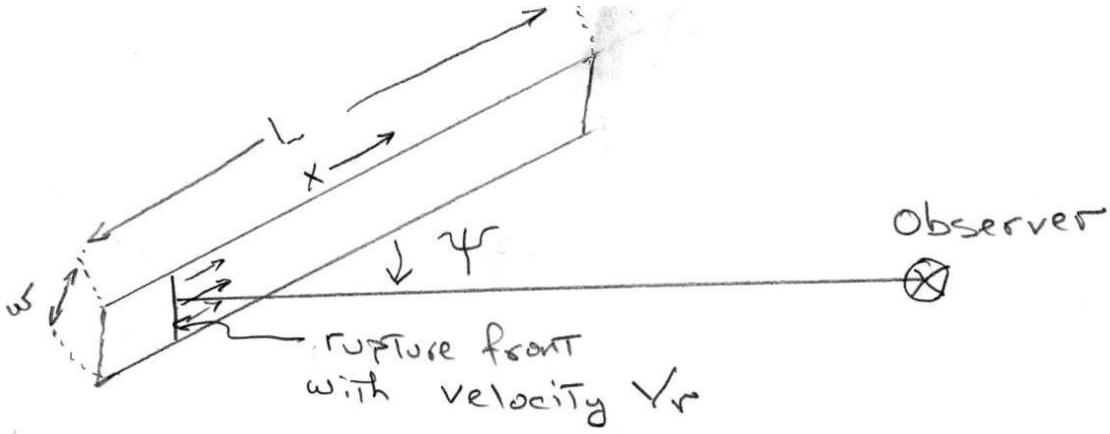


Figure 7.13. An idealized long and narrow fault.

To keep things simple, consider an observer in a homogeneous elastic whole space who is at a distance that is large compared to the fault length. Further consider the rupture geometry shown in Figure 7.13. The rupture is assume to be **unilateral,** which means that the rupture starts at one end and rupture to the other end. Some earthquakes rupture in both directions from the hypocenter and then they are called **bilateral**. Further assume that the slip at any point is given using a dimensionless along strike distance $\varsigma = x/L$.

$$D(\varsigma,t) = \bar{D}Lf(\varsigma)g\left(t - \frac{\varsigma L}{V_r}\right) \tag{7.65}$$

Where V_r is the rupture velocity, $\bar{D}$ is the average slip on the fault, $f(\varsigma)$ is a dimensionless function that describes the amplitude of the slip as a function of distance that is normalized by the average slip (i.e., $\int_0^1 f(\varsigma)d\varsigma = 1$,

$\int_0^1 D(\varsigma)d\varsigma = L\bar{D}$, and *g(t)* is the time behavior of the slip at each point normalized so that $\int_0^\infty g(t)dt = 1$. By writing slip on the fault in this way, we are assuming that the time dependences of the slip histories are the same for every point on the fault except for a delay due to the rupture velocity. While it is not really possible to decompose the slip history in this way, it is convenient for our purposes so that we can demonstrate several phenomena that appear in finite ruptures. Now let's assume that we are only interested in the far-field S-wave, since this is the wave that produces the largest particle velocities. Then for any point on our rupture, we can approximate the response from a Heaviside step in slip as (see equation (7.45))

$$\begin{aligned} G_\xi(x,t) &= \frac{w(dx)\Re_\xi(\theta,\delta,\lambda,\phi,\zeta)}{4\pi}\frac{\delta\left(t-r(x)/\beta\right)}{\beta r(x)} * \frac{\partial H(t)}{\partial t} \\ &= \frac{w(dx)\Re_\xi(\theta,\delta,\lambda,\phi,\zeta)}{4\pi}\frac{\delta\left(t-r(x)/\beta\right)}{\beta r(x)} \end{aligned} \tag{7.66}$$

where $G_\xi(x,t)$ has units of displacement per unit slip, $\delta\left(t-r(x)/\beta\right)$ has units of inverse time, and where

$$r(x) \approx r_0(1 - x\cos\psi) \tag{7.67}$$

(7.66) and (7.67) can be written in terms of dimensionless along strike distance ς and $dx = Ld\varsigma$ as

$$G_\xi(\varsigma,t) = \frac{w(Ld\varsigma)\Re_\xi(\theta,\delta,\lambda,\phi,\zeta)}{4\pi}\frac{\delta\left(t-r(\varsigma)/\beta\right)}{\beta r(\varsigma)} \tag{7.68}$$

and

$$r(\varsigma) \approx r_0(1 - \varsigma L\cos\psi) \tag{7.69}$$

Since we are assuming an observer at a large distance, we assume that $\psi \approx \psi_0$, and we assume that the effect of distance on the amplitude (the r in the

denominator) is negligible, and we can assume that the radiation pattern is approximately constant as a function of position along the rupture so that

$$G_\xi(x,t) \approx \frac{w(Ld\varsigma)\Re_\xi(\theta,\delta,\lambda,\phi_0,\zeta_0)}{4\pi} \frac{\delta\left(t - r_0(1-L\varsigma\cos\psi_0)/\beta\right)}{\beta r_0} \tag{7.70}$$

Now we can write the response of the finite fault as

$$\begin{aligned}
u(t) &= \int_0^1 \bar{D}f(\varsigma)\dot{g}\left(t-\frac{\varsigma L}{V_r}\right) * G_\xi(\varsigma,t)(Ld\varsigma) \\
&= \int_0^1 \left[\bar{D}f(\varsigma)\dot{g}\left(t-\varsigma L/V_r\right) * \frac{w\Re_\xi}{4\pi\beta r_0}\delta\left(t-r_0/\beta+\varsigma L\cos\psi_0/\beta\right)\right](Ld\varsigma) \\
&= \frac{wL\bar{D}\Re_\xi}{4\pi\beta r_0}\dot{g}\left(t-r_0/\beta\right) * \int_0^1 f(\varsigma)\delta\left(t+\varsigma L\cos\psi_0/\beta-\varsigma L/V_r\right)d\varsigma \\
&= \frac{wL\bar{D}\Re_\xi}{4\pi\beta r_0}\dot{g}\left(t-r_0/\beta\right) * \int_0^1 f(\varsigma)\delta(t-\varsigma T)\,d\varsigma \\
&= \frac{wL\bar{D}\Re_\xi}{4\pi\beta r_0}\dot{g}\left(t-r_0/\beta\right) * f\left(\frac{t}{T}\right)
\end{aligned} \tag{7.71}$$

where

$$T = L\left(\frac{1}{V_r} - \frac{\cos\psi_0}{\beta}\right) = \frac{L}{\beta}\left(\frac{\beta}{V_r} - \cos\psi_0\right) \tag{7.72}$$

T is actually the difference in shear-wave arrival times from a point at the hypocenter $(x=0)$ and a point at the end of the rupture $(x=L)$, including the time it takes the rupture to get from one end of the fault to the other.

Equation (7.71) tells us that if the rupture velocity is constant on a line source, then the ground displacement from the **far-field shear-wave is the same function of time** $f\left(t/T\right)$ **that the slip was as a function of distance along the fault** $f\left(x/L\right)$ (see equation (7.65)), but with an amplitude that is scaled by the $1/T$. In Figure 7.14, I show how the far-field S-wave displacement depends on receiver location, assuming that the slip at each point is instantaneous (i.e.,

assuming that $g(t) = H(t)$). Notice that the area under the pulse is identical for every receiver, except for the radiation pattern and the geometric spreading. This solution tells us that **spatial variations in slip map into temporal variations in the seismic waves**.

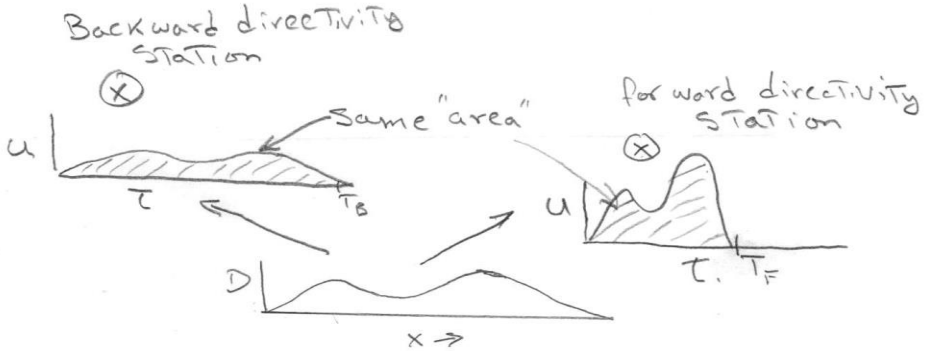


Figure 7.14. Schematic showing the relationship between the far-filed S-wave displacement time history and the spatial variation in slip along a line source in which the rupture proceeds from left to right at a constant velocity. The slip is considered to be a Heaviside function with time. Notice that the temporal variation of the far-field S-wave is identical to the spatial variation in the slip, except for a time contraction factor. This time contraction factor is specified by the time difference between the arrival of S-waves from each end of the fault.

If the slip at each point on the fault is very short (approximately a step function), and if you know the rupture velocity and fault geometry, the equation (7.71) tells us that we can derive the along-strike slip amplitude by a simple deconvolution process. Of course, we do not a priori know the rupture velocity and the duration of slip at each point. However, if we know the far-field S-wave at several stations then we can simultaneously determine $f\left(\frac{x}{L}\right), g(t),$ and V_r.

I won't go through the derivation here, but it can be shown that spatial variations in the rupture velocity also map into temporal variations in the far-field S-wave displacement. That is, complexity in the far-field S-wave can be caused by either spatial complexity in the amplitude of the slip, or in complexity in the rupture velocity.

At this point, it is useful to look at the characteristics of seismic waves that are produced if the slip on the fault is assumed to be spatially uniform, that is assume that

$$f\left(\frac{x}{L}\right)=\Pi\left(\frac{x}{L}\right)\equiv\begin{cases}0 & x<0\\ 1/L & 0<x<L\\ 0 & x>L\end{cases} \tag{7.73}$$

In this case equation (7.71) becomes

$$u(t)=\frac{wL\bar{D}\Re_{\xi}}{4\pi\beta r_0}\dot{g}\left(t-r_0/\beta\right)*\Pi\left(\frac{t}{T}\right) \tag{7.74}$$

We can now enquire about the energy flux of a radiated far-field shear wave through a patch of area $r_0^2 d\phi d\zeta$. We obtain this by integrating the energy flux per unit area $P=\rho\beta\dot{u}^2$ with respect to time (see equation 3.51). In this case

$$\begin{aligned}E&=r_0^2 d\phi d\zeta\int_0^{\infty}\rho\beta\left[\frac{wL\bar{D}\Re_{\xi}}{4\pi\beta r_0}\dot{g}\left(t-r_0/\beta\right)*\dot{\Pi}\left(\frac{t}{T}\right)\right]^2 dt\\ &=\rho d\phi d\zeta\frac{1}{\beta}\left(\frac{wL\bar{D}\Re_{\xi}}{4\pi}\right)^2\int_0^{\infty}\left[\dot{g}\left(t-r_0/\beta\right)*\frac{\delta(t)-\delta(t-T)}{T}\right]^2 dt\\ &=\rho d\phi d\zeta\frac{1}{\beta}\left(\frac{wL\bar{D}\Re_{\xi}}{4\pi}\right)^2\frac{1}{T^2}\int_0^{\infty}\left[\dot{g}(t)-\dot{g}(t-T)\right]^2 dt\end{aligned} \tag{7.75}$$

At this point we see that the density of radiated energy depends strongly on the duration of the far-field S-wave *T*. In order to evaluate (7.75), we will assume that the slip at every point is a ramp of duration τ_r, which is usually called either the "rise time" or the "slip duration." That is assume that

$$g(t)=\Pi\left(t/\tau_r\right)\equiv\begin{cases}0 & t<0\\ t/\tau_r & 0<t<\tau_r\\ 1 & t>\tau_r\end{cases} \tag{7.76}$$

In this case,

$$\dot{g}(t)-\dot{g}(t-T)=\frac{1}{\tau_r}\left[H(t)-H(t-\tau_r)\right]-\left[H(t-T)-H(t-T-\tau_r)\right] \tag{7.77}$$

There are two cases to consider; i) $\tau_r<T$, and ii) $\tau_r>T$. The first case, $\tau_r<T$, corresponds to a case in which the duration of slip is short compared to the rupture time. This type of rupture model is often referred to as a Haskell-like model (Haskell, 1964, BSSA, 1811-1841) and it has sometimes been criticized

since the slipping at any point on the fault arrest without any inherent knowledge of the ends of the rupture. However, more recent work suggest that rupture may propagate as a solitary wave of slip that is called a slip pulse (also sometimes referred to as a Heaton pulse, Heaton, T.H., 1990 Evidence for and implications of self-healing pulses of slip in earthquake rupture, *Phys. Earth Planet Int., Vol 64.* 1-20). The second case $\tau_r > T$ has rupture durations that are longer than the rupture time and they are referred to as "crack-like" models since the amplitude and duration of the slip are controlled by the dimension of the rupture. We will come back to further discuss these two models later in Chapter 8. Figure 7.15 shows a sketch of $\dot{g}(t)-\dot{g}(t-T)$ for these two cases.

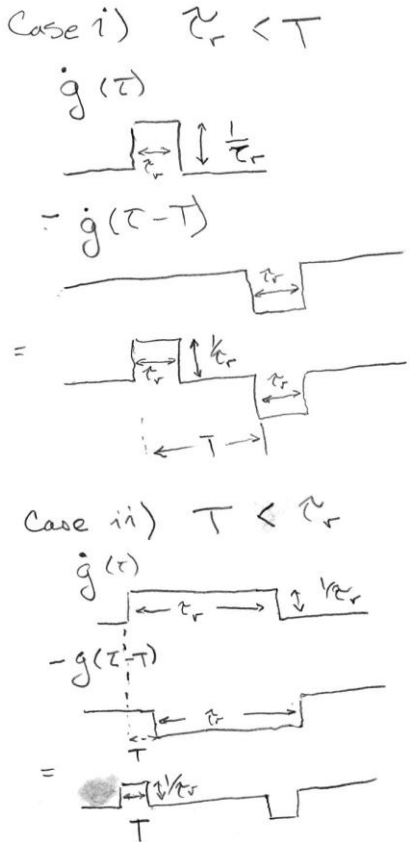


Figure 7.15 Schematic showing $\dot{g}(t)-\dot{g}(t-T)$, which is used for evaluating energy in equation (7.75) for a slip pulse model (case i) and for a crack-like model (case ii).

Let's first see what happens for the slip pulse model (case i).

$$
\begin{aligned}
E &= \rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\int_0^\infty \left[\dot{g}(t)-\dot{g}(t-T)\right]^2 dt \\
&= \rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\int_0^\infty \left[\Pi\left(t/\tau_r\right)-\Pi\left(t/\tau_r - T\right)\right]^2 dt \\
&= \rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\int_0^\infty \left[\Pi^2\left(t/\tau_r\right)+\Pi^2\left(t/\tau_r - T\right)\right] dt \\
&= \rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\int_0^\infty 2\Pi^2\left(t/\tau_r\right) dt \\
&= 2\rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\frac{1}{\tau_r}
\end{aligned}
\tag{7.78}
$$

We see that the radiated energy flux is very sensitive to T and τ_r. We can substitute (7.72) into (7.78) to obtain

$$
E = 2\rho d\phi d\zeta \left(\frac{w\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{\tau_r}\frac{\beta}{\left(\frac{\beta}{V_r}-\cos\psi_0\right)^2} \tag{7.79}
$$

Notice that the length of the fault drops out of this expression. This is because the only part of the radiated far-field S-wave is from the beginning of the rupture (a starting phase) and the end of the rupture (a stopping phase). As we have noted before, **there is no far-field radiation from a rupture that is propagating in a steady state.** Of course, this does not mean that there is no motion in the medium, it just means that the far-field terms (which comprise the radiated energy) are zero. The rest of the motion is comprised of the near-field terms.

We can now evaluate our crack-like model (case ii) in a similar way to obtain

$$E = \rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\int_0^\infty \left[\dot{g}(t)-\dot{g}(t-T)\right]^2 dt$$

$$= \rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\int_0^\infty \left[\Pi\left(t/T\right)T/\tau_r - \Pi\left(t/T-\tau_r\right)T/\tau_r\right]^2 dt$$

$$= \rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T^2}\int_0^\infty 2\left(T/\tau_r\right)^2 \Pi^2\left(t/\tau_r\right)dt$$

$$= 2\rho d\phi d\zeta \frac{1}{\beta}\left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{T}\frac{1}{\tau_r^2} \tag{7.80}$$

As before we can substitute (7.72) into (7.80) to obtain

$$= 2\rho d\phi d\zeta \left(\frac{wL\bar{D}\Re_\xi}{4\pi}\right)^2 \frac{1}{L\left(\frac{\beta}{V_r}-\cos\psi_0\right)}\frac{1}{\tau_r^2} \tag{7.81}$$

Notice that fault length is back in the expression, and that the radiated energy becomes very small as the rise time becomes large. In general, pulse-like ruptures radiate far more energy than do crack-like ruptures given the same fault dimensions and slip (that is the same potency).

Static Offsets from Finite Faults

The last section covered the problem of radiated waves, that is, they obey the homogeneous form of Navier's equation. However, there are other important terms to the solution for ground motions close to a finite rupture. Although it is common to use the word "near-field" for any site close to a rupture, I personally like to reserve the word "near field" to refer to the non-radiated parts of the solution. I suggest using the word "near-source" when the observer is close to a fault. Of course, the definition of close is subjective, but generally speaking it refers to sites that are less than a fault width away from the rupture (assuming that rupture width is the shorter of the two principal axes that span the rupture surface). As mentioned earlier, there are several types of near-field terms (see Figure 7.4), but the combination of terms that comprise the static solution are generally the most important.

Chinnery (The deformation of the ground around surface faults, Bull. Seism. Soc. Am., 1961, v. 51, 355-372) described the static solution for a rectangular strike-slip finite fault in a homogenous half-space with uniform slip. Figure 7.16 shows the geometry of the fault. Please note that the notation is different here; D is the

bottom depth of the fault and not the fault slip. Figure 7.17 schematically shows the displacements (in units of 10^{-3} times the fault slip) at the free surface assuming a fault that extends to the surface, i) and a bottom depth of 0.1 times L, and ii) a bottom depth that is 2.0 times L (a square fault). The motion right at the fault surface is symmetric about the fault plane and it is ½ of the fault slip, which would correspond to a contour of 500. This contour is not seen in the plots since this only occurs right at the fault. If d=0, L>>y1, y2, and D, then

$$u_1 \approx \frac{U}{\pi}\tan^{-1}\frac{D}{y_2} \qquad (7.82)$$
$$u_2 \approx u_3 \approx 0$$

By fitting (7.82) to geodetic data close to fault slips, we can approximately determine the bottom depth of slip. At distances that are large compared to the fault dimensions, the source can be approximated as a point-source, and the amplitude of the static displacements then decrease as R^{-2}. Coincidentally, since strain is proportional to spatial derivatives of displacement, the static change in strain (and also stress) decreases as R^{-3} when the distance is large compared to the rupture dimension.

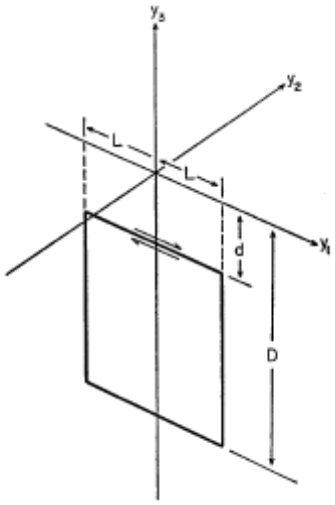


Figure 7.16. Chinnery's (1961) description of a vertical strike-slip fault embedded in an elastic half space.

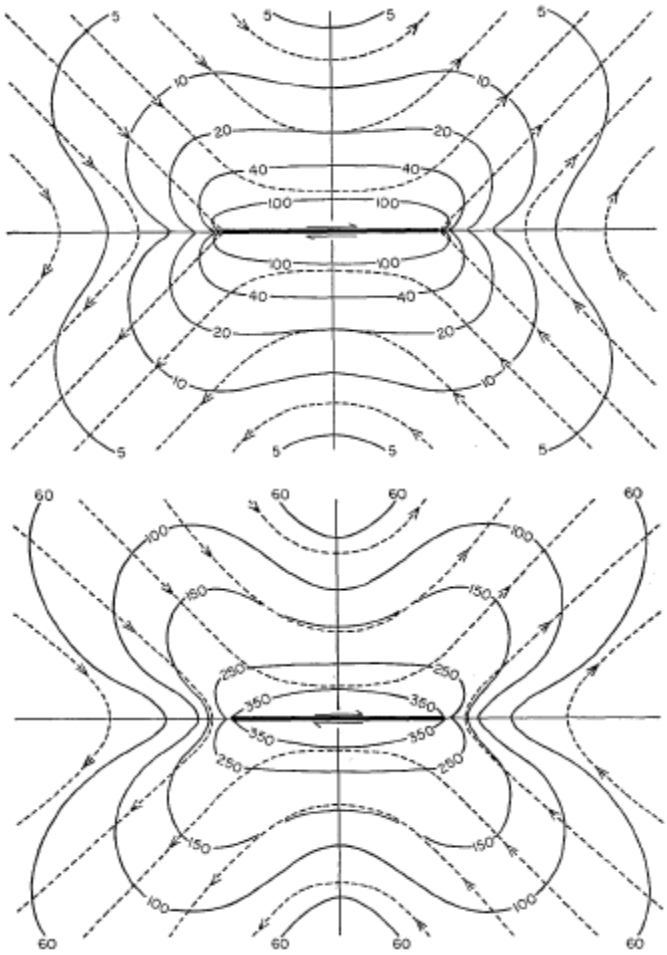


Figure 7.17. Map of the static displacement vectors from a vertical strike-slip fault with uniform slip that extends to the earth's surface and as observed at the free surface. The directions of the displacement vectors are parallel to the dotted lines and in the direction of the arrows. The contours give the amplitude of the vectors in units of 10^{-3} times the uniform slip assumed on the fault. The top panel shows a fault with a width that is 5% of its length and the bottom panel shows a fault with a width that is equal to its length.

Because of the symmetries in Chinnery's problem, the static deformation for a vertical strike-slip fault that extends to the surface of a half space is the same solution as for a vertical fault in a whole space, but with the width of the fault doubled. The problem becomes a little more complex for a dipping fault in a homogeneous half-space. Mansinha and Smylie (1971, The displacement fields of inclined faults, Bull. Seism. Soc. Am., 61, 1433-1440) presented the closed-form solution for a uniform slip on a dipping fault in a half-space. Figure 7.18 shows the geometry of the fault that they assumed. Figure 7.19 shows the displacements

at the free surface along a line that is perpendicular to the fault strike and that bisects

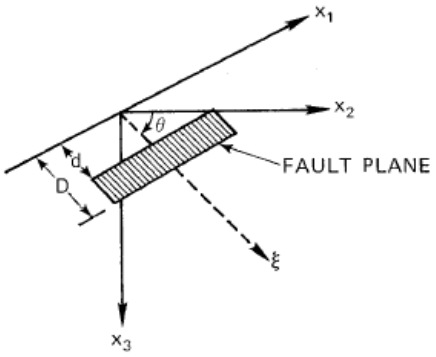


Figure 7.18. Mansiha and Smylie (1971) derived the static displacement field caused by a uniform slip on a dipping fault in an elastic half space.

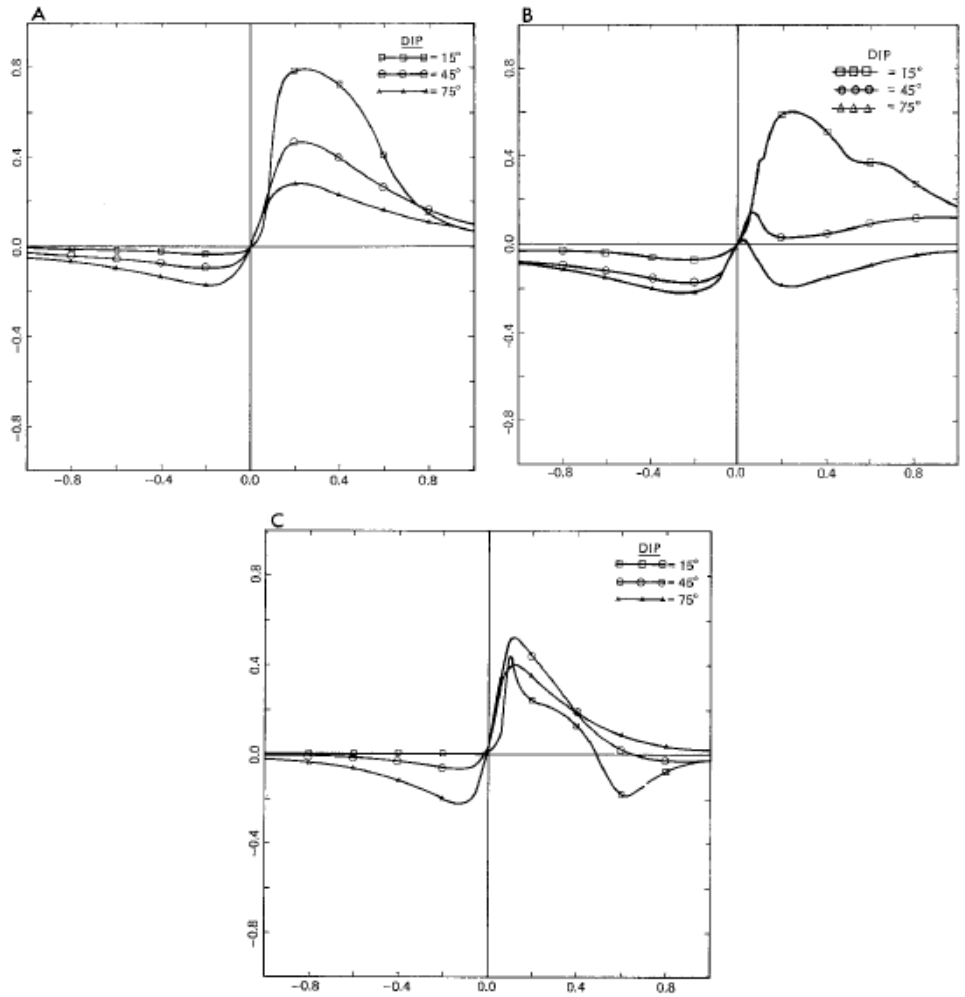


Figure 7.19. (from Mansinha and Smylie, 1971). Static ground displacements (one unit is fault offset) as a function of distance

along the axis x2 (line perpendicular to the fault strike) where one unit is L, and the fault length is 2L. Panel A is fault parallel motion from a dipping fault with strike slip motion. Panel B is fault-perpendicular horizontal motion for with dip-slip motion. Panel B is vertical motion for dip-slip motion.

the middle of the fault. Panel A is horizontal fault-strike-parallel static motion for a dipping fault with strike-slip motion (the other components of motion are identically zero along this line). Panel B is horizontal fault-strike-perpendicular motion for a dipping fault with dip-slip motion. Panel C is the vertical motion for a dipping fault with dip-slip motion.

-Chapter 8 Earthquake Scaling

Earthquake Intensity

Earthquakes occur in many different places in the Earth, some are shallow in the continental crust, others are along the interface between a subducting oceanic plate and a continental margin, and yet others are deep within the Earth (up to 650 km). Before the invention and deployment of seismographs, these earthquakes were largely described by their perceived maximum shaking intensity. There have been several shaking intensity scales developed through the years; the Rossi-Forel was defined in 1883 and it was the most common scale until the **Mercalli Intensity scale (MI)** was defined in 1902. This scale was modified in 1931 (intensities XI and XII were added to describe very damaging shaking and the definition of Modified **Mercalli Intensity scale (MMI)** is given in Table 8-1. Earthquakes are sometimes characterized by the maximum MMI reported. However, since large MMI values are based largely on observed damage to structures, maximum MMI is not a consistent way to characterize the size of earthquakes. In particular, the strongest shaking from an earthquake most commonly occurs in areas that do not have a sufficient density of buildings to characterize the intensity of the shaking. Furthermore, building construction changes dramatically with time and from region to region.

Modified Mercalli Intensity Scale

I. People do not feel any Earth movement.

II. A few people might notice movement if they are at rest and/or on the upper floors of tall buildings.

III. Many people indoors feel movement. Hanging objects swing back and forth. People outdoors might not realize that an earthquake is occurring.

IV. Most people indoors feel movement. Hanging objects swing. Dishes, windows, and doors rattle. The earthquake feels like a heavy truck hitting the walls. A few people outdoors may feel movement. Parked cars rock.

V. Almost everyone feels movement. Sleeping people are awakened. Doors swing open or close. Dishes are broken. Pictures on the wall move. Small objects move or are turned over. Trees might shake. Liquids might spill out of open containers.

VI. Everyone feels movement. People have trouble walking. Objects fall from shelves. Pictures fall off walls. Furniture moves. Plaster in walls might crack. Trees and bushes shake. Damage is slight in poorly built buildings. No structural damage.

VII. People have difficulty standing. Drivers feel their cars shaking. Some furniture breaks. Loose bricks fall from buildings. Damage is slight to moderate in well-built buildings; considerable in poorly built buildings.

VIII. Drivers have trouble steering. Houses that are not bolted down might shift on their foundations. Tall structures such as towers and chimneys might twist and fall. Well-built buildings suffer slight damage. Poorly built structures suffer severe damage. Tree branches break. Hillsides might crack if the ground is wet. Water levels in wells might change.

IX. Well-built buildings suffer considerable damage. Houses that are not bolted down move off their foundations. Some underground pipes are broken. The ground cracks. Reservoirs suffer serious damage.

X. Most buildings and their foundations are destroyed. Some bridges are destroyed. Dams are seriously damaged. Large landslides occur. Water is thrown on the banks of canals, rivers, lakes. The ground cracks in large areas. Railroad tracks are bent slightly.

XI. Most buildings collapse. Some bridges are destroyed. Large cracks appear in the ground. Underground pipelines are destroyed. Railroad tracks are badly bent.

XII. Almost everything is destroyed. Objects are thrown into the air. The ground moves in waves or ripples. Large amounts of rock may move.

Table 8.1. Definition of the Modified Mercalli Intensity Scale. Full descriptions are from: Richter, C.F., 1958. ***Elementary Seismology.*** W.H. Freeman and Company, San Francisco, pp. 135-149; 650-653.

U.S. earthquakes that occurred from the 1920's into the 1970's were assigned MMI's by the U.S. Coast and Geodetic Survey (merged with the USGS in 1978). The USCGS conducted routine surveys to prepare MMI intensity maps that were published regularly in a series called U.S. Earthquakes. The practice of compiling these data was largely

discontinued in the 1980's (too much manpower), and earthquakes were primarily described by other parameters (magnitude, pga, spectral acceleration). However, the widespread use of the Internet has allowed MMI compilations to be automatically managed through a USGS project known as "*did you feel it*?" https://earthquake.usgs.gov/data/dyfi/. In addition to shaking intensity observations, it has become common to characterize shaking intensity with observations of pga and pgv derived from strong-motion seismic data. In particular, relationships between reported MMI's and nearby recorded pga's and pgv's were used to derive quantitative relations that predict MMI from pga and pgv (Wald, D., V. Quitoriano, T. Heaton, and H. Kanamori, 1999, Relationships between peak ground acceleration, peak ground velocity, and Modified Mercalli Intensity in California, Earthquake Spectra, 15, 557-564 pdf). This is now referred to as Instrumental Modified Mercalli Intensity (IMM), which is given by

$$IMM = 3.66Log\left(PGA\right) - 1.66 \text{ for V} \le \text{IMM} \le \text{VIII}$$
$$IMM = 3.66Log\left(PGV\right) - 1.66 \text{ for V} \le \text{IMM} \le IX \qquad 8.1$$

$$IMM = 2.2\log(pga) + 1 \qquad IMM \le V \qquad 8.2$$

IMM uses the principle that smaller intensities are based on human perception of shaking. In these cases, humans seem to be better at discriminating pga than they are pgv. In contrast, higher levels of intensity are based on observations of structural damage; this seems to correlate best with pgv (see Chapter 6).

Japan has a long history of important earthquakes, and a Japanese intensity scale has evolved through time. The Japan Meteorological Agency (JMA) intensity scale (aka, Shindo scale) is summarized in Figure 8.1

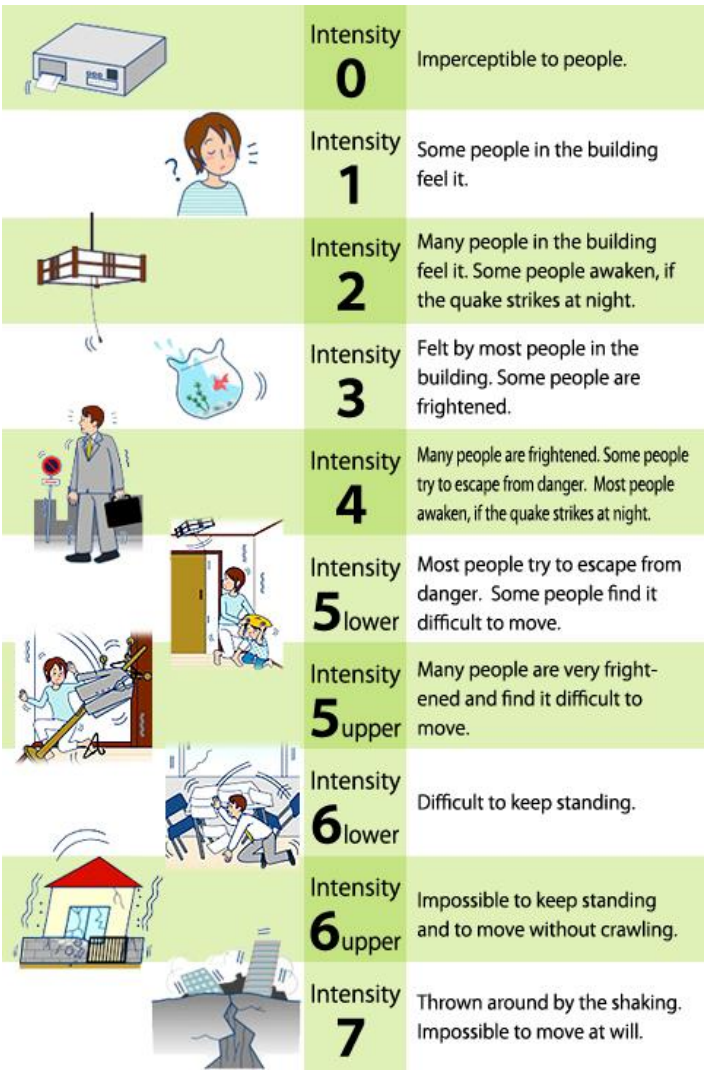


Figure 8.1. JMA intensity scale

Shaking intensity scales measure ground shaking at a point, BUT there must be structures to characterize the damage, and locally intense shaking may not reflect the overall size of an earthquake. One method to measure overall size using intensity data has been to estimate the area that experienced at least a given level of intensity. In particular, the area that experienced MMI ≥ IV has been helpful in quantifying the size of earthquakes for which only intensity data is available (pre 20th century). Intensity IV is convenient, since it is small enough to occur over a large area while still large enough to be reported in newspapers. (Estimating Earthquake Location and Magnitude from Seismic Intensity Data, 1997, Bakun W. and C. M. Wentworth, Bulletin of the Seismological Society of America, Vol. 87, No. 6, pp. 1502-1521).

Earthquake Magnitude

Charles Richter is usually identified as the designer of the first practical earthquake magnitude scale, which is now referred to as M_L (or local magnitude). When a seismologist tells you the "Richter magnitude" of an earthquake, he or she often intend to mean "local magnitude." Unfortunately, the non-seismologists seems to think that earthquakes are only assigned Richter magnitudes, which is definitely not true. Richter's magnitude scale was designed in the 1930's to assign a relative size to earthquakes that were recorded on the seismic network operated by Caltech's Seismological Laboratory in southern California. This network was comprised of horizontal-component, optical, torsion, seismometers designed by Harry Wood (Caltech seismologist) and John Anderson (an astronomer with the Hale Observatory). This instrument has a natural period of 0.8 s.; 80% damping, a displacement transducer (it recorded light deflected by a mirror onto photographic paper on a rotating drum) with a maximum gain of 2,800 (see chapter 2). The definition of M_L
is

$$M_L \equiv \log\left(\frac{A}{A_0}\right) = \log A - \log A_0 \qquad 8.3$$

Where A is the maximum amplitude (in mm) of a Wood-Anderson seismogram. The original definition only applies for a seismogram recorded at an epicentral distance of 100 km. Richter arbitrarily chose M = 3.0 to correspond to a peak seismogram amplitude of 1 mm. That is he defined that $A_0\left(\Delta = 100 \text{ km}\right) = 3.0$. Since earthquakes were recorded at a range of distances, Richter derived a distance correction factor, $A_0\left(\Delta\right)$, to correct the amplitudes to the standard distance of 100 km (see Fig. 8.2).

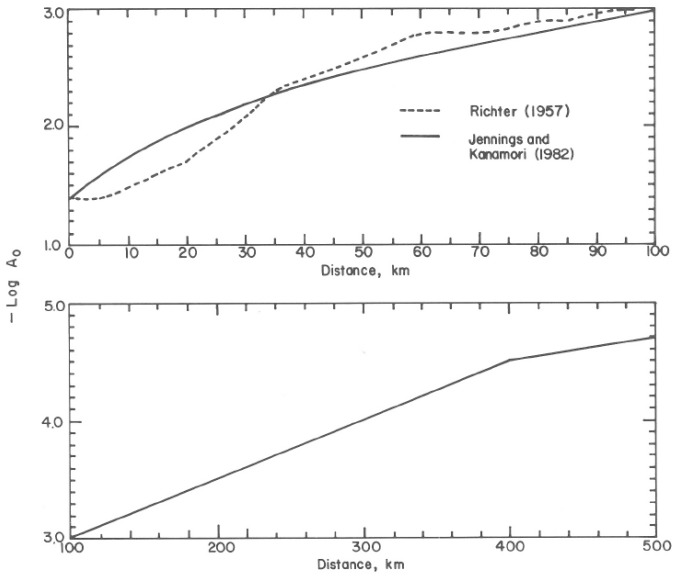


Fig. 8.2 Distance correction factor for Richter's local magnitude scale, M_L. The solid line shows Richter's definition, while the dotted line is from a later study by Kanamori. H. and Jennings, P.C., 1978. Determination of local magnitude, M, from strong-motion accelerograms. Bull. Seismol. Sot. Am., 68: 471-485.

Since a Wood-Anderson seismometer is a linear single-degree-of-freedom oscillator, the measured peak amplitude is identical to a 0.8 s, 80 % damped displacement-response spectrum. Since the damping is much larger than is typically used in structural engineering, the amplitude used in M_L correlates well with pgv, which also seems to be a good measure of shaking intensity; there is a detailed discussion of the relationship between pgv and 80% damped response spectra in Chapter 6. M_L was used from the 1920's through the 1980's as the standard magnitude to prepare catalogs of seismic activity in California (it has been adopted in other countries as well).

There were some important limitations to the use of M_L; in particular, earthquakes larger than M_L 5, tended to over-drive Wood Anderson seismometers. In response, low amplification versions of this seismometer were deployed. However, it was expensive to maintain a set of instruments that recorded only during the rare occurrence of $M_L > 5$. An even more serious limitation was that the local magnitude scale was only defined for California earthquakes that were recorded in California; it was not possible to compare the size of earthquakes worldwide.

In 1945 Beno Gutenberg (Caltech) introduced the Surface Wave Magnitude M_S which is defined as

$$M_S \equiv \text{Log}\, A_{R20} - A(\Delta) \qquad 8.4$$

A_{R20} is the maximum amplitude (in microns) of the 20-s Rayleigh wave (typically on the vertical component) recorded at distances exceeding 20° (about 2,200 km). Table 8.2 contains the values of A for a variety of distances

Δ (degrees)	−log A	Δ (degrees)	−log A
20	4.0	90	5.05
25	4.1	100	5.1
30	4.3	110	5.2
40	4.5	120	5.3
45	4.6	140	5.3
50	4.6	160	5.35
60	4.8	170	5.3
70	4.9	180	5.0
80	5.0		

Table 8.2. Distance correction factors for determination of M_S.

Rayleigh waves are often the largest wave group at these distances. Thus, A_{R20} is relatively easy to obtain using a photographic paper record, a meter scale and the response curve of standard seismographs (see Chapter 2). In principle, the measurement should be made after the record is narrow-band filtered at 20 s. Filtering these records would have been impractical in the 1940's; but fortunately 20 s Rayleigh waves are typically easy to identify on records (this arises from strong velocity dispersion in this period band). It's important to recognize that M_S is a measure of the Fourier amplitude spectrum at 20 s, while M_L is a maximum of the transient response of a 0.8-s sdof; that is these two scales measure different things. Nevertheless, Gutenberg and Richter adjusted A so that the two scales gave similar magnitudes for earthquakes where they could estimate both M_L and M_S. This meant that the scales converge at about $M_S \approx M_L \approx 6$. This can be seen in Figure 8.3, where M_L and M_S are plotted against each other. In particular, it appears that $M_L \sim \frac{2}{3} M_S$, or since these magnitude scales are logarithmic, $A_{Wood-Anderson} \sim A_{R20}^{2/3}$. That is, **M_L and M_S are fundamentally different measures of an earthquake**.

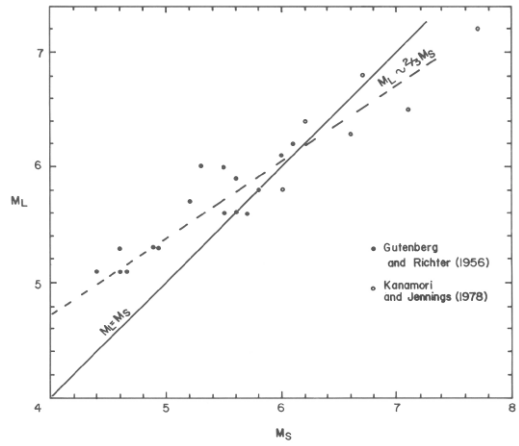


Fig. 8.3. Comparison of M_L and M_S for earthquakes reported for both scales. M_S is defined so that it gives comparable magnitude to M_L for earthquakes of about 6. From Heaton, T.H., Tajima. F., and Mori, A.W., 1986, Estimating ground motions using recorded accelerograms, Surveys in Geophysics, V. 8, pp 25-83.

Surface wave magnitude was a successful and commonly used scale for a large catalog of global earthquakes recorded in the 20th century (many older events were assigned magnitudes based on historic recordings. However, since the excitation of surface waves decreases exponentially with the depth of an event, M_S is not appropriate to characterize the size of earthquakes deeper than 20 km. To deal with this problem, Gutenberg, Richter, and Hugo Benioff (yet another famous Caltech seismologist) devised the **body wave magnitude scale, m_B**. This scale is based on teleseismic body waves (especially the P-wave between 30° and 90°). In this case, the maximum amplitude of the P-wave A_P and the duration of the pulse with the maximum amplitude T_P are measured.

$$m_B \equiv \log\left(\frac{A_P}{T_P A(\Delta, z)}\right) \qquad 8.5$$

Where $A(\Delta, z)$ is a complex correction that depends on distance and source depth; it was calibrated to give comparable magnitudes to M_S for magnitudes near 7. Notice that the amplitude is normalized by the pulse duration. Richter says that this is a measurement of particle velocity which is used in the computation of kinetic energy.

The short-period body wave magnitude m_b is also commonly encountered in catalogs of earthquakes. This magnitude is also logarithmic and it depends on the maximum amplitude of teleseismic P-wave amplitudes as measured in the first 3 seconds of a short-period standard vertical seismometer (1-s natural period). m_b was designed to help discriminate the seismic signals from underground nuclear explosions from seismograms of natural earthquakes. That is, nuclear explosions happen very quickly, and they have very impulsive P-waves, whereas earthquakes are typically more emergent; they take some

time to get going; that is events with high m_b / M_S were marked as possible underground nuclear tests.

The development of magnitude scales has a long and complex history. There have been many other magnitude scales that have been devised. Unfortunately, they all suffer from the fact that they are unit less; they measure relative size, but they are not related to physical parameters of the earthquake. Hiroo Kanamori (yet another Caltech Seismologist) defined **energy magnitude $\mathbf{M_W}$** to be

$$M_W \equiv \frac{\mathrm{Log} W_0 - 4.8}{1.5} \qquad 8.6$$

Where W_0 is the total radiated energy in joules (the W is for work). Kanamori defined M_W in this way so that it would be compatible with the energy-magnitude relation developed by Gutenberg and Richter (1956, BSSA, 32, 163-191). Unfortunately, estimation of radiated energy is a difficult problem, which makes the use of M_W cumbersome in practice. Kanamori noted that on average

$$W_0 \approx \frac{M_0}{2\times 10^4} = \frac{\mu \bar{D} S}{2\times 10^4} = \sigma_{eff} \bar{D} S = \sigma_{eff} P \qquad 8.7$$

Where μ is the average crustal rigidity, about 40 GPa, which means that $\sigma_{eff} \approx 2MPa$. σ_{eff} is the **effective stress** that converts slip per unit rupture area into radiated energy per unit rupture area. By substituting 8.7 into 8.6, Kanamori was able to approximate $\mathbf{M_W}$ as

$$M_W \approx \frac{\log M_0 - 9.1}{1.5} \qquad 8.8$$

Coincidentally, Thatcher and Hanks (1973, JGR, 78, 8547-8576) derived a general relation between M_0 and M_L for southern California earthquakes

$$M_L \approx \frac{\log M_0 - 9}{1.5} \qquad 8.9$$

Hanks noted that the M_L relation (eq.8.9) is very similar to Kanamori's approximate relation for M_W (8.8), and he convinced Kanamori to collaborate on a new definition of magnitude they called **moment magnitude $\boldsymbol{\mathcal{M}}$** defined as

$$\boldsymbol{\mathcal{M}} \equiv \frac{\log M_0 - 9.05}{1.5} \qquad 8.10$$

Where the 9.05 was obviously a political compromise (Hanks and Kanamori, 1979, JGR, 84, 2348-2350). The choice of the distinctive bold script M meant that it was awkward to find the appropriate symbol on word processors and hence many researchers began to call M_W moment magnitude. Unfortunately, some have mistakenly confused energy and moment (see Chapter 7). On a personal note, it drives me crazy that one of the most

referenced papers in geophysics (Hanks and Kanamori, 1979) is just a simple 2-page paper to define moment magnitude, but most researchers who cite it use it incorrectly; M_W is defined to be Energy Magnitude. M_W **is NOT Moment Magnitude**.

The easy form for moment magnitude (technically it's Potency Magnitude) is that given in 8.8

$$\begin{aligned} \mathcal{M} &\approx 1+\frac{2}{3}\log S\bar{D} \\ &= 1+\frac{2}{3}\log P \end{aligned} \qquad 8.11$$

Where the units are in meters. I have assumed that the average crustal rigidity is 35 GPa in deriving 8.11. This can cause some confusion, since some researchers use 40 GPa and others use 35 GPa. However, since the rigidity is not really a source parameter (see Chapter 7), this is not a significant issue. Since it makes the definition of magnitude simple, I suggest using 35 GPa.

As a convenient rough rule of thumb, a 30 km by 30 km rupture with a 1-m slip gives a potency of about 1 cubic km, which corresponds to a moment magnitude of about 7. From now on, I will refer to Potency Magnitude as simply magnitude. While it is, for all intents and purposes, the same as moment magnitude, potency is a much simpler concept and its use simplifies the math.

Figure 8.4 is a comparison of a variety of different magnitude scales (Heaton, Tajima, Mori, 1986, Surveys in Geophysics, 8, 25-83). It shows the average relationship between these scales and moment magnitude. This plot clearly shows that most magnitude **scales saturate** with increasing total energy (or potency). The scales based on shorter-period measurements saturate at smaller magnitudes and the scales based on long-period measurements saturate at larger magnitudes.
Unfortunately, magnitude saturation was poorly understood prior to the 1980's and there are important inconsistencies in reporting the sizes of older large earthquakes. For example, the 1906 San Francisco earthquake was assigned a surface wave magnitude, M_S, of 8.3 based on teleseismic recordings. In comparison, the 1964 Alaskan earthquake was assigned an M_S of 8 ¼ . However, modeling of geodetic data from these earthquakes has indicated that the 1906 earthquake was $\mathcal{M}$ = 7 ¾, while 1964 was $\mathcal{M}$ = 9 ¼ ; that is the Alaskan earthquake was 1.5 units greater, or $10^{3/2 1.5} = 180$ times larger. The use of the moment magnitude scale revolutionized the understanding of the sizes of the greatest earthquakes. You can find a table listing the largest historic earthquakes sorted according to their moment magnitudes at https://en.wikipedia.org/wiki/Lists_of_earthquakes.

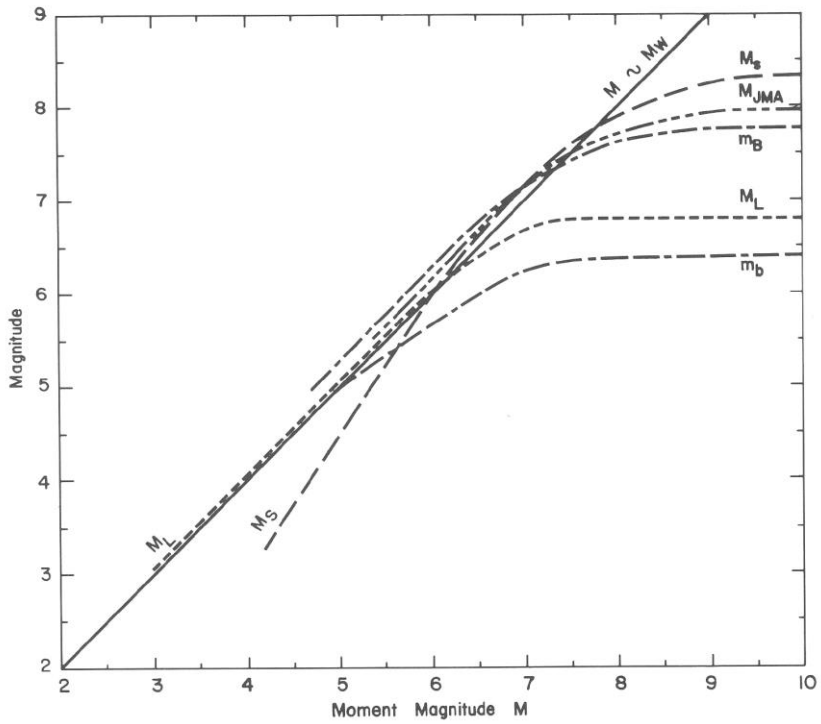


Figure 8.4. Comparison of a variety of magnitude scales (Heaton, Tajima, Mori, 1986, Surveys in Geophysics, 8, 25-83).

At this point, you should be asking yourself about this factor of $^2/_3$ that shows up in definition of moment magnitude; is there some physics behind this factor? One way to think of this problem is to consider a far-field P- or S-wave pulse of duration T and amplitude A. The potency (or moment) of the earthquake is then related to the area of the pulse, i.e., $P \sim AT$ (see Chapter 7). This implies that $A \sim {}^{P}\!/_{T}$. The assumption that most magnitude scales are logarithmic in amplitude, A, means that the definition of potency magnitude is consistent with other scales if $P \sim T^3$, in which case $A \sim P^{2/3}$, which implies that

$$\mathcal{M} \sim \log A \sim {}^{2}\!/_{3} \log P \qquad 8.12$$

As we will see in the following discussion, there are other ways to derive the $^2/_3$. Furthermore, this factor was in the relationship between magnitude and radiated energy that was derived by Gutenberg.

Gutenberg-Richter Relationship

Since the earliest days of compiling catalogs of earthquake magnitude, it has been noted that the number of earthquakes decreases with increasing magnitude and can be described as a power law known as the Gutenberg-Richter frequency-magnitude relationship (often shortened to just "Gutenberg-Richter").

$$\log(N) = A - bM \quad \text{or}$$
$$N = 10^{A}10^{-bM} \qquad 8.13$$

This can be rewritten in natural logs as

$$\ln(N) = \log e \log(N) = A\log e - (b\log e)M$$
$$\text{or } N = \exp(A\log e)\exp\left[(b\log e)M\right] \qquad 8.14$$

Where N is the number of earthquakes larger than M, and A and b are constants that are typically called "the A value" and the "b value". This equation is the first form of Gutenberg-Richter and it is called the "cumulative form of Gutenberg-Richter" . A second form of this relationship describes the statistics of the number of events $N'\Delta M$ with magnitudes between M and $M + \Delta M$. In this case, $N' = \frac{dN}{dM}$. Now differentiating 8.13 with respect to M , $\frac{d(\log N)}{dM} = -b$. However, we also know that $\frac{d(\log N)}{dM} = \frac{1}{N}\frac{dN}{dM} = \frac{N'}{N}$, so $N' = \frac{-bN}{\log e}$ from which it follows that $\log N' = \log N - \log b - 0.362$ which provides the second form of the Gutenberg-Richter relation (called the interval form) that is written

$$\log(N') = a - bM \qquad 8.15$$

where $a = A + 0.362 + \log b + \log \Delta M$. Notice that the two forms, 8.13 and 8.15, only differ by a constant. If a large data set exactly fits the G-R relation, then the a- and b-values can be estimated using a visual inspection of a log-linear plot ($\log N$ vs M). However, it is more common to encounter catalogs that do not precisely fit this law, in which case it is easier to estimate the constant when plotting the first form. Most commonly, people use the first form. However, if you wish to visually judge whether a data set fits G-R, it is probably best to use the second form; it reveals the details better.

Interpreting the meaning of the G-R relation has been the subject of a huge number of papers, but to this day, there still seems to be a wide variety of speculative explanations.

The a-value is a constant that gives the overall rate of activity. The b-value describes the relative frequency of large and small events. Later in this chapter, I argue that the b-value arises from the chaotic dynamics that are caused by the low dissipation energy in earthquakes.

It is important to understand that, given a suite of earthquakes, the estimates of the a- and b- values is dependent on the magnitude scale that was used to parameterize the events. For example, I argued that $M_L \sim \frac{2}{3} M_S$. This means that using M_L to determine b-value will give a different estimate than if the catalog uses M_S . Fortunately, most networks that prepare earthquake catalogs attempt to enforce the use of moment magnitude. Unfortunately, many catalogs of historic earthquakes contain a variety of different magnitude estimates.

The G-R relation is also referred to as a *Pareto Distribution*. This name is common in economics, where it has long been observed that personal wealth is described by a Pareto distribution; a few people have most of the overall wealth in society. Pareto distributions often arise in self-organizing systems. I have found a very nice overview lesson on power laws by Mark Newman (U. of Mich.) which can be found at http://www-personal.umich.edu/~mejn/courses/2006/cmplxsys899/powerlaws.pdf

We can use the Gutenberg Richter relation to investigate the following common question. Which magnitude earthquakes are responsible for most of the motion of the Earth's tectonic plates? Is it the infrequent large events, or the far more numerous smaller earthquakes? To estimate the total inelastic strain integrated over the volume (i.e., total Potency or P_{total}) , resulting from suite of earthquakes between magnitudes $M_{\min}$ and $M_{\max}$, we perform the following integral

$$P_{total} = \int_{M_{\min}}^{M_{\max}} N' P(M) dM \qquad 8.16$$

Where $P(M) = 10^{\frac{3}{2}(M-1)}$ (see 8.8) and N' is given by 8.15. Performing the integral, we obtain

$$\begin{aligned} P_{total} &= 10^a \int_{M_{\min}}^{M_{\max}} 10^{(1.5-b)(M-1)} dM \\ &= \frac{10^a}{(1.5-b)\ln 10} \left[10^{(1.5-b)M} \right]_{M_{\min}}^{M_{\max}} \end{aligned} \qquad 8.17$$

If $M_{\min} = -\infty$, then

$$P_{total} = \frac{10^{a+(1.5-b)M_{\max}}}{(1.5-b)\ln 10} \sim C10^{(1.5-b)M_{\max}} \qquad 8.18$$

Analysis of the global catalog of earthquakes provides estimates that $b \approx 1$, so $P_{total} \sim C10^{\frac{1}{2}M_{\max}}$. This means that the total slip increases by a factor of 3.16 for every 1-unit increase in $M_{\max}$; clearly **the large earthquakes account for the majority of the deformation. This fact means that it is difficult to use historic seismicity catalogs to estimate tectonic deformation rates. The largest events dominate the total deformation estimate, but they are so infrequent that there are too few events to do statistics. Furthermore, Gutenberg Richter converges to a finite tectonic rate only in the event that the power law is truncated at the largest events. If there is no largest size event, then 8.18 is divergent. Although attempts have been made to estimate the maximum event, there is currently no accepted methodology to truncate the Gutenberg-Richter Relation.** As you read further in this chapter, I think that you will see that estimation of a largest event is not a solvable problem.

Notice that if $b > 1.5$, then the situation reverses and more of the deformation is caused by the small events. Taken literally, $b > 1.5$ means that there is so much deformation from the smallest events that there does not need to be any larger events, that is, the plate boundary is creeping.

One notable aspect of the G-R relation is the way it scales rupture area. In particular, it has been observed that seismic moment scales with rupture area as

$$M_0 \approx 1.2x10^{15} S^{\frac{3}{2}} \qquad 8.19$$

Where S is rupture area in km^2 , and moment is in N-m see Figure 8.5 from Kanamori (1980). Writing this in terms of potency,

$$P \approx 3x10^{4} S^{\frac{3}{2}} \qquad 8.20$$

Where *S* is still in km^2. Converting S to m^2, we obtain.

$$P \approx 3x10^{-5} S^{\frac{3}{2}} \qquad 8.21$$

This allows us to infer the total rupture area of an ensemble of earthquakes as a function of M. Combining 8.21 and 8.11, we obtain

$$\begin{aligned} M &\approx 1 + \tfrac{2}{3}\log P \approx 1 + \tfrac{2}{3}\log\left(3\times10^{-5}\right) S^{\frac{3}{2}} \\ &= \log S - 2.02 \end{aligned} \qquad 8.22$$

Or

$$S \approx 10^{M} 10^{2.} \qquad 8.23$$

This means that the sum of the rupture areas of earthquakes with magnitudes between M and M+ΔM is

$$S' = N' \Delta M S(M) = 10^{a-bM} 10^{M+2.} = 10^{a+2.} 10^{(1-b)M} \qquad 8.24$$

This means that S' is a constant if $b = 1$. That is, **the total rupture area is the same for all earthquakes in a magnitude interval** ΔM . That is, if the b-value is 1, then the total rupture area is the same for integrated area of each magnitude. That is, the sum of the rupture area of all 2's is the same as the rupture area of all 3's, is the same as all 4's, etc. What this means is that given a b-value of 1, and given that a point has just experienced slip, then it is equally likely that it could have come from any magnitude earthquake. Given that different magnitude earthquakes have different slips, any slip is as likely as any other. This is only true in a logarithmic sense. That is, a fault is equally likely to experience slip between D and C x D, regardless of the value of D and a constant, C (e.g. given, slips between 1 and 2 mm are just as common as slips between 4 and 8 m.

For example, if you sum the rupture areas of 10 M 6 earthquakes, they would equal the rupture area of a M 7. This allows us to answer the following question. "Given that a point on a fault experiences a rupture, what is the most likely magnitude of earthquake that caused the rupture?" Apparently, if $b = 1$, then all magnitude earthquakes are equally likely to have been the cause of the rupture. If $b < 1.0$, then the given rupture is most likely caused by a large event, whereas if $b > 1.0$, then it's most likely that the event was caused by a small event.

From the previous discussion, we can conclude that if there is a point on a fault that fails in an earthquake, then it is equally likely that the causative event was any magnitude. Later I show that the average slip in an event increases as $10^{\frac{1}{2}M}$ (see 8.74), which means that, "**Given that a point on a fault experiences a rupture, then all values of slip at that point are equally likely**."

Stress Drop and Strain Change

The term, stress drop, is often used in seismology and unfortunately, it has been defined in many ways (often confusing). In the following discussion, **I will define stress drop to be the change in shear traction on a fault that is the result of shear displacements that occur in an earthquake**. That is,

$$\Delta\sigma(x,y) \equiv \sigma_0(x,y) - \sigma_1(x,y) \qquad 8.25$$

Where σ_0 is the initial stress (often called the prestress) and σ_1 is the final stress. This definition of stress drop is often referred to as the **static stress drop**. As far as I am concerned, this is the only unambiguous definition for stress drop, although it seems that many (perhaps most) of my colleagues use other definitions.

The easiest way to think of static stress drop at a point on a fault is that it is the change in elastic strain times the rigidity, or $\Delta\sigma \equiv \sigma_0 - \sigma_1 = \mu(\Delta\varepsilon)$, where $\Delta\varepsilon$ is the amplitude of the change in shear strain in the volume immediately adjacent to the rupture surface. Unfortunately, things are not quite so simple since $\Delta\sigma$ and $\Delta\varepsilon$ are actually tensors that vary over space such that they satisfy the equations of static equilibrium ($\sigma_{ij,j} = 0$). However, in these class notes, I am not very interested in the details and I will use appropriately simple definitions. In particular, I **assume that the average stress drop of spatially complex ruptures is approximated by rectangular faults with homogeneous stress drop**. The adequacy of this approximation is discussed by Das (1988, Relation between average slip and average stress drop for rectangular faults with multiple asperities, BSSA, 78, 924-930) and Noda et. al. (2013).

Notice that all stress and strain changes scale linearly with slip. That is, whatever average stress drop corresponds to a particular distribution of slip, the average stress drop doubles when we double all of the slips.

I will use the relations between average slip, $\bar{D}$, rupture length, L, and rupture width, W, for a rectangular fault in a homogeneous half space as reported by Parsons (Parsons, Ian Dennis, 1988, *The Application of the Multigrid Method to the Finite Element Solution of Solid Mechanics Problems.* Dissertation (Ph.D.), California Institute of Technology. doi:10.7907/MVMM-ED69. **https://resolver.caltech.edu/CaltechETD:etd-08102006-090020**. This is actually a difficult solid mechanics problem and there are several flawed solutions that are in use. I believe that the Parsons solution is reliable.

In these discussions I will assume that the rupture surface can be approximated by a rectangle with a long dimension of L and a short dimension of W. In many cases, L becomes the along-strike dimension, whereas W is the down-dip dimension. Further, I will assume that

$$\overline{\Delta\sigma} \approx \mu\left(C\frac{\bar{D}}{W}\right) = \mu\left(C\frac{P}{LW^2}\right) \qquad 8.26$$

Or alternatively,

$$\overline{\Delta\varepsilon} \approx \left(C \frac{\bar{D}}{W} \right) = \left(C \frac{P}{LW^2} \right) \quad 8.27$$

Where $P = LW\bar{D}$ is the Potency (the volume integral of inelastic shear strain in an earthquake, see Chapter 7), C is a dimensionless constant that depends on the aspect ratio of the rupture, $\frac{L}{W}$, the depth of burial of the rupture z_{top}, and the rupture dip and rake angles. Of course, since I am assuming a homogeneous half-space, it is a trivial change to write 8.26 in terms of seismic moment $M_0 = \mu P$. Parsons (1988) shows that for a fault with rupture that reaches the Earth's surface,

$$C \approx 1.39\left[2.3^{\left(1-L/W\right)} \right] + 0.65 \quad 8.28$$

And for a deeply buried fault

$$C \approx 1.27\left[2.3^{\left(1-L/W\right)} \right] + 1.28 \quad 8.29$$

For ruptures with shallow burial depth, C changes quickly as a function of $\left(z_{top}/W \right)$; C is within 95% of its value for deeply buried when $\left(z_{top}/W \right) = 0.16$. This rapid change in C with burial depth is caused by the fact that, in an elastic model, there are very large strains and stresses in the region between the top of the fault and the free surface. In reality, it is highly questionable whether such large stresses develop in the shallow region just above a rupture that does not quite reach the surface. For simplicity, I classify events as either "surface rupturing" or "deeply buried"; that is, I assume that C is described by 8.29 for any earthquake that did not have surface rupture. I assume a uniform rigidity of $\mu \approx 35GPa$, which is taken to represent average properties in the upper Crust.

Notice that 8.21 can be rewritten as

$$\bar{D} \approx 3\times10^{-5}\sqrt{S} = 3\times10^{-5}\sqrt{LW} \quad 8.30$$

If I assume that ruptures are geometrically similar, then $L/W \sim a$, where a is the aspect ratio between rupture length and rupture width. In this case,

$$\bar{D} \approx 3\times10^{-5} W\sqrt{a} \quad 8.31$$

If typical ruptures are as long as they are wide, then $\bar{D}/W \approx 3\times10^{-5}$. **That is, the typical slip to length ratio of earthquakes is about 3 parts in 100,000.** Substituting this into 8.26 gives

$$\Delta\sigma \approx \mu C \times 3\times10^{-5} \qquad 8.32$$

And

$$\overline{\Delta\varepsilon} \approx 7.6\times10^{-5} \qquad 8.33$$

If the rupture is deeply buried and if the aspect ratio is about 1, then $C \approx 2.55$. If $\mu \approx 35GPa$, then we conclude that $\Delta\sigma \approx 2.7MPa$ and $\overline{\Delta\varepsilon} \approx C\times3\times10^{-5}$. While most current analysis uses stress drop as a parameter, I find that it's easier to use strain drop since 1) it's directly related to the observables (Potency and Rupture area), and 2) it is a unitless quantity.

If you inspect Fig. 8.5, you will see diagonal lines with labels of different stress drops given in bars. A bar is defined to be mean atmospheric pressure at sea level, and it is approximately equal to 100 kPa. The use of bars to describe stress was common decades ago, but almost all research journals now require the use of Standard International (SI) units, where stress has units of Pascals, Pa ≡ N/m^2.

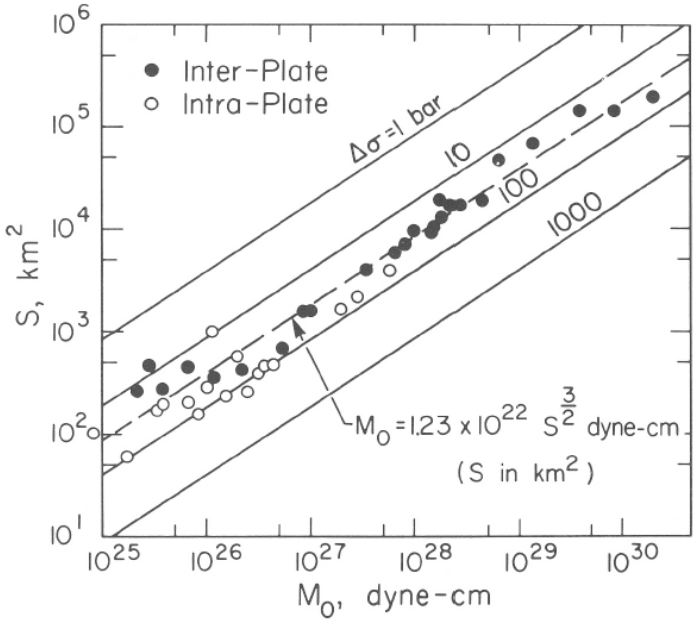


Fig. 12. Relationship between fault area and seismic moment. The solid lines denote constant stress drops and the dashed line ($M_0 = 1.23 \times 10^{22}\, S^{3/2}$ dyne-cm; S in km^2) signifies a stress drop of about 30 bars. Solid points denote interplate earthquakes and open circle denote intraplate earthquakes. This figure is taken from Kanamori (1980) and is used with permission from the author.

Figure 8.5. This is a copy of a Figure in Heaton, Tajima, and Mori,1986, Surveys in Geophysics, V. 8, pp 25-83, which is a copy of a very famous figure by Kanamori, 1980.

Earthquakes modeled as Brittle Cracks

In the discussion in the previous section about stress drop, we approximated an earthquake as having a constant change in stress on a planar area. This problem is almost identical to the problem of introducing a frictionless crack into a uniformly stressed elastic material. This allows us to borrow the extensive research developed to describe the failure of brittle materials. Crack-like models of dynamic rupture were common in the 1960's and 70's (e.g., Kostrov, 1966, Self-similar problems of propagation of shear cracks, J. App. Math. Mech., 28, 1077-1087; Madariaga, 1976,).

In Chapter 3, I described the solution of an instantaneous pressure change in a spherical cavity (3.82). In that solution, the cavity expanded when the pressure increased. In addition, the expansion looked like the solution to an impulse of momentum on a 58% damped sdof. That is the cavity oscillated about its new static equilibrium and it radiated wave energy as it oscillated. In the case of a frictionless crack with a step change in shear traction, the crack also oscillates about its new equilibrium. I don't know any closed form solution to this problem, but Brad Aagaard made some finite element simulations of a frictionless crack and he found that it had radiation damping of about 20% (unpublished research). Fault friction differs from the frictionless case in that once the slip drops to zero, the friction returns and presumably the motion stops. Thirty years ago there was discussion about whether the slip would stop at its dynamic maximum, or whether it would stop at its static equilibrium. This was referred to as the "overshoot" problem. In the past few decades, researchers construct numerical simulations of faults with very specific friction laws that are more complex than a simple instantaneous change in traction.

True analytic solutions of spontaneously propagating ruptures are rare. One simple approach to the problem of a steadily propagating semi-infinite crack (shown in Figure 8.6) is to start with the spatial solution for a static semi-infinite crack. To obtain the solution for a propagating crack, merely move the observer at a steady rate opposite to the direction of the crack propagation; the transforming velocity is scaled by the shear-wave velocity (similar to a relativistic Lorentz transformation).

The static solution for a semi-infinite shear crack with a constant stress drop is

$$\sigma_{xz} \sim \frac{K}{\sqrt{r}} f_{xz}(\varphi) \tag{8.34}$$

Where K is the **stress intensity factor** (a measure of the sharpness of the crack tip), r is distance from the crack tip, and $f_{xz}(\varphi)$ is a scalar function to account for the polar coordinate φ , which is the angular position of the observer as measured with respect to the rupture plane. This solution assumes that everything is uniform in the *y* direction. That is the crack extends infinitely in *y* direction. For a mode III crack, $K_{III} = \sqrt{2\mu G}$, where G is the ***fracture energy***, which is the energy per unit length required to extend a semi-infinite crack. If the fracture energy becomes small, then the tip of the crack approaches the linearly elastic solution, which has infinite stress at the crack tip. That is, if the crack tip is linearly elastic, then even small applied stresses will cause very large stresses at the crack tip.

The work expended on the crack of length L by slippage D is just $DL\left(\sigma_f + \frac{1}{2}\Delta\sigma\right)$. In a true crack, the sliding surface is frictionless $(\sigma_f = 0)$ and the total change in potential energy is

$$\Delta E_{potential} \approx DL\left(\frac{1}{2}\Delta\sigma\right) \approx \left(C/2\mu\right) L\Delta\sigma L\Delta\sigma = \left(C/2\mu\right) L^2\Delta\sigma^2 \tag{8.35}$$

When a mode III crack is extending in a steady-state way (constant rupture velocity), then the rate change of potential energy is balanced by the rate of change of fracture energy, or $\dot{E}_{fracture} = \dot{E}_{potential}$. The rate of change of fracture energy is just

$$\dot{E}_{fracture} = GV_R \tag{8.36}$$

where V_R is the rupture velocity. The rate of change of potential energy is calculated by differentiating 8.35 with respect to time, or

$$\Delta\dot{E}_{potential} \approx \left(C/\mu\right) L\dot{L}\Delta\sigma^2 = \left(C/\mu\right) LV_R\Delta\sigma^2 \tag{8.37}$$

which means that

$$G = \left(C/\mu\right) L\Delta\sigma^2 \tag{8.38}$$

For an equilibrium crack, the shear stress on the fault plane scales as

$$\sigma_{xz} \sim \frac{L\Delta\sigma}{\sqrt{x}} \tag{8.39}$$

For a propagating crack, this transforms to

$$\sigma_{xz} \sim \frac{L\Delta\sigma}{\sqrt{x'}} \tag{8.40}$$

Where x' is time-transformed distance given by

$$x' = \frac{1}{\sqrt{1-\left(\frac{V_R}{\beta}\right)^2}}\left(x - V_R t\right) \tag{8.41}$$

As the rupture velocity approaches the intrinsic shear wave velocity, the solution appears to be compressed in space. Stress at the equilibrium crack tip is singular. When the crack runs at the shear-wave velocity, it becomes very singular.

In crack theory, an expanding crack under uniform load tends to continue to expand indefinitely. That is, once it becomes unstable, the only way to stop it is with high fracture energy. The longer the crack length, the harder it is to stop, since the required fracture energy increases as $L\left(\Delta\sigma\right)^2$.

The laboratory-measured values of fracture energy of materials depend strongly on the brittleness of the material. Although mono-crystalline diamond is the hardest known material, it is also extremely brittle and only 5.5 J/m^2 is required to propagate a mode I fracture along a cleavage plane. This means that diamond is very strong at the microscale, but you wouldn't want to construct anything large out of a single diamond crystal; if a crack was initiated, then it would not take much average stress to propagate a crack through the rest of the material. Glass is another brittle material and its measured fracture energies range from 3.5 to 5.5 J/m^2.

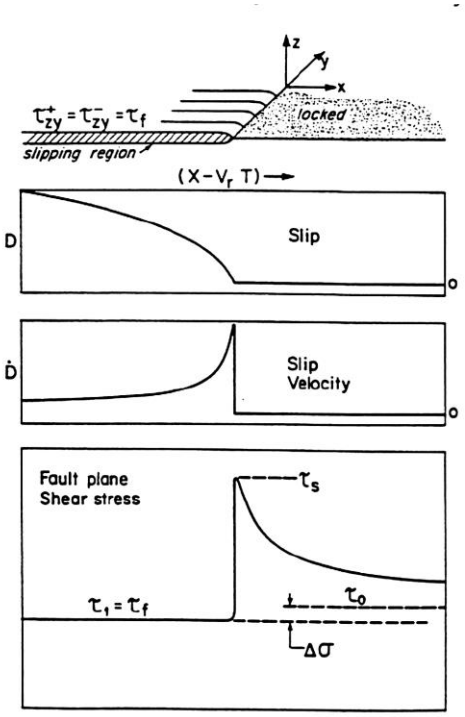


Figure 8.6. Simplified view of an earthquake as an expanding crack. This is often called a semi-infinite crack since it is uniform in the y coordinate.

Polycrystalline materials tend to have higher fracture energies because the fracture surfaces are complex and more molecular bonds must be severed than for a similar dimension of single crystalline material. Concrete has a fracture energy of about 190 J/m^2, and most polycrystalline rocks (e.g., granite) have fracture energies between 100 and 1,000 J/m^2.

In contrast, earthquake energies are much, much larger than measured values of G. For example, 8.7 gives estimates of radiated energies of

$$\frac{E_R}{S} \approx \sigma_{eff} D \approx 2D \ \left(\mathrm{MJm^{-2}}\right) \qquad 8.42$$

That is, the radiated energy for a typical 1-m slip is about 2 MJ/m^2, which is at least 1,000 times larger than measured fracture energies measured in a laboratory. Or, alternatively, a 10 km by 10 km rupture with a stress drop of 3 MPa has a fracture energy of 30 MJ/m^2 (see 8.38).

This is a **MAJOR PROBLEM** with interpreting earthquakes using simple constant-stress-drop cracks. That is, **constant stress drop cracks are NOT SCALE FREE since they imply that fracture energies (the work to stop ruptures) increases as** L . The notion that the earthquake dynamics are

controlled by fracture energy processes at the crack tip that can be 10's of km away seems totally implausible to me. Unfortunately, this idea is central to many popular earthquake physics studies as I will discuss shortly.

Earthquake Similarity

Notice that Fig. 8.5 shows that the average stress drop of 2.7 MPa is a kind of universal relation that applies over a very wide range of earthquake sizes. In essence, the observation is that average slip correlates with rupture dimension as stated by 8.31. The apparent universality of this relationship was the motivation for a class of earthquake scaling relations that are often referred to as "self-similar rupture models." The use of the term, self-similar, has created significant confusion, and I recommend using it only when you mean it.

Self-similar is usually used in physics and mathematics to describe **objects that look similar, independent of the magnification** at which the object is observed. In particular, **fractal objects that have similar complexity at all scales are referred to as self-similar.** That is, they look like themselves independent of the magnification. In contrast, two objects are **called similar when they look identical after an appropriate scale change**. For example, squares are all geometrically similar; all squares are identical if the length of a side is normalized to unity (remember similar triangles that you learned in high school).

Another related term that you might encounter is **self affine**. Objects that are mapped in multiple dimensions are typically called self-affine when linear scale changes in some of the dimensions result in objects that are self-similar. Again, this term is common in the study of fractals.

Aki (1967, Scaling law of seismic spectrum, JGR, 72, 1217-1231) proposed that all earthquakes were similar; he did not call then self-similar. That is, he suggested that different ruptures could be seen to be identical except for the length of the rupture. Aki's argument was simple and compelling. He argued that all earthquakes were fundamentally controlled by the same 3 MPa sudden drop in shear stress on the fault surface. That is, he argued that the stress drop is a fundamental (and universal) description of the friction on the fault. The initial stress σ_0 is assumed to be the strength of the fault, and the sliding friction, σ_f , is assumed to be equal to the final stress σ_1. Later this general model has been called "self-similar " [sic] constant-stress-drop model of earthquake ruptures. It has been the basis of numerous models to explain seismic data (especially the shape of Fourier amplitude spectra). According to Aki's conjecture, the primary difference between different earthquakes is the area and aspect ratio of a rupture.

That is, once the rupture length is chosen, then all other dynamic parameters can be inferred. That is, all of the dynamics are controlled by physics that is independent of the scale length of the earthquake. In this way, these models are **scale invariant** (not really, though, because fracture energy increases with rupture dimension).

In the following pages I will describe the standard "self-similar" [sic] source model so that you will understand the origin of numerous terms that are commonly used. However, I warn you now that I view that these models are seriously flawed, and their use typically decreases real understanding. Furthermore, I will develop arguments that show that earthquakes are not scale invariant.

As we saw in Chapter 7, the displacements $u(t)$ from far-field S-waves radiated by a sliding fault is proportional to the potency rate, $\dot{P}$. Now so long as $P(t)$ is monotonically increasing with time, then $\dot{P}(t)$ *is a strictly positive function*. Now we can conclude that

$$P = \int_0^T \dot{P} dt \sim \int_0^T u dt \qquad 8.43$$

Where T is the total duration of the earthquake. While we could evaluate P by a simple integration of $u(t)$ in time, it has become customary to perform this integration in the Fourier transformed domain. In particular,

$$\lim_{\omega \to 0} \left| \tilde{u}(\omega) \right| = \left| \int_{-\infty}^{\infty} u(t) dt \right| \sim P \qquad 8.44$$

That is, the earthquake potency is proportional to the zero-frequency-amplitude of the Fourier transform of the far-field S-wave. Of course, one must correct for radiation pattern and path effects (e.g., transmission and reflection coefficients and geometric spreading). In fact, the seismograms of what we usually call the S-wave are actually comprised of a combination of many different rays that sum together to comprise the "S-wave group." In order to deduce potency, P, from this data, one should really account for all of these different rays. However, it's far simpler to just take a Fourier transform of the S-wave group and to then find the amplitude at zero frequency. It's usually assumed that the complexities from many rays serves to make the phase spectrum look random while there is little systematic effect on the overall amplitude spectrum. While this assumption is pervasive, I am unaware of any study that provides convincing evidence to support it. For the purposes of the following discussion, though, I will assume that the spectra of real earthquakes are the same as the spectra of far-field waves

radiated from a shear fault located in a homogeneous, isotropic whole space (***caveat emptor***).

Suppose that $u(t)$ is a strictly positive function of time and having a duration of T_c, then $|\tilde{u}(0 \le f \le f_c)| = \int_0^{T_c} u(t)dt \sim P$ and $|\tilde{u}(f > f_c)| < |\tilde{u}(f_c)|$, where $f_c = 1/T_c$ is the "corner frequency." As an example, consider a far-field S-wave that consists of a simple rectangle, or

$$u(t) = \Pi(t) \equiv \begin{bmatrix} 0 & t < -\frac{1}{2} \\ 1 & -\frac{1}{2} < t < \frac{1}{2} \\ 0 & t > \frac{1}{2} \end{bmatrix} \qquad 8.45$$

(see chapter 1 for discussion of the rectangle function). Now $\tilde{u}(f) = \text{sinc}(f) \equiv \frac{\sin f}{f}$.

$\tilde{u}(f=0) = 1$, which is the integral of the rectangle function. Furthermore, $|\tilde{u}(f \gg 1)| \to \frac{1}{f}$. The high-frequency asymptote, $|\tilde{u}| = \frac{1}{f}$ intersects the low-frequency asymptote, $|\tilde{u}| = 1$, at the corner frequency, $f_c = 1$, which is the reciprocal of the duration of the rectangle function. While f_c is well known for a rectangle function, it is something that must be determined when it is used to characterize an S-wave group. It is customary to plot the amplitude spectrum of the wave group on a log-log graph and to fit straight lines to the low frequencies and the high frequencies. The intersection of the two lines provides a measurement of the corner frequency.

We can use the corner frequency measurement to estimate the rupture dimension,

$$L \approx \frac{1}{2} T_c V_R = \frac{1}{2} \frac{V_R}{f_c} \qquad 8.46$$

where I have assumed that the earthquake duration is approximately twice the time it takes for a rupture front to sweep across the rupture surface. The factor of two is to account for the average duration of the slip. Assuming that $L \approx W$, $C \approx 2.55$, and applying 8.26 we obtain

$$\overline{\Delta\sigma} \approx 20\mu P\left(\frac{f_c}{V_R}\right)^3 \qquad 8.47$$

Or alternatively,

$$\overline{\Delta\varepsilon} \approx 20P\left(\frac{f_c}{V_R}\right)^3 \qquad 8.48$$

Notice that we made a questionable assumption that the rupture velocity is a constant, independent of the earthquake size[1]. Furthermore, the stress drop is very sensitive to our funky measurement of the duration of the S-wave, f_c^{-1} , since the stress drop depends on the cube of the corner frequency; if your measurement of corner frequency is off by a factor of two, then your stress drop estimate is off by a factor of eight. Although I don't recommend it, 8.47 can be used to estimate average stress drop (or strain change) by simply measuring the asymptotic behavior of Fourier amplitude spectra. If you are genuinely interested in stress drop, then I strongly recommend that you obtain it from the types of finite-fault slip models that were described in Chapter 7. In particular, finite-element models can be used to calculate the spatial distribution of stress drop (or strain change) based on the spatial distribution of slip.

Aki's conjecture of universal similarity (1967) suggested that all earthquakes were identical dynamic processes except for the rupture dimension. That is, he suggested that all events had the same stress drop and that everything else could be deduced from this stress drop and the rupture dimension. In terms of the Fourier amplitude spectrum, this means that everything is controlled by the corner frequency, which is a description of the rupture dimension. Now if there is only one variable in our dynamics problem (the rupture length), then the shape of the Fourier amplitude spectrum must be described by only one variable, f_c . In this case, length is mapped to corner frequency through the rupture velocity. Aki argued that the spectrum must be of the following form

$$\left|\tilde{u}\right| \sim \frac{1}{\beta r}\frac{P}{1+\left(\frac{f}{f_c}\right)^{\alpha}} \qquad 8.49$$

where α is a constant that is determined by the dynamics of the rupture process. Note that Aki used M_0 instead of P, so his equations look a little different. The r^{-1} is to account for geometric spreading of a far-field body wave, and the β^{-1} is an

[1] This is an important assumption of the Brune source model, but the evidence is that the duration of slip at any point is small compared to the time it takes a rupture to propagate on a rupture surface.

impedance factor that tells the size of radiated waves relative to the near-source term (look back at the far-field terms in 7.45). We can now estimate the energy in the radiated wavefield as follows. The power $\dot{E}_R$ flowing through an increment of area ΔS in the radiated S-wave is (see Chapter 3, eqn 3.55)

$$\frac{\dot{E}_R}{\Delta S} = \sigma \dot{u} = \rho \beta \dot{u}^2 \tag{8.50}$$

Since this energy is traveling radially to great distance, We can calculate the radiated energy traveling through an element of surface area dS by integrating 8.50 with respect to time. That is,

$$\frac{E_R}{\Delta S} = \rho \beta \int_0^\infty \dot{u}^2 dt = \rho \beta \int_0^\infty \tilde{\dot{u}}^2 df = \rho \beta (2\pi)^2 \int_0^\infty (f\tilde{u})^2 df \tag{8.51}$$

Our integral over time is transformed to an integral over frequency using Parseval's theorem. The integral over frequency can be broken into two separate integrals, the first over the constant amplitude at frequencies less than f_c , and the second over the frequencies higher than f_c . Combining 8.49 and 8.51, we obtain

$$\left| \frac{E_R}{\Delta S} \right| = \frac{\rho}{r^2 \beta} (2\pi)^2 P^2 \int_0^\infty \frac{f^2}{\left[1 + \left(\frac{f}{f_c} \right)^\alpha \right]^2} df \tag{8.52}$$

Unfortunately, the solution to this integral is a hypergeometric function, which is not very useful for this discussion. Fortunately, the integral can be solved in closed form if $\alpha = 2$ (the commonly used f^{-2} model). In this case

$$\frac{E_R}{\Delta S} = \frac{\pi^3 \rho}{\beta r^2} P^2 f_c^3 \tag{8.53}$$

provided that $\alpha \geq 3/2$; the energy at high-frequencies is infinite for any $\alpha < 1/2$. Notice that 8.47 can be used to conclude that

$$f_c \approx \left(\frac{V_R^3}{20\mu} \frac{\overline{\Delta\sigma}}{P} \right)^{1/3} \tag{8.54}$$

Or even more conveniently,

$$Pf_c^3 \approx \frac{V_R^3}{20}\overline{\Delta\varepsilon} \qquad 8.55$$

or

$$\overline{\Delta\varepsilon} \approx \frac{20Pf_c^3}{V_R^3} \qquad 8.56$$

so 8.53 can be written

$$E_R \approx \pi^3 \rho V_R \left(\frac{V_R}{\beta}\right)^2 \Delta\sigma P \qquad 8.57$$

While the issue of spectral scaling for $f < f_c$ is well studied and non-controversial, the spectral characteristics for $f > f_c$ have been the focus of a huge number of studies, and unfortunately there is a depressing lack of consensus about spectral scaling for radiated high frequencies. Aki (1967) presented both an $\alpha = 2$ model and also an $\alpha = 3$ model. He argued that the apparent geometric similarity of ruptures (approximately scale independent stress drop) was a sign that models should be appropriately simple, which he interpreted to mean that spectral scaling should have power-law scaling with a simple integer in the power law. He said he did not know whether f^{-2} or f^{-3} was more appropriate (f^{-1} leads to infinite energy), but based on sparse data, he had a preference for f^{-2} (often called the omega squared model) . Figure 8.7 shows Aki's two hypothesized spectral models.

While Aki's hypothesized model was scale independent in many ways, it is clearly not scale independent when it comes to fracture energy. That is, Aki's model is essentially a constant stress drop crack model. This is the class of models that require a fracture energy that increases with rupture dimension. Later I will show that slip-pulse models can be constructed in a way that they produce scale-independent stress drop and also scale-independent fracture energy.

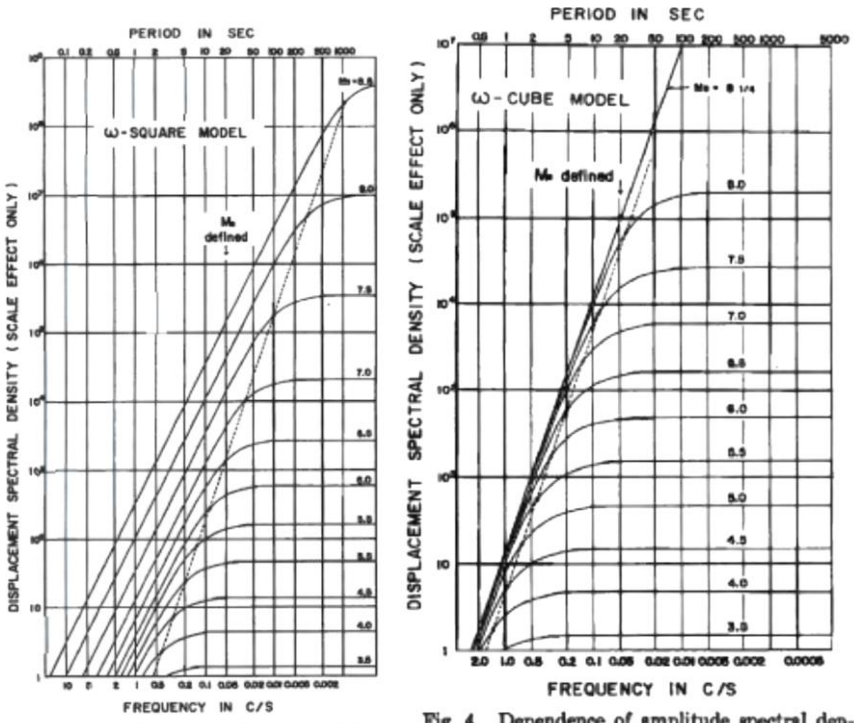


Fig. 3. Dependence of amplitude spectral density of earthquake magnitude M, for the ω-square model.

Fig. 4. Dependence of amplitude spectral density on earthquake magnitude M, for the ω-cube model.

Figure 8.7 Aki's spectral models. The f^{-2} model on the left is often referred to as a self-similar (sic.) model. (Aki, K., 1967, J Geophys. Res. Scaling law of Seismic Spectrum, v. 72, 1217-1231)

Brune's Spectral Scaling Model

Brune's 1970 paper, Tectonic stress and the spectra of seismic shear waves from earthquakes (J. Geophys. Res., v.75, 4997-5009) is one of the most cited papers in all of earthquake research (approximately 5,000 citations according to Google Scholar). This paper is basically an extension of Aki's 1967 paper on universal scaling. In this paper, Brune argues that the dynamics of earthquake slips are controlled by an effective shear stress, σ_e, that accelerates the faces of an earthquake fault. That is, he argued that the fault appears as if it's a free surface with an instantaneously applied shear traction that is equal to σ_e. Brune then introduced the slip history for a fault that experiences an instantaneous step in shear traction. He uses this solution to infer the spectral characteristics of radiated far-field waves. That is, Brune argues that he introduced dynamics into the types of spectral models developed by Aki. Brune argues that one can derive the effective stress by appropriate modeling of the envelope of the Fourier amplitude spectra of seismograms. Notice that in the previous discussions, I was careful to describe the solutions using either change in strain, or in terms of stress drop. If the material is linearly elastic, strain and stress are linked by the shear modulus. The seismograms used to construct spectra only

measure lengths and time, so the parameter strain change is natural. In contrast, Brune's analysis is formulated in terms of stress drop. In Brune's model, the stress drop defines the shear tractions that accelerate the sides of the fault.

I will now try to describe Brune's 1970 spectral model, although that is a difficult task. In particular, what is currently called the Brune model is different from what is presented in his 1970 paper. In addition, Brune's induction about rupture physics has some rather glaring errors that are seldomly acknowledged.

I will begin by saying that the spectral envelopes of a wide range of seismograms can be approximately described with the Brune model; this is what makes this model so popular. Unfortunately, the glaring physics errors also make this paper confusing; there have been countless misuses of this paper (*caveat emptor*).

Brune relies heavily on a particularly simple solution for the displacements that occur due to a step in shear traction on a free surface. This problem was described back in Chapter 3 (see 3.58). In particular, if an instantaneous change in shear traction, $T_x = \sigma_e H(t)$ occurs uniformly on the surface of a uniform half space, where the plane is described by $z = 0$, then a uniform shear stress propagates into the medium (in the z direction) at the shear-wave velocity. That is, the solution is just $\sigma_{xz} = \sigma_e H\left(t - \frac{z}{\beta}\right)$. This means that $\varepsilon_{xz} = \frac{\sigma_e}{\mu} H\left(t - \frac{z}{\beta}\right) = \frac{\partial u_x}{\partial z}$. In equation (3.68), we derived that $\varepsilon_{xz} = \frac{1}{2}\frac{\dot{u}_x}{\beta}$. So $\dot{u}_x = 2\beta\varepsilon_{xz} = \frac{2\beta}{\mu}\sigma_e H\left(t - \frac{z}{\beta}\right)$. This can be integrated to give the displacement,

$$u_x = \frac{2\beta}{\mu}\sigma_e\left(t - \frac{z}{\beta}\right)H\left(t - \frac{z}{\beta}\right) \qquad 8.58$$

This is a simple linear ramp in time that starts at $t = \frac{z}{\beta}$. The Fourier amplitude spectrum of this displacement can be easily deduced by recognizing that a linear ramp is the second time integral of an impulse. Now recall that integration in the time domain is the same as dividing by $i\omega = 2\pi i f$ in the frequency domain. Now since $\left|\tilde{\delta}(f)\right| = 1$, $\left|\tilde{u}_x\right| = \frac{-2\beta}{\mu}\sigma_e\left(\frac{1}{2\pi f}\right)^2$. **That is, a simultaneous step in stress on the fault should produce ground displacement with an f^{-2} spectrum.** Brune argued that the accelerations experienced adjacent to the rupture were a measure of the effective stress accelerating the sides of the fault.

The solution that Brune used **assumes uniform stress applied instantaneously on an infinitely large fault**, and the resulting fault slip increases linearly with time indefinitely. That is, the slip trends to infinity. Brune reasoned that the final slip should approach the slip expected from fault dimensions and the appropriate stress drop (similar to Aki's argument). He hypothesized that, in the very near-source region,

$$u_x \approx \frac{\sigma_e}{\mu}\beta\tau\left(1-e^{\frac{t}{\tau}}\right)H\left(t-\frac{z}{\beta}\right) \qquad 8.59$$

where $\tau = r/\beta$, and r is the radius of an approximately circular rupture area. This leads to particle velocities given by

$$\dot{u}_x = \frac{\sigma_e}{\mu}\beta e^{-t/\tau}H\left(t-\frac{z}{\beta}\right) \qquad 8.60$$

The Fourier amplitude spectrum of 8.59 is

$$\left|\tilde{u}_x\right| = \frac{\sigma_e}{\mu}\beta\frac{1}{\left(1+i2\pi f/f_c\right)^2} \qquad 8.61$$

As before, this spectrum is f^{-2} , **which is the inevitable result of the discontinuous jump in particle velocity at the start of the motion**. Brune argues that this jump is fundamental to the dynamics of earthquakes. Furthermore, he argued that the dynamic stress controls the particle accelerations in the very near-source region. Notice that the particle acceleration that is derived by differentiating 8.60 is

$$\ddot{u}_x = \frac{\sigma_e}{\mu}\beta e^{-\left(t-z/\beta\right)/\tau}\left[\left(\frac{1}{\tau}\right)H\left(t-\frac{z}{\beta}\right)+\delta\left(t-\frac{z}{\beta}\right)\right] \qquad 8.62$$

The **impulse function comes from differentiation of the step function. This indicates infinite accelerations at the arrival time of the shear wave. Brune argued that structural heterogeneity scattered these high-frequency waves and that the result is random white noise whose amplitude is determined by the effective stress**.

For frequencies between $f = 0$ and $f = f_c$, Brune's spectrum 8.61 is similar to the Aki spectral model. Brune's argument about instantaneous stress drop led him to conclude that the high-frequency decay was naturally f^{-2} and that the accelerations had a white-noise spectrum whose amplitude scaled linearly with the effective stress. When Brune imagined this model, he was aware that earthquake engineers were

simulating recorded near-source accelerograms as Gaussian white noise modulated by an envelope function in time. **Brune then appealed to conservation of radiated energy (through different enclosing spheres) to conclude that if the high-frequency S-wave is f^{-2} in the near-source region, then it must also be f^{-2} in distances that are large compared to the source dimension.**

At this point, Brune had used inductive reasoning to describe a Fourier amplitude spectrum for a distant observer. This spectrum had a long-period amplitude that is scaled by the potency (he used moment), a corner frequency that described the dimension of the rupture (used to determine stress drop), and an f^{-2} high-frequency radiation whose overall amplitude is controlled by the effective stress. In his 1968 paper, Brune stitched these different features together to hypothesize that the displacements in seismograms observed at distances large compared to the rupture dimension should be of the form

$$u \sim f\left(\frac{r}{R}\right)\frac{\sigma}{\mu}\beta t \exp(t\alpha) H(t) \qquad 8.63$$

and that the distant spectral scaling should be

$$\left|\tilde{u}(f)\right| \approx \mathcal{R}_{\theta\varphi}\frac{\sigma_e \beta}{\mu}\frac{r}{R}F(\varepsilon)\frac{1}{4\pi^2 f^2 + \alpha^2} \qquad 8.64$$

where $\mathcal{R}$ is a radiation pattern, r is the radius of the rupture area, R is the distance of the observer, $\alpha = {2.21\beta}/{r}$ (the corner frequency). and $\varepsilon \equiv \frac{\Delta\sigma}{\sigma_e}$, which Brune called the fraction of the stress drop. $F(\varepsilon) = \left\{(2-2\varepsilon)\left[1-\cos\left({2.42\pi\varepsilon f}/{\alpha}\right)\right]+\varepsilon^2\right\}^{1/2}$; this factor joins the high-frequency levels to the corner with a section of the spectrum that decays as f^{-1}. The effect of partial stress drop ($\varepsilon < 1$) is shown in figure 8.8.

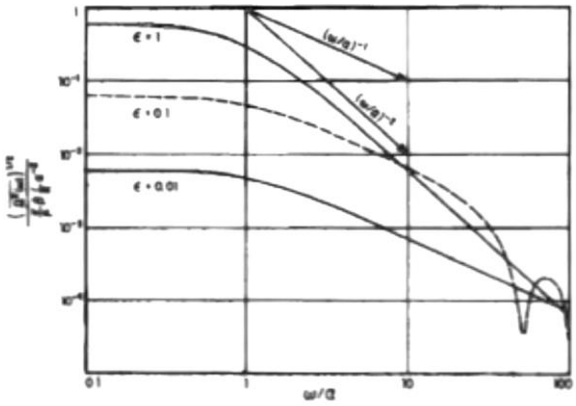


Fig. 5. Average (rms) far-field spectral density curves.

Figure 8.8. From Brune (1970). This is the "Brune far-field spectral model." If the stress drop is the effective stress then it's an f^{-2} with a single corner. If the stress drop is less than the effective stress, then there are both f^{-2} and f^{-1} falloffs with two corners.

Although Brune introduced the concept of **partial stress drop,** it is very rarely used, and most researchers use a simpler version of the **Brune spectrum** as given by

$$\left|\tilde{u}(f)\right| \sim \frac{P}{1+\left(\frac{f}{f_c}\right)^2} \qquad 8.65$$

This simplified form assumes that the effective stress (which regulates high-frequency radiation) equals the final change in stress (which is controlled by the rupture dimension and the Potency). This simplified form of the model links the high-frequency near-source radiated energy to the overall size of the earthquake (P) and the stress drop. This can be seen in the following argument, which I developed in 2013.

Alternate Derivation of Brune Spectral Scaling.

Although I read Brune's paper many times, I find that its logic is difficult to follow. To better understand the "essence" of the Brune model, I devised an alternate derivation of his spectral law for waves observed at distances larger than the rupture distance. My goal is to start the derivation with a minimum number of simple assumptions. The following is my derivation. It begins with several key assumptions that may be stated somewhat differently than Brune did.

1) **Assume that the duration of the S-wave group is proportional to the dimension of the fault** (that is, assume constant rupture velocity), this assumption can be restated $f_c \sim \frac{1}{V_R\sqrt{S}}$. Brune also made this assumption.
2) **Assume size similarity of the form** $P \sim S\bar{D} \sim S^{3/2}\Delta\sigma$**, or alternatively** $LW \sim P^{2/3}\Delta\sigma^{-2/3}$. Brune also made this assumption.
3) **Assume that the Fourier amplitude spectrum is a constant at low frequencies and** f^{-2} **at high frequencies as shown by equation** 8.65**.**

Assumptions 1 and 2 imply that $f_c \sim P^{-1/3}\Delta\sigma^{1/3}$. Substituting into 8.65, we obtain

$$\left|\tilde{u}(f)\right| \sim \frac{P}{1+\left(\frac{f}{P^{-1/3}\Delta\sigma^{1/3}}\right)^2} \tag{8.66}$$

Which has asymptotes

$$\left|\tilde{u}(f)\right| \sim \begin{cases} P & f \ll f_c \\ P^{1/3}\Delta\sigma^{2/3}f^{-2} & f \gg f_c \end{cases} \tag{8.67}$$

This is an alternate form of Brune's spectrum that is written as an explicit function of stress drop. Variations of Equation **Error! Reference source not found.** are so often used (aka, assumed) that I will refer to this as **the Standard Source Model.** Now the radiated energy spectrum scales as the square of the Fourier amplitude spectrum, or at high frequencies

$$E_R^{f\gg f_c} \sim \left(\tilde{U}_R^{f\gg f_c}\right)^2 \sim P^{2/3}\Delta\sigma^{4/3} \sim \left(S^{3/2}\Delta\sigma\right)^{2/3}\Delta\sigma^{4/3} = S\Delta\sigma^2 \tag{8.68}$$

That is, **Brune's spectral model predicts that high-frequency radiated energy scales with the rupture area and the square of the stress drop.** This is consistent with the notion that the amplitude of very high frequencies close to the rupture are a constant that is related to the effective stress. This is a key feature of

Brune's spectrum; the near-source acceleration should scale with stress drop. That is, **at very close distances, pga should be independent of the earthquake magnitude**, but it should scale linearly with stress drop. As I will shortly show, near-source pga is indeed independent of magnitude (M > 6). However, **near-source accelerograms show that near-source pga is also independent of stress drop. This turns out to be a fundamental flaw of the Standard Source model**.

Stress Drop and Near-Source Ground Motions

I will now discuss the scaling of near-source ground motions and how they scale with magnitude and stress drop. The work in this section is modified from an unpublished manuscript by Tom Heaton and Masumi Yamada. This is work from 2013.

Brune (1970) hypothesized that near-source pga correlates linearly with effective stress, σ_e . Since it's not currently feasible to measure σ_e independently from measures of radiated seismic waves, it is not feasible to independently verify Brune's hypothesis. It is, however, possible to investigate the correlation between near-source motions (less than 10 km from the surface projection of the fault rupture) and spatially averaged static stress drop, $\overline{\Delta\sigma}$; this is a meaningful test since the most common form of Brune's model assumes that $\Delta\sigma = \sigma_e$. In this study, we show that there is no significant correlation between near-source peak ground accelerations (pga's) and spatially averaged static stress drop $\overline{\Delta\sigma}$. In contrast, near-source long-period motions do correlate with stress drop.

Our current study investigates the relationship between static stress drop and the intensity of near-source ground motion as measured by peak ground acceleration (pga), peak ground velocity (pgv), and peak ground displacement (pgd). This work is a follow-on to our previous study of the statistical characteristics of near-source motions, where we reported that there is no apparent correlation between near-source pga and pgd (Yamada, Olsen, and Heaton, 2009, *Statistical Features of Short-Period and Long-Period Near-Source Ground Motions,* Bulletin of the Seismological Society of America, 99 (6). pp. 3264-3274. ISSN 0037-1106). In this earlier study, we hypothesized that pgd should correlate with stress drop; if all other source parameters are equal, then larger slips are associated with both larger near-source pgd's and also with larger stress drops. However, the lack of correlation between near-source pga and pgd seems to imply that pga and stress drop are also uncorrelated.

In this study, we estimate the average static stress drop for 20 earthquakes for which there are also near-source strong motion accelerograms. We then

investigate the relationship between measures of shaking intensity and earthquake potency and stress drop.

We estimate average static stress drop using the relations between average slip, $\bar{D}$, rupture length, L, and rupture width, W, for a rectangular fault in a homogeneous half space as reported by Parsons et. al. (1988) and presented in equations 8.26 through 8.29. In particular, we assume that

$$\overline{\Delta\sigma} \approx \mu\left(C\frac{\bar{D}}{W}\right) = \mu\left(C\frac{P}{LW^2}\right) \qquad 8.69$$

Where $P = LW\bar{D}$ is the Potency (the volume integral of inelastic shear strain in an earthquake), C is a dimensionless constant that depends on the aspect ratio of the rupture $\frac{L}{W}$, the depth of burial of the rupture z_{top}, and rupture dip and rake angle. Approximations for *C* are given in the section on stress drop.
For ruptures with shallow burial depth, C changes quickly as a function of $\left(z_{top}/W\right)$; C is within 95% of its value for deeply buried when $\left(z_{top}/W\right) = 0.16$. This rapid change in C with burial depth is caused by the fact that, in an elastic model, there are very large strains and stresses in the region between the top of the fault and the free surface. In reality, it is highly questionable whether such large stresses develop in the shallow region just above a rupture that does not quite reach the surface. For simplicity, we classified events as either "surface rupturing" or "deeply buried"; that is, we assumed that C is described by 8.29 for any earthquake that did not have surface rupture. We assume a uniform rigidity of $\mu = 35\,\text{GPa}$, which is taken to represent average properties in the upper Crust.

Table 8.3 lists the source parameters for the earthquakes in this study. Most of these models are described in the ETH source data base. For some of the events there are multiple source models and, in these cases, we assume that the models are log-normally distributed about the geometric mean of the models. Figure 8.10 shows the relationship between average static stress drop and Moment Magnitude, $\mathbf{M} \equiv 1 + \frac{2}{3} LogP$. In order to show the distribution of the different models, each of the earthquakes is assigned a single $\mathbf{M}$ that is based on the geometric mean of the Potencies of the different models. In a similar manner, the static stress drop values for each event are the geometric mean of the static stress drops for multiple models for each earthquake. Of course, it's difficult to estimate the uncertainty in these values (especially stress drop), but we have

included an error estimate that is based on the standard deviation of all of the models about their mean. We calculate the standard deviation using Bessel's correction for small number statistics. Earthquakes with only one source model are considered to be more uncertain than those that represent the geometric mean of multiple models

Earthquake	Year	Date	NS	mean(Mw)	mean(Δσ)	std(Δσ)
Imperial Valley	1979	10/15	12	6.53	1.00	0.33
Loma Prieta	1989	10/17	6	6.99	5.47	1.27
Landers	1992	6/28	1	7.25	2.89	1.00
Northridge	1994	1/17	7	6.79	3.17	1.15
Kobe	1995	1/17	4	6.96	1.25	0.50
Izmit	1999	8/17	3	7.51	2.01	0.88
Chi-Chi	1999	9/20	38	7.68	2.39	0.47
Western Tottori	2000	10/6	5	6.84	2.11	0.42
Denali	2002	11/3	1	7.91	2.97	1.66
Parkfield	2004	9/28	48	6.05	0.15	0.02
Mid Niigata	2004	10/23	6	6.72	0.97	0.00
Noto-Hanto	2007	3/25	1	6.74	1.63	0.00
Niigataken-Chuetsuoki	2007	7/16	2	6.81	2.17	0.00
Wenchuan	2008	5/12	6	8.00	1.67	0.00
Iwate-Miyagi	2008	6/14	6	6.96	2.43	0.00
Surugawan	2009	8/11	1	6.43	1.19	0.00
Darfield	2010	9/4	7	7.00	0.83	0.00
Christchurch	2011	2/22	15	6.36	2.83	0.00
Northern Nagano	2011	3/12	4	6.35	1.92	1.08
Fukushima-Hamadori	2011	4/11	3	6.69	1.13	0.00
Total			176			

Table 8.3. Earthquakes used in this study have finite source models tabulated by the SRCMOD project (Mai, ETH). NS is the number of stations. All statistics are computed assuming log-normal distributions and Bessel's correction for small samples is assumed to calculate the standard deviation of $\overline{\Delta\sigma}$.

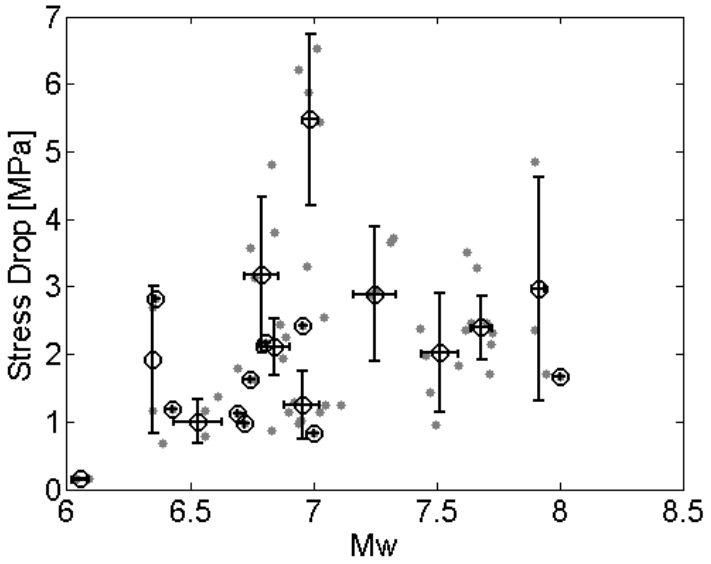


Figure 8.9. The values of $\overline{\Delta\sigma}$ plotted as a function of $\mathbf{M}$. Each dot represents a finite-source model compiled in the ETH database of finite source models. The circled points represent the mean model for a given earthquake. The "error bars" are the standard deviations for earthquakes that have more than one model.

We collected strong motion records from twenty earthquakes in table 8.3 from sites in the near-source region. We define the near-source region to be within 10 km of the surface projection of the fault rupture (often called the Joyner-Boore distance). We choose this number because it is small enough to demonstrate the source scaling, but large enough to include a sufficient number of records to characterize the statistics. To determine the pga, pgv and pgd of a recorded ground motion, we processed each time history. We first remove the bias from the acceleration record by subtracting the mean. Then we integrate once and remove the long-period component with a fourth-order high-pass Butterworth filter (13.3 sec corner period) to generate the velocity time history. The displacement records are obtained from the integration of velocity waveforms. Peak values are obtained from the square root of the sum of the squares of the north-south, east-west and up-down components at each time step of the record.

Figure 8.10 shows a log-log plot of near-source pga as a function of static stress drop. Figure 8.10 also shows a linear regression of log pga as a function of $\mathbf{M}$ and log $\Delta\sigma$. Compatible with the conjecture of Yamada and others (2009), we do not observe a significant correlation between near-source pga and either stress drop or Potency (Moment Magnitude). Notice that there is no magnitude dependence to the near-source pga. This is because we constrained the magnitude dependence to be positive only. Allowing decreasing near-source pga with increasing $\mathbf{M}$ actually provides a marginally better fit to the data, but including this possibility requires us to propose non-intuitive models that only fit highly scattered data marginally better.

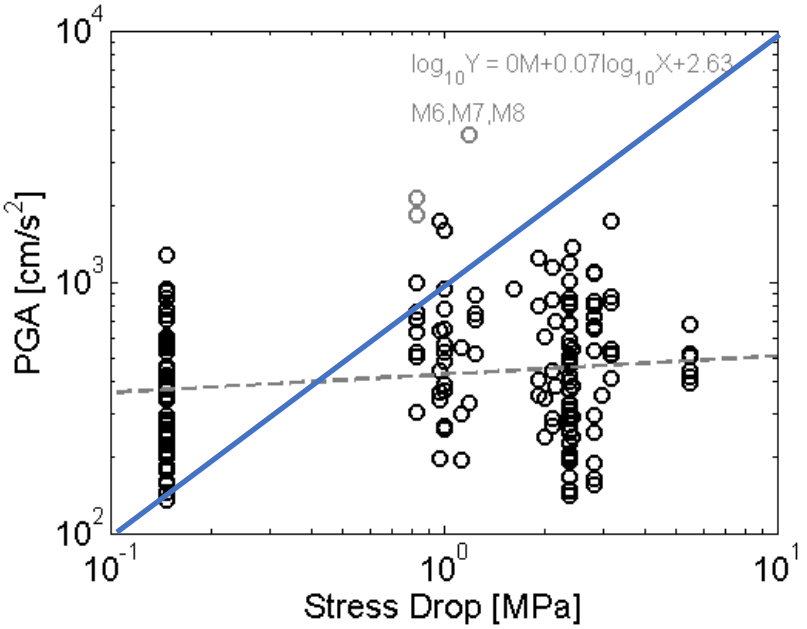


Figure 8.10. Near-source pga vs stress drop for all records used in this study. The dotted line represents the least-squares linear regression, $\log(pga) = 0.0M + 0.07\log\Delta\sigma + 2.63$. This seems to indicate that near-source pga is independent of magnitude (as Brune hypothesized) and also independent of stress drop (very different from Brune's hypothesis which is represented by the blue diagonal line).

Figures 8.11 and 8.12 are similar, but they are for near-source pgv and pgd, respectively.

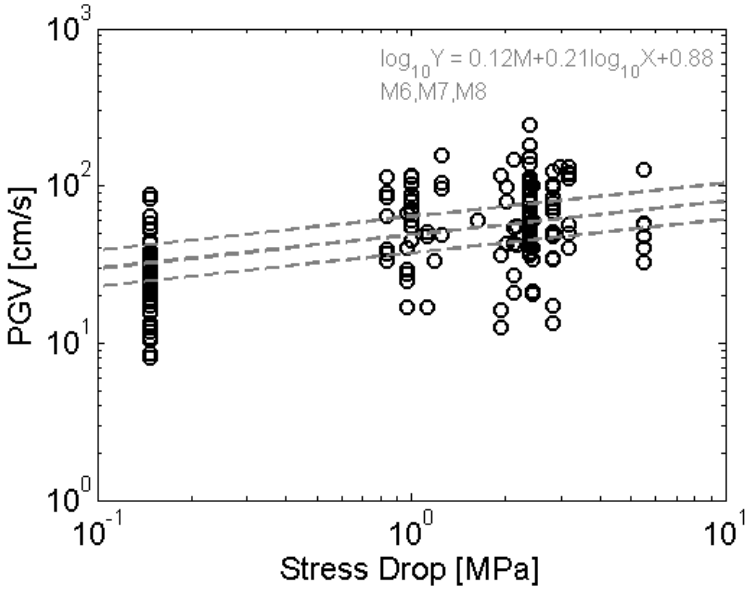


Figure 8.11. Near-source pgv vs stress drop for all records used in this study. The three dotted lines represent M 6, 7, and 8 in the linear least-squares regression

$$\log(pgv) = 0.12M + 0.21\log\Delta\sigma + 0.88$$

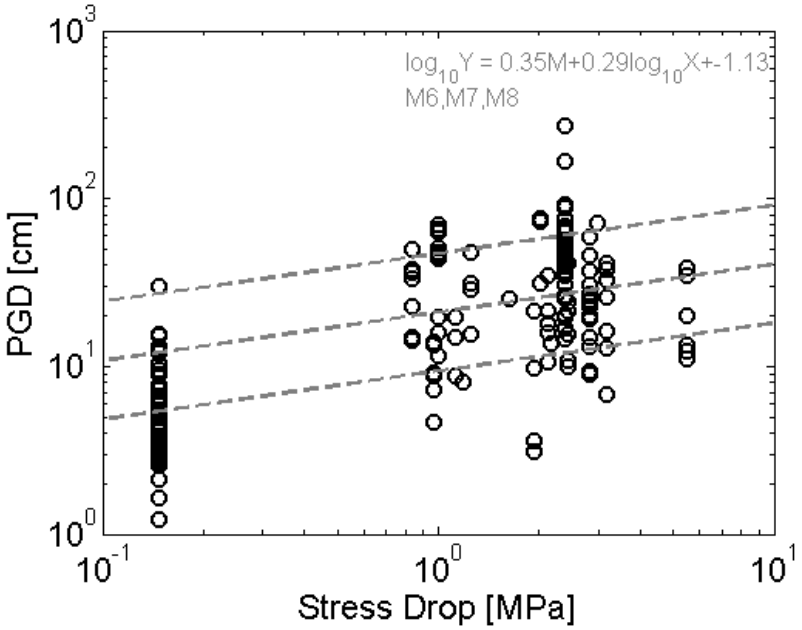


Figure 8.12. Near-source pgd vs stress drop for all records used in this study. The pgd is from high-pass filtered records with a 13 s corner. The three dotted lines represent M 6, 7,

and 8 in the linear least-squares regression $\log(pgd) = 0.35M + 0.29\log\Delta\sigma - 1.13$.

We can anticipate the following asymptotic behavior for any ground motion prediction equations. When the distance is large compared to the source dimension, and when the predominant periods of the ground motion are large compared to the source duration, we expect the ground motion amplitudes to scale with the seismic potency, or

$$\log U_{far\&lowfreq} \sim \log P \sim {}^{3}\!/_{2}\, M \tag{8.70}$$

For very near-source long-periods (e.g. displacement), we expect the peak amplitude to scale with the size of the slip on the nearby fault segment, or

$$\bar{D} = \frac{P}{S} \tag{8.71}$$

and

$$S = \left(\frac{P}{\Delta\sigma}\right)^{2/3} \tag{8.72}$$

So

$$D \sim P^{1/3}\Delta\sigma^{2/3} \tag{8.73}$$

So

$$\begin{aligned} \log U_{near\&lowfreq} &\sim \log \bar{D} \sim \log P^{1/3} + \log \Delta\sigma^{2/3} \\ &\sim {}^{1}\!/_{2}\, M + {}^{2}\!/_{3}\log\Delta\sigma \end{aligned} \tag{8.74}$$

Relationship 8.74 can be compared with Figure 5 that is derived from the ground motions and stress drop models

$$\log pgd_{near} \sim 0.35M + 0.29\log\Delta\sigma \tag{8.75}$$

This can be compared with the 10-s response spectral acceleration, sa_{10} , scaling reported by Campbell and Borzorgnia (2014).

$$Log(sa_{10}) \sim 0.333M \tag{8.76}$$

Cua and Heaton (2009) also produced a gmpe for pgd using data from both small- and large-magnitude earthquakes (see Figure 8.18). Their near-source pgd scales as

$$\log(pgd) \sim \tfrac{1}{3} M \qquad 8.77$$

Interestingly, Campbell and Borzorgnia (2007, PEER report) recognized that sa_{10} should approximately mimic the scaling of pgd, and that near-source pgd is controlled by average slip, which in turn, scales with $\Delta\sigma$. However, they state that "the resulting PGD ground motion model is intended for evaluation purposes only at this time and should not be used for engineering design until further empirical and theoretical verification becomes available." That is, the requirement that long-period near-source motions should scale with slip somehow seems to have never made it into the NGA2 gmpe's that were used to calculate the National Probabilistic Hazard Models. **It seems quite clear that $\Delta\sigma$ has a 1st order effect on near-source long-period motions. $\Delta\sigma$ should be included in the gmpe's for long-period motions. The current practice of only using magnitude and observer distance to predict long-period motions is flawed and is likely to seriously under predict the maximum motions from a suite of plausible events.**

As an example of how these relations work, consider the implications for pga and pgd for a large strike-slip earthquake similar to the 1906 San Francisco earthquake. This event has been assigned a moment magnitude of M=7.8. If we were to assume that the rupture dimensions in a future earthquake were the same, but that the slip in the event was doubled, then the Potency and the stress drop would both double. Doubling the potency increases the moment magnitude by 0.2 units and inserting these values into 8.75 gives a 0.3 unit increase in log u, which corresponds to a doubling of u.

Figure 8.13 clearly shows how differently near-source pga scales with size than does pgd. This figure is from Yamada, Heaton, and Olsen (2009) and it shows a log-log plot of pga vs. pgd for every strong motion record for earthquakes of M>6 that was available at the time of the study. The light grey points are for sites at JB distances > 10 km, whereas the bold circles are for JB distances<10 km. Although there is a lot of scatter in in the relationship between pga and pgd, it is obvious that pga and pgd have a linear correlation for distances>10 km, whereas there is absolutely no correlation for distances < 10 km. This observation is compatible with the hypothesis that near-source pga is unrelated to stress drop.

Figure 8.14 is also from Yamada and other (2009) and it shows the frequency distribution of near-source pga. This figure clearly shows that near-source pga for M > 6 is approximately log normal with a geometric mean of 4.64 m/s2. The

grey line shows the distribution without including a large number of records recorded in the 1999 M 7.6 Chi-Chi, Taiwan earthquake, whereas the bold solid line includes Chi-Chi data. Notice that the distribution of near-source pga for smaller previous events seems to fit the data that was observed in Chi-Chi. Again, this is convincing evidence that near-source pga completely saturates with magnitudes greater than 6.

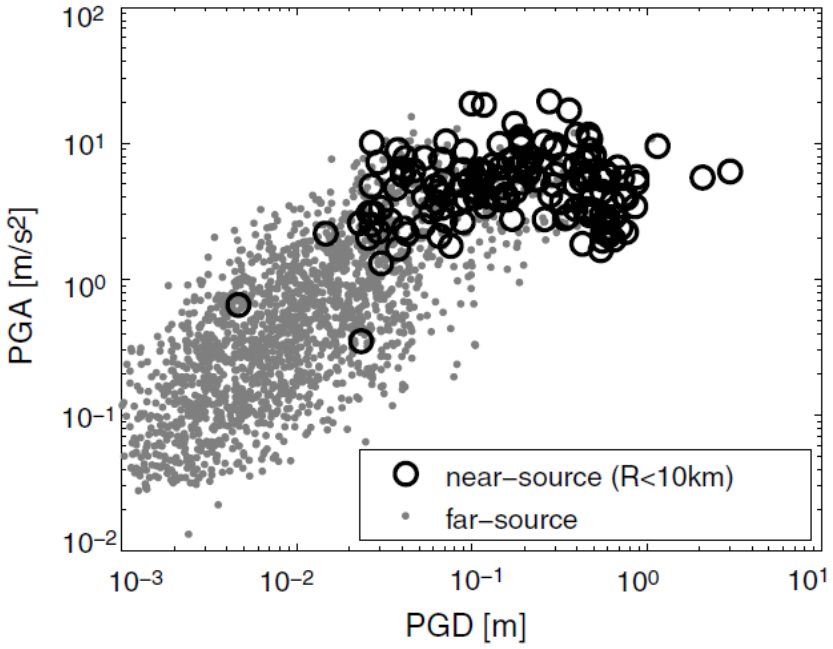


Figure 8.13. Pga vs. pgd for all strong motion records available as of 2008 (from Yamada, Heaton, and Olsen, 2009.

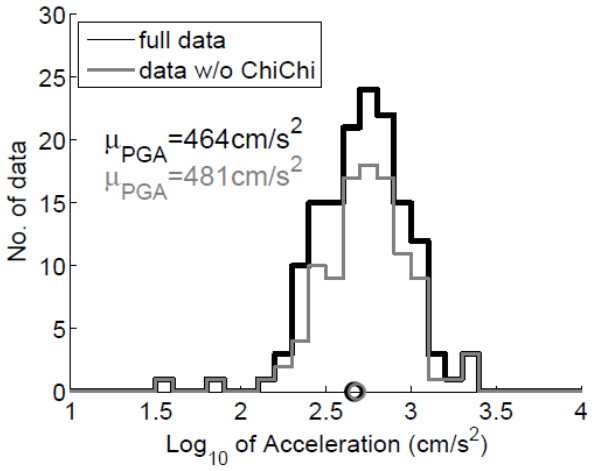


Fig. 8.14. Histogram of the number of near-source records as a function of log pga for earthquakes M.6. The figure is also from Yamada, Heaton, and Olsen (2009).

The saturation of near-source pga can also be clearly seen in the ground motion prediction equations (gmpe's) of Cua and Heaton (2009, *Characterizing Average Properties of Southern California Ground Motion Amplitudes and Envelopes.* Earthquake Engineering Research Laboratory, Earthquake Engineering Research Laboratory , Pasadena, CA. EERL report 2009-05). These gmpe's were created for use in earthquake early warning systems that require the ability to predict over a very wide range of magnitudes. Most gmpe's in engineering seismology are intended for use in predicting damaging motions (M > 5). The Cua and Heaton gmpe was created using a large data set recorded by the Southern California Seismic Network. In addition, strong motion data compiled by PEER was also included. The solid lines are gmpe's derived using only sites on rock, whereas the dotted red line was derived using only data from soil sites.

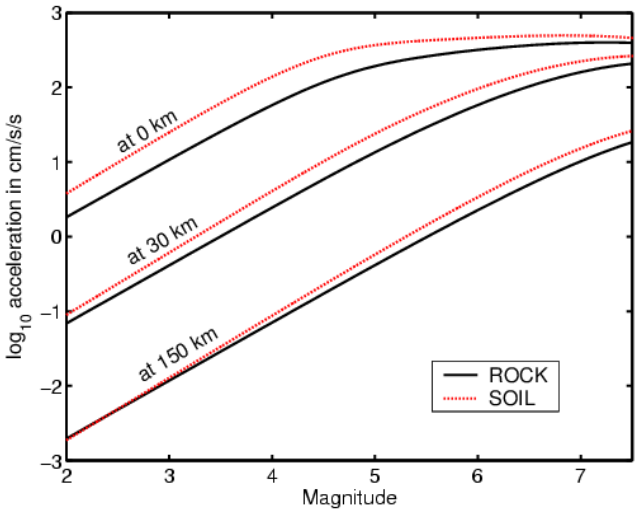


Figure 8.15. Relationship between pga and magnitude as observed at different Joyner-Boore site distances. From Cua and Heaton (2009)

Notice that the 0-distance gmpe seems to indicate that pga saturates for M>5 and the saturation level is consistent with Fig. 8.14. This saturation becomes less apparent for larger distances. This is most likely caused by the fact that near-source pga is primarily the result of frequencies greater than 3 Hz, whereas the important frequencies become lower the further away that you are. Similar plots are shown for pgv and pgd in Figures 8.16 and 8.17, respectively. Notice that log(pgd) scales approximately linearly with magnitude for near-source and M < 5, whereas it scales as approximately $\frac{1}{2}M$ for M > 5; this is consistent with 8.75. The linear scaling of log(pgd) with M seen at either smaller magnitudes or larger distances is consistent with 8.12.

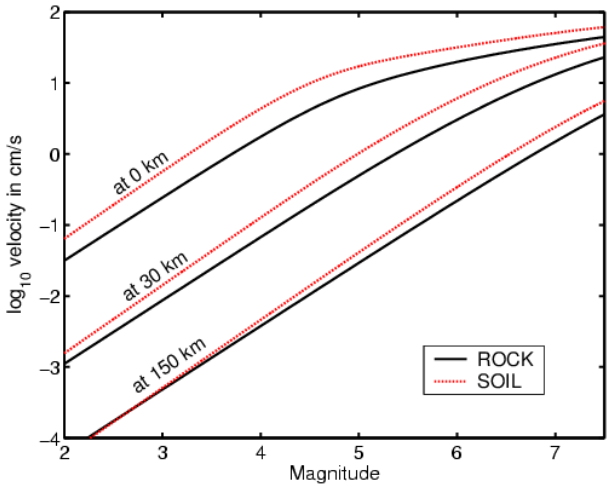


Figure 8.16 Same as 8.15, except for pgv.

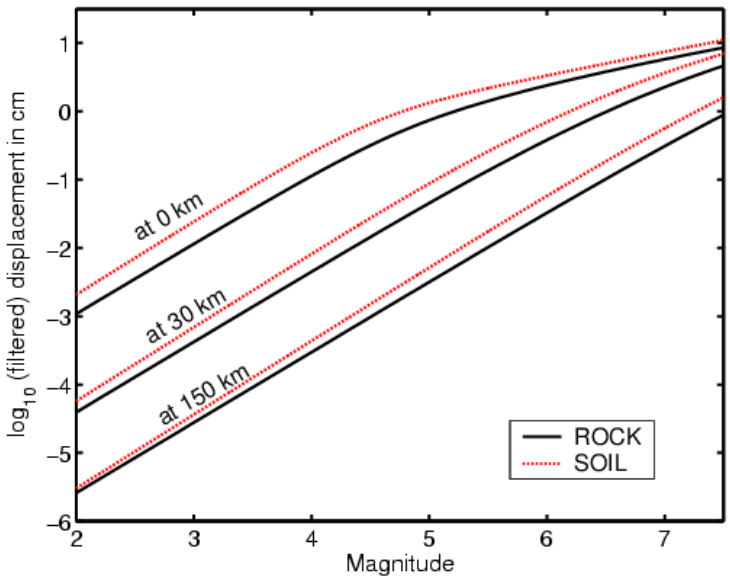


Figure 8.17 Same as 8.15, except for pgd.

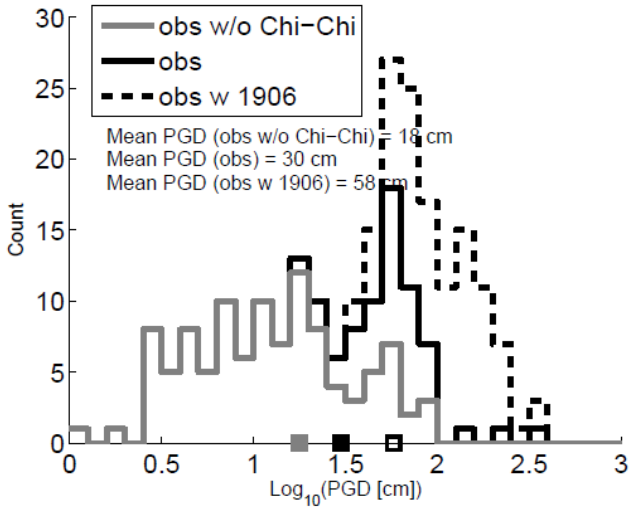


Fig. 8.18. Same as Fig. 8.14, except for peak ground displacement (from Yamada, Olsen, and Heaton, 2009). The solid lines are the distributions with and without Chi-Chi data. The dotted line includes synthetic data at existing strong motion stations for a re-enactment of the 1906 San Francisco earthquake. The analysis in equation 8.24 suggest that this distribution may be log-uniform.

Figure 8.18 shows the number of strong motion records vs. log (pgd) as of 2009. Notice the contrast with log (pga) (see Fig. 8.15). Whereas the log (pga) data

appeared to be log normally distributed about ½ g, the log(pgd) data is clearly not a normal distribution. As was argued in eq. 8.24, this distribution may be **log uniform**. The smaller pdg's are mostly from relatively frequent earthquakes of 6 < M < 7. Notice that including the data from just several large earthquakes (i.e., Chi-Chi and a re-enactment of 1906 San Francisco) can significantly alter this distribution, including the mean. Figure 8.18 is an example of a heavy-tailed power law distribution. Later I will present a more extensive development of the physics associated with probabilities that are defined by a power law. A common feature of power-law distributed data is that the mean of the data typically increases as more data is added; this is certainly the case for the data set shown in Figure 8.18.

A log uniform distribution is an example of **Benford's law**, which is also sometimes referred to as the **first-digit law**. Benford's law was reported in 1938 and it is based on the observation that the first digit of numbers in tables of physical constants are often distributed logarithmically; numbers that start with the digit 1 are much more common than numbers that start with a 9. The formal definition of Benford's law is

$$\begin{aligned} P(N) &= Log(N+1) - \log(N) \\ &= \log\left(\frac{N+1}{N}\right) = \log\left(1 + 1/N\right) \end{aligned} \qquad 8.78$$

Where N is the first digit of the numbers in tables of physical measurements. To derive that this is, in fact, a power law, we can begin by assuming that $P(N) = 1/N$, then the probability of digits between N and $N+1$ is

$$P(N) = \int_N^{N+1} \frac{dN}{N} = \int_N^{N+1} d(\log N) = \log\left(\frac{N+1}{N}\right) \qquad 8.79$$

which is Benford's law as given by 8.78.

It may seem surprising that pga, which is a measure of high-frequency motions, is more or less independent of both $\mathbf{M}$ and $\Delta\sigma$. However, as I will now show, saturation of high-frequency near-source motions with $\mathbf{M}$ is compatible with the well-known Brune (1970) spectral scaling model. **Unfortunately, it seems that Brune scaling is fundamentally inconsistent with the observation that near-source pga's that are independent of** $\Delta\sigma$. In the next section, I will show that it is not possible to formulate a simple spectral scaling law (single power law with

one corner frequency) that has near-source pga that is independent of magnitude and which is also independent of stress drop.

In the variation of the Brune model where the static stress drop is assumed to be the effective stress, stress drop becomes the fundamental parameter that links statics and dynamics. However, many of the kinematic rupture models used in the Yamada and Heaton study actually reported slip pulse models. That is, the slip is already known at most points long before the rupture has stopped propagating. In slip pulse models, there is no correspondence between effective stress and stress drop. Two ruptures could be identical up to the point that one stopped and the other continued. The radiation from the ruptures would also be identical up to that point. However, since the rupture that continued would have a longer rupture length, it would have a different stress drop. **Stress drop is not a fundamental parameter in slip pulse models**; it's just a statistical accident that some ruptures continue on for long distances and have a lower stress drop.

In his 1970 paper, Brune assumed that the slip velocities (and accelerations) scale linearly with effective stress. He also hypothesized that near-source peak accelerations (pga) should scale linearly with effective stress. When using the Brune model, it is common to assume that effective stress is equal to stress drop. This logical construction is self-consistent only if 1) near-source pga's saturate with magnitude (they do), and 2) near-source pga's also scale linearly with stress drop (they seem to be independent of stress drop).

The conclusion of this discussion is that 1) the 1970 Brune model is incompatible with observed seismograms, and 2) there is no scale-free model similar to Aki's conjecture that fits available data.

Although Aki's similarity conjecture is appealing, the most serious issue is that in order to maintain constant stress drop in crack-like ruptures, the fracture energy must grow as L (see 8.35). Clearly this violates the basic assumption that material properties are independent of scale.

Brune's spectral model has been an appealing and powerful conceptual framework to interpret seismic data. This framework consists of: 1) fit power-law envelopes to the Fourier amplitude spectrum , and 2) then derive the dynamic stresses in an earthquake. **Unfortunately, there are many major problems with Brune's approach.** I have already mentioned many of these problems in the previous discussion. I now gather these issues together in one list.

Problems with the Brune Spectral Model

I now list a number of specific objections that I have discovered. The first four objections are mistakes in continuum mechanics that are included in Brune's paper. The next five objections are descriptions of how earthquake observations are inconsistent with the Brune model.

1. *Near-field energy vs far-field energy*
   Perhaps the most fundamental error in the Brune paper is the use of the solution to an instantaneous traction applied to the surface of a half-space (equation 8.58) to infer f^{-2} spectral amplitudes for the far-field motion (equation 8.62). It's true that 8.58 is the solution to an instantaneous step in shear traction on the boundary of a half-space, but this solution is entirely comprised of near-field terms (see Chapter 7). In fact, the assumption of an infinitely large rupture surface means that there is no far-field radiation for this somewhat pathological problem. While it is true that energy is conserved for radiated far-field energy (except for anelastic attenuation), there is no such energy conservation between near-source ground motions (comprised of both near- and far-field terms) and far-field radiation (see Chapter 7). Also, recall from Chapter 7 that far-field terms are the time derivative of the near-field terms. That is, if the motion is a ramp in the near field (f^{-2}), then it is a step in the far field (f^{-1}).

2. *Infinite rupture velocity is impossible*
   Brune argued that "*the particle velocities are always much smaller than the rupture propagation velocity (since the stresses are much less than the shear modulus), and thus the interaction of particle motion velocity with the rupture velocity will always be small, i.e., the average energy spectrum for instantaneous application of stress will be approximately valid.*" Later analysis showed that the ratio of the rupture velocity to the shear wave speed is of interest to mode III ruptures, and the ratio of the rupture velocity to the Rayleigh wave speed is of interest for mode II ruptures. Since typical observations of rupture velocities are in the range of 0.8β to 0.9β , the approximation of $V_R \approx \infty$ is clearly inappropriate. In fact, rupture velocity plays a key role in almost all models of dynamic rupture.

3. *Ramp-like slip produces f^{-3} far-field radiation.*
   This objection is related to objection 1. In Chapter 7, I demonstrated that a line source with an instantaneous step in slip produces a rectangular far-field time function, which has a f^{-1} high-frequency spectrum. A rectangular source with an instantaneous slip has a far-field time function that is the convolution of two rectangles, which is a trapezoid that has a high-frequency decay of f^{-2} (one for timing associated with length and the other with

width). A rectangular source with fault slip that is a linear ramp has a far-field time function that is the convolution of three rectangles (f^{-3}). The third rectangle is associated with Brune's slip function. The solution describing the displacements from an instantaneous change in stress on a circular fault (the problem Brune claimed to be solving) is actually a difficult mechanics problem and I am not aware of any closed-form solutions. Madariaga (1976, Dynamics of an Expanding Circular Fault, Bull Seisms Soc. Am, 66, 639-667) used finite elements to simulate this problem and he concluded that the far-field radiated S-waves have a high-frequency spectral decay of $f^{-2.5}$. In Madariaga's study, most of the high frequencies were radiated at the circumference of the rupture.

4. *The Brune model is a uniform stress drop crack model.*
I have already mentioned this important issue several times. Constant stress drop cracks produce stress concentrations at the crack tips that grow linearly with the rupture dimension. These models are not scale independent.

5. *Slip is not ramp-like*.
Ruptures that propagate at sub-shear rupture velocities produce slip functions that are similar to a time-transformed version of slip on a crack. The slip near the tip of a static shear crack increases as $\Delta\sigma K\sqrt{x}$, where x is the distance from the crack tip and K is a stress intensity factor. For a steadily propagating crack, x transforms to $x - V_R t$. This means that for any fixed point on the rupture, $D \sim \Delta\sigma K\sqrt{t}$. To obtain the far-field time function, you need to differentiate to slip velocity, or $\dot{D} \sim \Delta\sigma K / \sqrt{t}$. Convolving with $1/\sqrt{t}$ is equivalent to multiplying with $f^{-1/2}$. This is sometimes called a **fractional integral**.

6. *Predicted slip durations much longer than observations*.
The Brune model predicts that slip in the near-source region is given by 8.59, which predicts that slip continues until shear waves arrive from the outer edge of the rupture. That is, the slip at any point is comparable to the entire duration of an earthquake. As I will describe shortly, there is convincing evidence that the duration of slip at a point is short (less than 10%) compared to the overall rupture time; this is known as a slip-pulse model.

7. *Real earthquake far-field time functions are much more complex than Brune's model.*

Brune's far-field time function (8.63) predicts a relatively smooth single pulse (see Figure 8.19). In contrast, observed far-field time functions are typically very complex. See the far-field time functions shown in Figure 8.7 as an example.

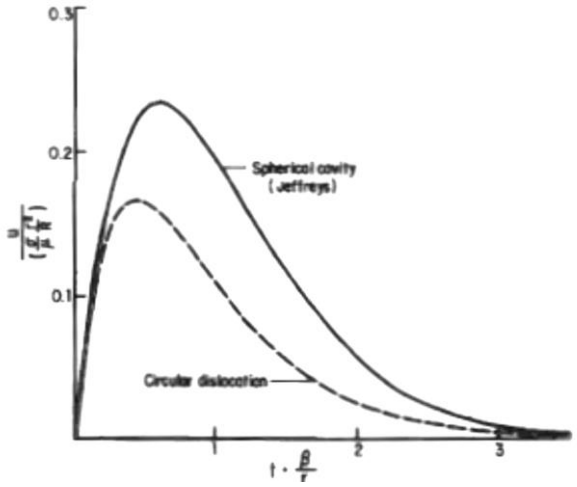


Figure 8.19 (from Brune, 1970) This is the time behavior of Brune's far-field time function. It has the same ramp-like beginning as his near-source ground motion which gives it a f^{-2} spectral decay.

Hartzell and Heaton (1093, Bull. Seism. Soc. Am. 73, 1553–1583) compiled far-field time functions from large earthquakes. More recently, Meier, Ampuero, and Heaton, (2017, *The hidden simplicity of subduction megathrust earthquakes.* Science, 357 (6357). pp. 1277-1281. ISSN 0036-8075) investigated characteristics of a compilation of time functions for large earthquakes and these are shown in Figure 8.20. The time functions are scaled to have the same total durations. The red curve at the bottom is the median of all of the time functions and it has a simple triangular form. However, the individual time functions are very complex and do not resemble Brune's time function.

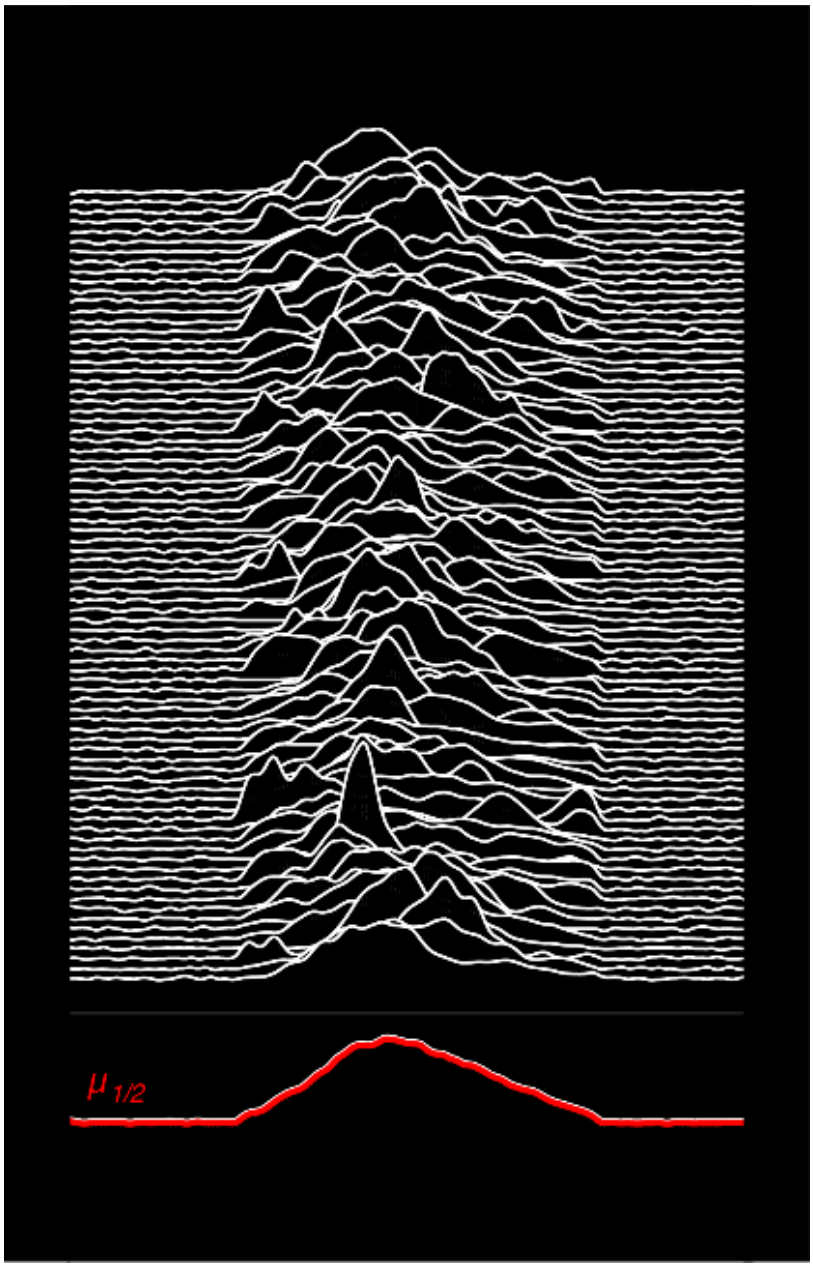


Figure 8.20. Teleseismic time functions from large subduction earthquakes. Notice the temporal complexity which contrasts with Brune's time function shown in Figure 8.7. The red line at the bottom is the median of all of the functions above it. This is an unpublished figure from the study of Meier, Ampuero, and Heaton (2017).

8. *Observed far-field spectral decays of large earthquakes (M > 6) are close to* $f^{-1.5}$.

   Although it is widely assumed that far-field radiated energy is characterized by f^{-2} spectral decay, recent studies using far more data than was available in the 1960's indicate that the average spectral decay between f_c and 0.5 Hz (the highest frequency observable at Teleseismic distances) is actually $f^{-1.5}$ as shown in Figure 8.21 from Hartzell and Heaton (1988, *Failure of self-similarity for large (Mw > 8 1/4) earthquakes.* Bulletin of the Seismological Society of America, 78 (2). pp. 478-488. ISSN 0037-1106). Brune's conjecture that Fourier amplitude spectra of near-source high-frequency motions (frequency band from 2 Hz to 15 Hz) is approximately f^{-2} seems to be consistent with a very large volume of strong motion data. As I will discuss later, the physics of near-source radiation of high frequencies is still an unsolved problem and Brune's model is not a plausible explanation of the observation that near-source acceleration look like Gaussian white noise.

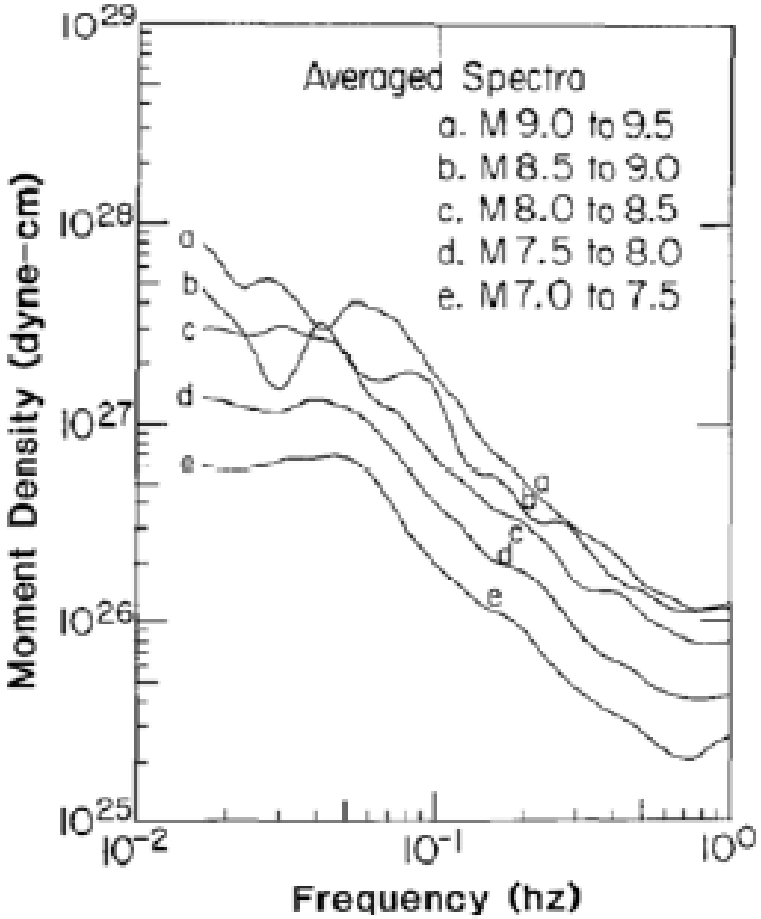


Figure 8.21. These are the Fourier amplitude spectra of Teleseismic P-waves from large subduction earthquakes. The average spectral decay is $f^{-1.5}$ for the frequency band between 0.5 Hz and 0.08 Hz. Anelastic attenuation prevents study at frequencies higher than 0.5 Hz using Teleseismic records. From Hartzell and Heaton (1988)

9. There is no correlation between stress drop and pga

Brune hypothesized that the spectral amplitude of near-source accelerations should scale linearly with effective stress, σ_e, which is equal to stress drop in the most commonly used form of Brune's model . This is an important issue since current predictions of strong shaking often use variations in regional stress drops to predict corresponding variations in high-frequency shaking. More fundamentally, I show in Appendix A that there is no

possible power law with a single corner frequency that produces bear-source pga's that both saturate with magnitude and which are also independent of stress drop.

Simulating ground motions using records from smaller earthquakes.

As much as I dislike the Standard Source model, it is simple and easy to use. For example, it allows us to use the record, $u_1(t)$, from a smaller potency magnitude M_1 earthquake to simulate the motion, $u_2(t)$, from a larger M_2 earthquake. The expected ratio of the amplitude spectra of the two events can be estimated from 8.65 .

$$\begin{aligned}\frac{|\tilde{u}_2(f)|}{|\tilde{u}_1(f)|} &\sim \frac{P_2}{\left(1+\left(\frac{f}{f_{c2}}\right)^2\right)}\left(\frac{1+\left(\frac{f}{f_{c1}}\right)^2}{P_1}\right) \\ &= \left(\frac{10^{\frac{3}{2}M_2}}{1+\left(\frac{f}{f_{c2}}\right)^2}\right)\left(\frac{1+\left(\frac{f}{f_{c1}}\right)^2}{10^{\frac{3}{2}M_1}}\right) \\ &= 10^{\frac{3}{2}(M_2-M_1)}\frac{1+\left(\frac{f}{f_{c1}}\right)^2}{1+\left(\frac{f}{f_{c2}}\right)^2} \\ &= 10^{\frac{3}{2}(M_2-M_1)}\left(\frac{f_{c1}^2+f^2}{f_{c2}^2+f^2}\right)\left(\frac{f_{c2}^2}{f_{c1}^2}\right)\end{aligned} \qquad 8.80$$

Assuming stress drop invariance (that is $\Delta\sigma_1 \approx \Delta\sigma_2$), we can use 8.54 to estimate the corner frequencies of the two events. Or,

$$
\begin{aligned}
f_c &\approx \left(\frac{V_R^3}{20\mu} \frac{\overline{\Delta\sigma}}{P} \right)^{1/3} \\
&\approx \left(\frac{\left(2.8\times10^3 \, m/s\right)^3 2.7\times10^6 \, Pa}{7\times10^{10} \, Pa \times 10^{3/2 M - 1} \, m^3} \right)^{1/3} \\
&= \left(\frac{2.2\times10^{10} \left(m/s\right)^3 2.7\times10^6}{7\times10^9 \times 10^{3/2 M} \, m^3} \right)^{1/3} \\
&= \left(\frac{8.5\times10^7}{10^{3/2 M} \, s^3} \right)^{1/3} = 440\times10^{-1/2 M} \\
&= 10^{2.64 - 1/2 M} \, s^{-1}
\end{aligned}
\qquad 8.81
$$

Substituting 8.81 into 8.80 gives

$$
\begin{aligned}
\frac{\left|\tilde{u}_2(f)\right|}{\left|\tilde{u}_1(f)\right|} &\approx 10^{3/2 (M_2 - M_1)} \left(\frac{f_{c1}^2 + f^2}{f_{c2}^2 + f^2} \right) \left(\frac{f_{c2}^2}{f_{c1}^2} \right) \\
&\approx 10^{3/2 (M_2 - M_1)} \left(\frac{10^{3.3 - M_1} + f^2}{10^{3.3 - M_2} + f^2} \right) \frac{10^{3.3 - M_2}}{10^{3.3 - M_1}} \\
&= 10^{1/2 (M_2 - M_1)} \left(\frac{10^{3.3 - M_1} + f^2}{10^{3.3 - M_2} + f^2} \right)
\end{aligned}
\qquad 8.82
$$

Therefore, if we have recorded the motion $u_1(t)$ for a magnitude M_1, then we can obtain an estimate of the motion expected at the same site from a M_2 event at the same source location using the following approximation

$$
u_2(t) \approx 10^{1/2 (M_2 - M_1)} FT^{-1} \left[\tilde{u}_1(f) \left(\frac{10^{3.3 - M_1} + f^2}{10^{3.3 - M_2} + f^2} \right) \right] \qquad 8.83
$$

This approximation assumes that the observing distance is large compared to the source dimensions; that is, both events are approximately point sources. This requirement can be relaxed by assuming that a large rupture is

approximately described by the appropriate sum of smaller ruptures. The responses of the smaller ruptures are assumed to describe the response of different parts of the finite fault. This technique is referred to as the **Empirical Green's Function** technique. I will not go through the details here (it's considerably more complex than what I've just derived), but it is discussed in detail in Appendix B of Heaton and Hartzell, 1989, Estimation of strong ground motions from hypothetical earthquakes on the Cascadia subduction zone, Pacific Northwest, Pure and Applied Geophysics, 129, 131-201.

Self-Healing Slip Pulses

Early in my research career, I investigated the physics of strong ground motions from significant earthquakes. In the 1970's this was a new field of study. Although there had already been strong motion recordings of several earthquakes, the 1971 M 6.7 San Fernando earthquake was the first earthquake with dozens of near-source strong motion records. Furthermore, there was an extensive project at Caltech to digitize the analog film accelerometer records and to integrate these records into ground velocity and displacement. One of the most significant records was from the Pacoima dam site that was directly up-dip from the hypocenter. This record had two key features of great interest to earthquake engineering (see Figure 8.22). The first was a peak acceleration of 1.15 g that occurred relatively late in the record. The second key feature was a distinct velocity pulse with an amplitude of 1.25 m/s. The acceleration records were too high-frequency to be interpreted using a deterministic model of rupture on a finite fault. However, the ground velocity pulse was very different. I interpreted it as a far-field S-wave that was strongly enhanced by directivity. Rather remarkably, the duration of this pulse was less than 1 s. Since the duration of the S-wave from slip on any point on the rupture is the duration of the slip at that point, the implication is that the duration of slip would need to be less than 1 second as well. When the rupture is propagating towards the receiver at close to the S-wave velocity, then all of the far-field S-waves arrive simultaneously, thereby producing this velocity pulse.

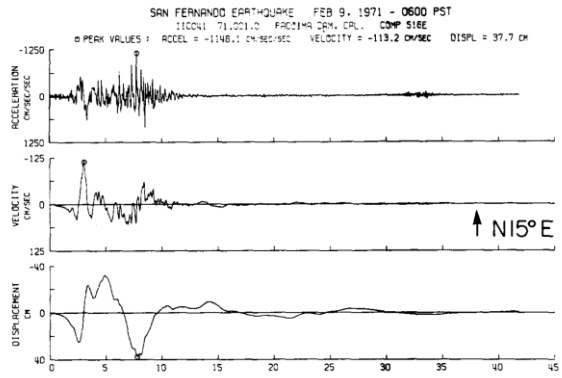


Figure 8.22. North component of ground motion from the Pacoima Dam recording of the 1971 San Fernando earthquake. The strong velocity pulse at 3 seconds is interpreted as a directivity enhanced far-field S-wave. Modeling of this pulse indicates that the duration of slip was less than 1 second at individual points on the 15-km length rupture, which took more than 5 seconds to occur.

One of my colleagues, Steven Hartzell, was also working in this same research direction and we collaborated on constructing finite-fault models to explain the near-source strong motion data from a number of earthquakes. All of these models shared the common feature that the duration of slip at any point was short compared to the time it took for the rupture to propagate over the rupture surface. For example, Figure 8.23 shows our preferred rupture model of the M 6.5 1970 Imperial Valley Earthquake. The hatched region shows our estimate of the part of the fault that was rupturing at a particular instant.

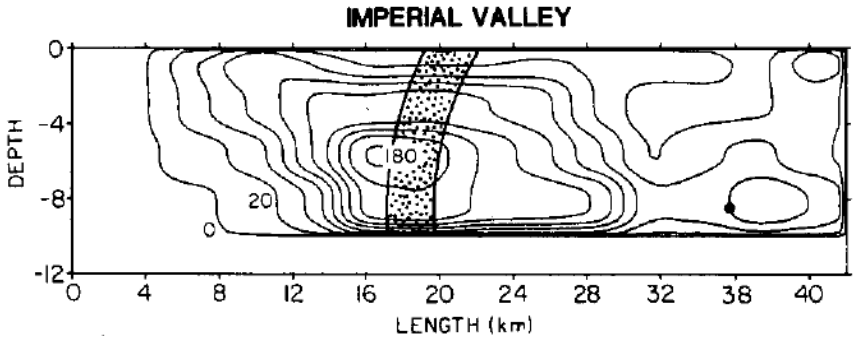


Fig. 8.23. Final slip distribution used to match strong motion and teleseismic data for the 1979 Imperial Valley Earthquake. The stippled region shows the area inferred to be slipping at just one instance in time (from Heaton, 1990).

It was only later in 1988 that I realized that our rupture models were fundamentally inconsistent with crack-like models, which were the prevailing models for earthquake dynamics. In crack-like models, slip at a point starts when the rupture arrives and then continues until S-waves arrive from the boundary of the rupture.

Fig. 8.24 shows the slip histories for a uniform stress drop on a circular fault embedded in a homogeneous whole space (from Madariaga, 1978). The rupture expands radially from the center of the circle with a rupture velocity of 0.9β. Notice that the slip at every point continues until a shear wave arrives from the circumference of the fault (sometimes referred to as a stopping phase). This shear wave contains the information that causes points on the fault to stop slipping. In general, uniform stress drop ruptures in which the slip duration is determined by the end of the rupture have slip durations that are on the order of $2\sqrt{S}/3V_R$. These long slip durations contrast sharply with much shorter slip durations used to model seismic data from significant earthquakes.

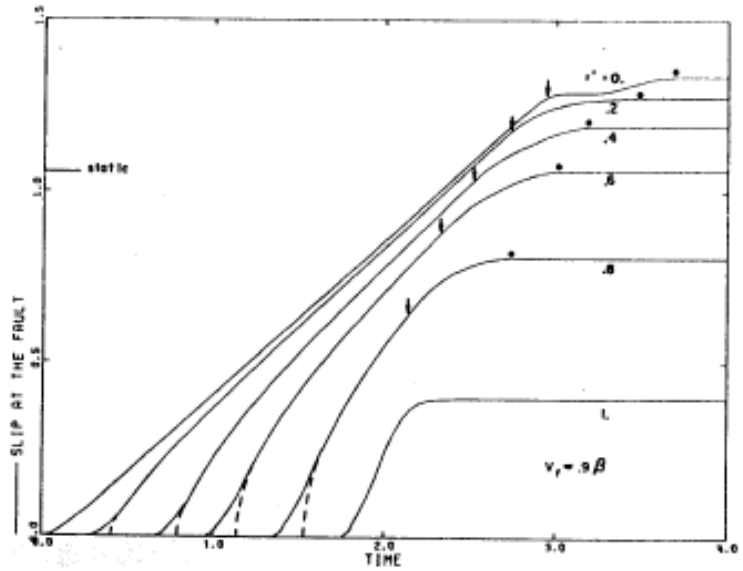


Figure 8.23. Crack-like rupture. Slip histories for a variety of points on a circular fault in a homogeneous whole space. The upper diagonal is near the hypocenter (also the center of the circle) and the lower right history is near the outer edge of the circle. The stress drop is homogeneous and the rupture velocity is 0.9β. The arrows represent the time that a stopping phase comes from the perimeter of the rupture and the dots represent the cessation of rupture. (from Madariaga, 1976).

The primary question was to explain why slipping would stop before the rupture stopped propagating. This issue was the subject of my 1990 paper, Evidence for and implications of self-healing pulses of slip in earthquake rupture (Phys. Of Earth and Planet. Int., 64, 1-20). In addition to the source modeling of seismic records, there were also eyewitness reports of relatively short times for surface fault scarps to develop.

My slip-pulse paper was extremely controversial. In fact, I eventually withdrew it from the BSSA since the reviewers would not accept it without a wide range of fundamental changes. Since the time of the publication, there have been many other well recorded earthquakes and I think it's safe to say that it is now relatively well accepted that earthquake rupture durations at a point are small compared to the overall duration of the rupture.

I coined the phrase self-healing pulses of slip. I suggested that dynamic friction was strongly rate weakening. I argued that, similar to crack models, there is very high shear stress at the propagating crack tip and that this led to very high slip velocities just behind the propagating tip of the pulse. I suggested that the slip velocity decreased as $1/\sqrt{x_p}$, where x_p is the distance from the leading edge of the propagating pulse (see Figure 8.24). I hypothesized that the shear stresses in the vicinity of the crack tip were comparable to Byerlee friction and that the dynamic friction was very low in the high-slip velocity region just behind the rupture front. I hypothesized that as the slip velocity decreased with distance from the crack tip, the dynamic friction increased. I suggested that this increase in dynamic friction caused the slip to arrest.

At the time of my paper, one of the most robust observations of earthquake scaling was the observation that average slip scales linearly with the rupture dimension (e.g., see Figure 8.5). In the slip-pulse model, slip at most points on the fault was already complete long before the final dimension of the rupture was determined. I suggested that there was a causal relationship between the amplitude of a slip pulse and the distance that the pulse was likely to propagate. That is, large earthquakes with long ruptures should have large slips that show up rapidly after the origin of the event. However, further study has shown that the characteristics of the early part of ruptures is not systematically different for larger ruptures.

I have now abandoned the hypothesis that larger earthquakes start with larger slips, and I now favor the notion that the correlation between slip and rupture length comes from the statistical properties of suites of spatially complex slip distributions. I will later argue that the spatial complexity of ruptures is related to spatial complexity of the initial stress. In order to adequately discuss slip pulses, it is necessary to discuss friction and the

larger problem of maximum shear stress in the Earth's Crust. I will get to that shortly, but in the meantime, I have some comments about my hypothesis of strong rate weakening friction.

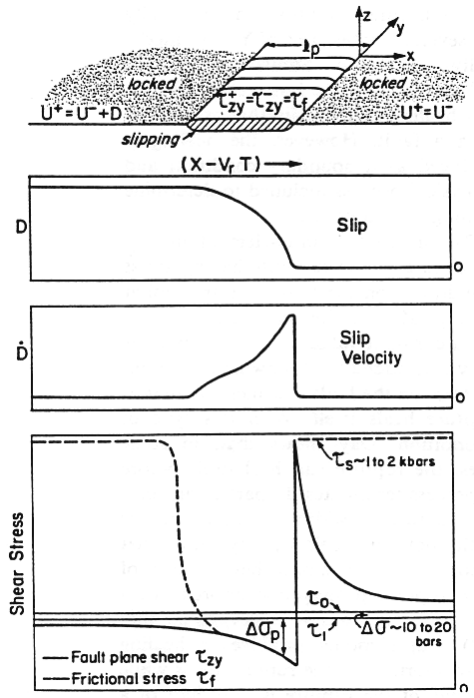


Fig.8.24. Sketch from Heaton (1990) of a slip-pulse rupture mode. I proposed that fault friction decreased with slip rate. This implied that dynamic friction is low just after the passage of a rupture. The dotted line is meant to show the friction necessary for sliding. This friction is comparable to standard friction (60% of the confining stress) when the fault is not sliding. This model has instantaneous restrengthening.

When I first suggested strong rate weakening friction, I discussed the situation with SDSU Prof. Steven Day who was a leading expert in finite element simulations of dynamic shear rupture simulation. Steve was intrigued by the hypothesis, and he attempted to simulate slip-pulse ruptures that arose from velocity weakening friction. He reported to me that he found it difficult to tune his friction and prestress in such a way that it would produce stably propagating slip pulses. In particular, they tended to either grow without bound as they propagated, or alternatively, they tended to die out fairly rapidly without propagating very far.

In retrospect, his finding was prescient, and should have been anticipated. That is, a velocity weakening friction law produces a positive-feedback dynamic system. In particular, the faster the sliding, the lower the friction. Lower friction produces

faster sliding which leads to even lower friction. Conversely, slow sliding produces high friction, which causes even slower sliding. In fact, it has been shown that while strong velocity weakening can produce slip pulses, the pulses are inherently unstable. That is, there are no steady-state slip pulses caused by strong rate weakening friction (Elbanna, Lapusta, Heaton).

Since there are no stable steady-state solutions to strong rate-weakening slip pulses, they can only be studied using numerical techniques, such as finite elements or boundary-integral-elements. The fact that there are strong positive feedbacks in the dynamics means that the numerical solutions are sensitive to the details of the rupture dynamics, which means that high precision is required to make reliable numerical simulations.

Before I can discuss the dynamics of slip pulses, it is important to understand the characteristics of shear stress in the Earth's crust. I will shortly describe how understanding the characteristics of chaotic dynamics is key to understanding how it's possible to have 1) a static coefficient of friction of about 0.6, 2) low crustal heat flow and a paucity of frictional melts, 3) average static stress drops of about 1.5 MPa (1% of the overburden pressure), 4) radiated energy of about $2P\,\mathrm{MJ}/\mathrm{m}^3$.

ESTIMATING SHEAR STRESS IN THE CRUST

An enduring problem in geophysics focuses on estimating the size of shear stress that causes earthquakes. This problem has been approached in several ways. One approach is to estimate the stress necessary to cause shear failure. This approach typically leads to the conclusion that the frictional stresses at the depth of earthquakes are a substantial fraction of the pressure from the weight of the overburden rocks. As it turns out, these stresses far exceed any of the shear stresses that are estimated using seismic shaking data. This inconsistency has been debated for decades.

I have given numerous lectures about this issue and I often frame the problem as, "why are earthquakes so gentle?" Although this may at first sound like an odd question, it is a rather serious issue when one is asked to estimate the maximum shaking that may occur in an earthquake.

In particular, laboratory test equipment that causes frictional sliding of surfaces that are forced together with the overburden pressure that occurs 10 km deep in the earth (about 200 MPa) are large and powerful machines. In the laboratory, yielding at these stresses can be dangerous even for relatively small samples; for safety, operators typically go to the next room to control these experiments. If

you scale these experiments up to the size of significant earthquakes, then none of us would survive earthquakes, the ground motion would be too violent.

To better describe this problem, I will assume that earthquakes can be understood in terms of an expanding frictional crack. This will allow me to demonstrate fundamental inconsistencies that arise from this view. I will then present an alternate view that describes earthquakes as propagating slip pulses. Unfortunately, there are no analytic dynamic solutions to the slip pulse problem; by its nature, it is fundamentally a chaotic system.

A simple analysis of energy partitioning is appropriate at this point. Noting that an earthquake is the process of removing potential energy, E_W , from the crust and converting it into radiated seismic energy, E_R , frictional heat, E_f , and fracture energy, E_G , conservation of energy dictates that

$$\Delta E_W = E_R + E_f + E_G \tag{8.84}$$

We can express the change in potential energy by integrating the traction on the fault over the slip, or

$$\Delta E_W = \iint_S \left(\int_0^{D(x,y)} \bar{\sigma}(x,y;D)\,dD \right) dx\,dy \tag{8.85}$$

Where $D(x,y)$ is the final slip as a function of position on the fault plane, and $\bar{\sigma} = \frac{1}{2}(\sigma_0 + \sigma_1)$ is the average shear stress between the initial state, $\sigma_0(x,y)$ and the final state $\sigma_1(x,y)$. Remember that the continuum is considered to be linearly elastic except on the slip surface. Therefore, the introduction of a reverse slip from the final state back to the original state would take the final stress back to the original stress. Since the system is linear, the average stress during that transition is the traction that the slip works against.

8.85 can be rewritten as

$$\begin{aligned} \Delta E_W &= \iint_S \left(\int_0^{D(x,y)} \frac{1}{2}\left[\sigma_0(x,y) + \sigma_1(x,y)\right] dD \right) dx\,dy \\ &= \iint_S \left(\int_0^{D(x,y)} \left[\sigma_0(x,y) - \frac{1}{2}\Delta\sigma(x,y) \right] dD \right) dx\,dy \end{aligned} \tag{8.86}$$

Where $\Delta\sigma(x,y) \equiv \sigma_0(x,y) - \sigma_1(x,y)$. At this point, it is tempting to assume that the initial stress and the stress drop are approximately constant as a function of x and y. In fact, Kanamori and Heaton (2002) made this assumption. I will make it now because it helps to provide context for many other researcher's work. However, I warn you strongly that this is a *critical assumption that is almost certainly very wrong*. Because this is a key assumption (and probably wrong), I will introduce a notation for the energy estimated in this way. I will call it ΔE_W^{crack} , which I define as

$$\Delta E_W^{crack} \equiv \iint_S \left(\int_0^{D(x,y)} \left[\overline{\sigma_0} - \frac{1}{2}\overline{\Delta\sigma} \right] dD \right) dxdy \qquad 8.87$$

Where the overbars signify spatial average over the rupture surface. Stress drop and slip are both spatially smooth (except at the crack tip) for crack models, so the approximation is appropriate for cracks. In contrast, the slip in actual earthquakes is typically heterogeneous. Heterogeneous slip implies that prestress and stress drop are also spatially complex. Furthermore, the spatial complexity correlates between these quantities that specify the net energy change caused by an earthquake.

Examination of once-active faults that have been uplifted and exhumed by erosion shows that there is a paucity of frictional melting along rupture surfaces. This observation allows us to provide upper limits on the energy of frictional heating. In addition, seismic data allows us to estimate the radiated energy in earthquakes.

Note that $D(x,y)$ can be inferred from modeling seismic data. Knowing the spatial distribution of slip allow us to also infer $\overline{\Delta\sigma}(x,y)$. Now if we assume that slip amplitude is spatially smooth for crack-like ruptures, then $D^{crack}(x,y) \approx \overline{D}$; note that the averaging in this case is spatial averaging over the rupture surface. Using these assumptions allows us to integrate 8.87 to obtain

$$\begin{aligned} \Delta E_W^{crack} &\approx \iint_S \left(\left[\overline{\sigma_0} - \frac{1}{2}\overline{\Delta\sigma} \right] \overline{D} \right) dxdy \\ &\approx \left[\overline{\sigma_0} - \frac{1}{2}\overline{\Delta\sigma} \right] \overline{D}S \end{aligned} \qquad 8.88$$

We can now estimate the average prestress as

$$\sigma_0^{crack} \approx \frac{1}{2}\overline{\Delta\sigma} + \frac{\Delta E_W^{crack}}{\overline{DS}} \qquad 8.89$$

At this point, we can use 8.84 to estimate $\Delta E_W^{crack} = E_f^{crack} + E_G^{crack} + E_R$. Although we need to assume spatial smoothness of stress and slip to estimate friction and fracture energy, radiated seismic energy is directly inferred from seismograms; assuming spatial smoothness is not necessary for this calculation. Although friction energy cannot be inferred from seismograms, we can nevertheless put upper limits on its size. The same goes for fracture energy.

FRICTION

Sliding friction during an earthquake is often called dynamic friction, σ_f . Friction is a dissipative process; it cannot add energy into the mechanical system. In the case of a uniform crack, $\sigma_f = \sigma_1 = \sigma_0 - \Delta\sigma$. Laboratory measurements of sliding friction for materials found in fault zones typically indicate that $\sigma_f \approx \mu_f \sigma_n$, where μ_f is the coefficient of friction and σ_n is the compressive stress normal to the fault. Most fault materials have measured coefficients of friction between 0.6 and 0.8. The normal stress on the fault is typically thought to be comparable to the pressure from the weight of the rocks above the earthquakes; this is referred to as lithostatic pressure and it is estimated to be $\sigma_n \approx \rho g z$ where z is depth beneath the Earth's surface. Assuming, the average density of crustal rocks (about 2,700 kg per cubic meter), this can be approximated as $\sigma_n \approx z \times \frac{30\text{MPa}}{\text{km}}$. Lithostatic pressure contrasts with hydrostatic pressure (the pressure of water in an open well) which is $\sigma_H \approx z \times \frac{10\text{MPa}}{\text{km}}$. If a material is porous and saturated with water, then the effective pressure is $p_{eff} = \sigma_n - \sigma_H$. The bottom line is that laboratory friction measurements indicate that the friction is on the order of $\sigma_f \approx z \times \frac{20\text{MPa}}{\text{km}}$. That is, at 10 km, we expect that frictional stress is on the order of 200 MPa. This expected frictional stress is much larger than the average measured stress drop of earthquakes (3 MPa).

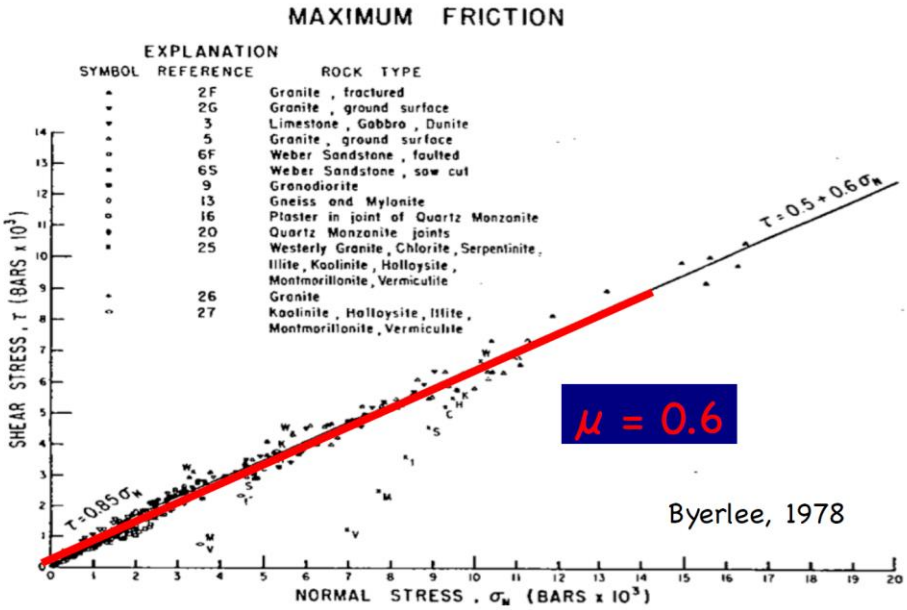


Figure 8.25. Shear stress of frictional yielding of a large range of experiments on different rocks and different normal stresses as reported by Byerlee. Although there are some notable exceptions (especially clay materials), most rocks have an apparent coefficient of friction of about 0.6. Note that the stress is given in units of kilobars (1kilobar= 100MPa).

Figure 8.25 is a famous figure that shows the apparent coefficient of friction for a wide range of Crustal rocks subjected to a large range of confining stresses. This observation leads to one of the most fundamental unanswered questions in geophysics. What is the average shear stress in the seismogenic crust? Is it near the laboratory observations (~ 200 MPa), or is it comparable to the stress drop in earthquakes (~ 3 to 10 MPa)?

One clear observation that crustal rocks can sustain high shear stress comes from the fact that there are deep mines (up to 3 km in South Africa). At 3 km, the overburden stress is about 100 MPa. On the other hand, the interior surface of a mine is traction free. The calculation of stress around a mine is a complex problem, but it is clear that the shear stresses associated with deep mines are at least 50 MPa, which is consistent with Figure 8.21. Deep mines cannot remain open unless there are high shear stresses in the vicinity of the mine faces. Tragically, sometimes these faces fail violently in a phenomenon known as rock bursts. Rock bursts kill many miners in the deep gold mines of South Africa.

Figure 8.26 shows a popular view of the strength of the Earth's crust. This cartoon is often called the "Christmas tree." The maximum shear stress before frictional sliding is often called the "Byerly limit;" when you hear this, it is

intended to refer to the Christmas tree strength model. I have often heard earth scientists describe the shallow Crust as both brittle and strong, but deep Crust as ductile and weak. If you remember our discussions about the properties of buildings (Chapter 6), you should be confused that brittle is used together with the word strong. Remember that in structural engineering, the goal is to avoid brittleness since local flaws can cause global collapse. Most structural engineers try to make their buildings ductile.

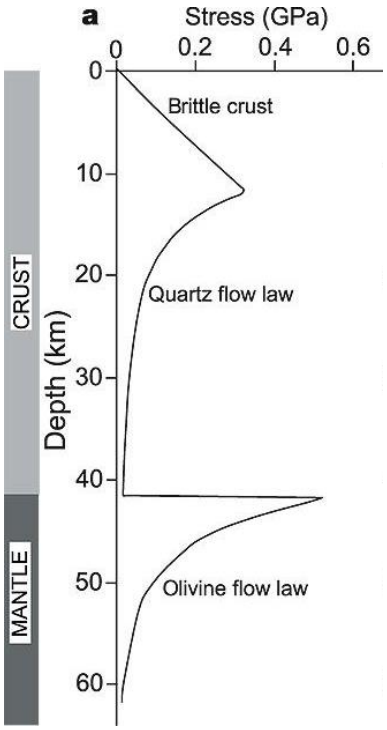


Figure 8.26 shows a simplified picture of the "strength" of the crust and is often called the "Christmas tree." The linear increase in the top 10 km assumes a constant coefficient of friction taken together with the assumption that pressure increases linearly with depth. The decreased stress in the lower crust is assumed to be the result of ductile yielding of rocks at high temperatures present in the lower crust.

Interestingly, mechanical engineers who work on fracture mechanics often describe the strength of brittle materials with either the fracture energy, or alternatively the stress intensity factor, $K_{III} = \sqrt{2\mu G}$. Note that fracture energy has units of $\frac{\text{force}\times\text{length}}{\text{rupture area}}$, whereas yield stress has units of $\frac{\text{force}}{\text{rupture area}}$. That

is, these two different descriptions of strength differ in units by a length scale. Interestingly, brittle materials typically fail on thin surfaces (length scale of L^2), whereas failure in ductile materials is typically distributed in a volume (length scale of L^3). In many real-world cases, faults are complex at all length scales (apparently fractal). That is we can anticipate that the complexity of fault surfaces implies that we will have to resort to descriptions of strength that are somewhere between brittle and ductile. These laws will have units of stress times length to a fractal power. If the power is zero, L^0 , then the material will have a simple yield stress (like a plastic material). If the power is 1, L^1 , then the material is brittle and its strength is described with a fracture energy. That is, we can already anticipate that there will be length dependencies to strength and that spatial complexity will be part of the story.

Friction Laws

Earthquakes appear to be due to shear sliding along fault planes; that is, inelastic strains associated with earthquakes occur in very narrow regions, which are often simplified as a planar interface. Although the inelastic strains almost certainly occur over a finite width, this width may be very small compared to other rupture dimensions. In order to construct simple models of fault rupture dynamics, we often assume that there is indeed slipping on a plane and that the shear stress on the plane is governed by a friction law. If the shear traction on a fault plane is less than the sliding friction, then the boundary is assumed to be a welded contact for which all displacements are assumed to be continuous across the fault plane. As long as the amplitude of stress is less than the friction law, the stresses on the fault plane are described by the standard laws of continuum mechanics (Chapter 3). However, the stresses cannot exceed the friction law; once the deformations are large enough, then the displacements on the fault are no longer continuous across the fault and the shear tractions on the fault are assumed to be equal to a friction law. These friction traction vectors are always assumed parallel to fault slip velocity, but in the opposite direction. The simplest possible friction law is just an instantaneous jump in friction from μ_1 to μ_0 . In this case, if the shear traction rises to $\mu_0\sigma_n$, then the shear traction instantaneously drops to $\mu_1\sigma_n$ and there is an instantaneous stress drop of $\Delta\sigma = \sigma_n\left(\mu_0 - \mu_1\right)$, that is, assuming that the normal stress is constant. This was the assumption that Brune made in his 1970 paper.

Most observations of friction show a linear increase in frictional traction with fault-normal compressive stress. Examinations of machined surfaces typically show that contact surfaces are never truly planar; instead they all have geometric

roughness (typically fractal). When two rough surfaces are placed into contact, then the actual contacts occur at only a few places (if the material is rigid, then only three contact points are enough to form a statically stable configuration). Of course, the normal force on the materials would be concentrated at these contact points. The result is that the contact points elasto-plastically deform into contact patches, that are typically called *asperities* (see Figure 8.23). The frictional shear tractions are also concentrated at these asperities. One common view is that the static coefficient of friction is simply describing how the total area of asperities changes as the normal stress is increased.

This view that friction laws are essentially describing the properties of surface asperities seems consistent with the observation that fault normal compressive stress is best measured by also including the effect of fluid pressures. That is, any interstitial fluids in the rough fault surface serve to push the surfaces apart. Obviously, this argument works well if the area of asperities is small compared to the overall area of the surface. If there are almost no void spaces, then the effect of fluid pressure is mitigated.

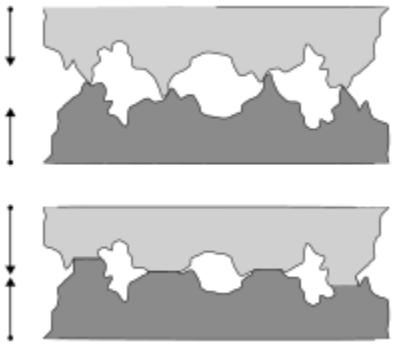


Figure 8.27. The contact area of geometric asperities are thought to be the key aspect of standard Byerly friction, The contact area increases with the normal stress

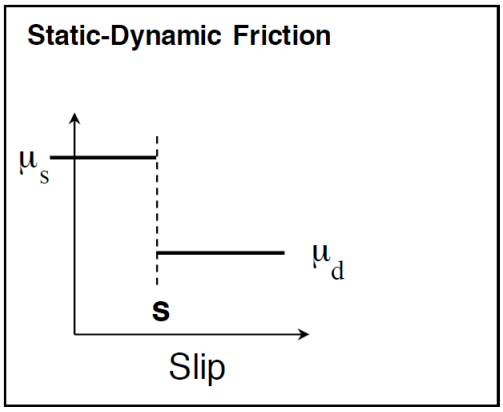


Figure 8.28. Instantaneous stress drop friction law where the friction drops instantaneously from a static coefficient of friction to a dynamic coefficient of friction. This law is unphysical because the properties of the material change without doing any work.

Although this "instantaneous stress drop" friction law is convenient for its simplicity, it has a fundamental and fatal flaw. That is, the material changes its physical properties without doing any work to the material. In the case of an instantaneous stress drop law, there is no fracture energy, which means that the material is infinitely brittle. A far more realistic situation is one in which the properties of a material change as a result of work that is done on the material. It may be that this work is small compared to other energies, but there must always be some work associated with changing the properties of a material. Realistic friction laws always involve parameters that describe the "state" of the material. In particular, as inelastic yielding occurs, the material's yield stress evolves as well. If we assert a simple friction law that changes from an initial friction to a final friction, it is necessary that there is some evolution of this change; it cannot be instantaneous.

Simple Slip-Weakening Friction

Slip-weakening friction as shown in Figure 8.29 is commonly used in simulations of earthquake rupture. It has the advantages that there is work associated with change in the frictional state of the material. That is, the work associated with the transition from static friction to sliding friction is just the area of the back-hatched region shown in 8.29. Slip-weakening friction is described by

$$\tau_f \le \sigma_n \begin{cases} \mu_S - (\mu_S - \mu_d)\dfrac{D}{D_0} & 0 < D < D_0 \\ \mu_d & D > D_0 \end{cases} \qquad 8.90$$

Where D_0 is the **slip-weakening distance**. The fracture energy (per unit of rupture area), G_0 , is given by

$$G_0 = \frac{1}{2}\sigma_n D_0 (\mu_s - \mu_d) \qquad 8.91$$

and the frictional heat energy (per unit of rupture area), E_F , is

$$E_F = \sigma_n \mu_d D \qquad 8.92$$

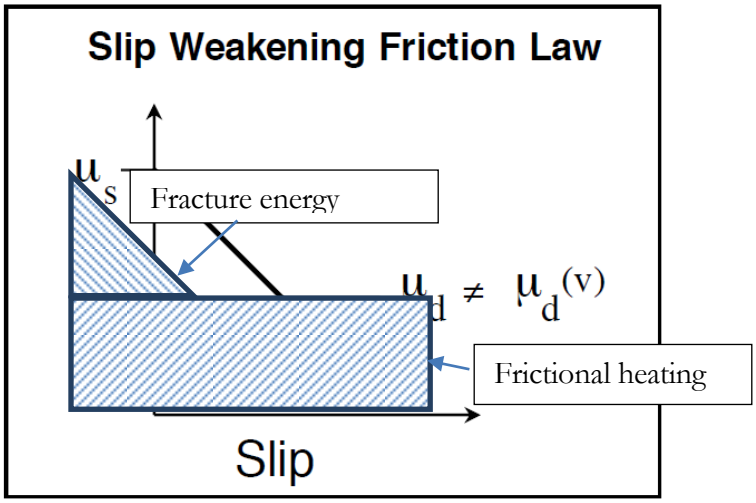


Figure 8.29. Slip-weakening friction law is common in rupture dynamics. The lower hashed area represents the total work done by frictional heating $\sigma_o D$, while the upper back hashed area represents the work done to drop the friction and is commonly interpreted as fracture energy.

As long as $D > D_0$, slip-weakening produces a simple stress drop. However, this law also includes a length scale that is associated with changing the frictional state from static to dynamic. In fact, any friction law that changes the internal energy of a material has included (either explicitly or implicitly) a characteristic length in the friction model. Rather interestingly, **Aki's conjecture of scale-free rupture dynamics is violated by any friction law with a scale-independent fracture energy**.

Notice that slip-weakening friction specifies the maximum shear that can occur before sliding, $\mu_s \sigma_n$. Sometimes this number is called the strength of the fault since it is the maximum shear stress that can be achieved locally. While this parameter is relevant to the conditions of crack nucleation, it turns out that it has virtually no effect on the dynamics of propagating failures. Instead, propagating failures are more related to the work necessary to extend a rupture, which is the fracture energy. Fracture energy is the product of peak stress and slip weakening distance, and neither quantity can be independently determined from dynamic modeling of rupture (Gutierri and Spudich).

A major problem with slip-weakening friction centers on the problem of rupture nucleation. That is, in order to have a seismic instability, the rate at which potential strain energy is extracted by the rupture must exceed the rate at which energy is dissipated by friction and fracture on the expanding rupture surface. We can approximate the change in potential energy due to faulting that is available for radiation, $\Delta E_{W-f}(t)$, as the change in potential energy minus the frictional energy as

$$\begin{aligned} \Delta E_{W-f}(t) &\approx S(t)\bar{D}(t)\sigma_{eff} \\ &\approx S^{3/2}\frac{\Delta\sigma}{C\mu}\sigma_{eff} \end{aligned} \qquad 8.93$$

Where S is the rupture area, $\bar{D}$ is the slip averaged over the rupture area (scales with L, see 8.69), and σ_{eff} is the difference between the prestress and the dynamic friction (see 8.7). That is, it is the effective stress that is available for dynamics.

I have written this as an approximate relationship because I am assuming that the effective stress is constant in space. Later we will see that the prestress, the slip, and the dynamic friction are spatially heterogeneous. 8.93 needs to be modified if it correctly considers this heterogeneity.

Now the energy that is available for radiation E_R (kinetic energy) is just

$$\begin{aligned} E_R &= \Delta E_{W-f} - E_G \\ &\approx S^{3/2}\frac{\Delta\sigma}{C\mu}\sigma_{eff} - SG_0 \\ &= S\left(\frac{\Delta\sigma}{C\mu}\sigma_{eff}\sqrt{S} - G_0\right) \end{aligned} \qquad 8.94$$

where G_0 is the fracture energy per unit rupture area. This means that in order **to have spontaneous sliding** , the change in potential energy must be greater than the energy required for fracture and friction, or,

$$\sqrt{S} > {G_0 C \mu}/{\Delta\sigma\sigma_{eff}} \qquad 8.95$$

That is, slip weakening friction implies that if the prestress is homogeneous, then there is a minimum dimension for earthquakes. If we use the laboratory measures of fracture energy (say 200 J/m^2), and effective stress of 2 MPa (estimated from seismic energy radiation), a typical stress drop of 2.7 MPa, then the minimum rupture length for an earthquake is about a meter, which is very small. However, if we use fracture energies that are required to stop a large-dimension crack in a uniform stress field (see 8.35) then we end up with fracture energies of tens of MJ/m^2 . In this case, it is necessary to use high prestress in the nucleation region just to get things started. Currently, many researchers insert patches of high pre-stress in order to nucleate instabilities. While inserting high-stress patches is easy on a computer model, it requires a religious miracle to realize in the real world.

Unfortunately, many researchers report the fracture energies that are employed in their dynamic modeling and interpret this energy as having physical significance. They rarely mention that using small fracture energies causes extreme numerical instability in their dynamic calculations. That is, in many cases the reported fracture energy is the energy that is required to produce stable finite-element models (a function of the grid spacing and time step). Small fracture energies produce singular stresses at the crack front. Discrete element codes cannot handle these singular stresses, and the modelers increase the fracture energy in order to soften the singularity at the crack tip. In addition, the use of small fracture energy means that this parameter has almost no impact on the details of the fracture dynamics. Basically, fracture energy is used to stop fractures. However, it is impossible to stop long cracks without very large (stupidly large) fracture energies.

Another consequence of using fracture energy together with crack-like dynamics is the fact that discrete numerical simulations are not only poorly resolved, they also typically jump to super-shear rupture velocities as the rupture dimension grows. In discrete models (e.g. finite elements) with low fracture energy, the expanding crack fronts evolves into such large stresses that elements several grids in front of the crack can begin to slip even before the crack front arrives. That is, the rupture jumps ahead of the crack front. This is super-shear rupture velocity and once it starts, it can propagate stablely. Although there have been reports of evidence for super-shear rupture velocities in some earthquakes, most well recorded earthquakes appear to have sub-shear rupture velocities. Given the

behavior I just described, the real mystery is to explain why all significant earthquakes don't have super-shear rupture velocities. I will shortly show that a better way to control the rupture dynamics is with spatially heterogeneous prestress.

Rate and State Friction

Although instantaneous weakening friction and slip-weakening friction have been commonly used in earthquake dynamics because they are so simple, they are not actually based on experimental observations of frictional sliding. Frictional sliding is actually a specialized type of localized inelastic shear straining. In general, the inelastic shear strain rate depends on the applied shear stress and it also depends on the past history of inelastic strain (the "state" of the interface) and it can also depend on the current rate of sliding. Friction laws that have these two dependencies belong to a class of laws called **rate and state friction laws.** The use of variables that characterize the state of a deforming material are quite common in material science; for example, the stress-strain relationship for reinforced concrete is very complex. In particular, it is important to know if the concrete has been cracked in previous straining. In that sense, the amount of fracturing is a state variable for reinforced concrete.

In the Earth Sciences, the term, "Rate and State friction" has been used almost exclusively to mean a particular friction law that was developed to describe the temporal behavior of a number of low-slip-rate (less than 1 mm/s) frictional sliding experiments. I personally prefer to use the term "Dieterich-Ruina friction" when referring to friction laws of the type described below. Since rate and state dependent rheologies are common in mechanics, it's easy to be confused when using the phrase rate-state friction.

In a foundational set of friction experiments, James Dieterich (1979, Modeling of rock friction 1. Experimental results and constitutive equations, JGR, 84, 2161-2168) observed the time variation of slip along the interface of machined blocks of granite that were subjected to time-dependent shear stress and constant fault-normal compression. Dieterich discovered that if a block was forced to slide at a steady slip rate, then an instantaneous step in the shear traction caused a change in the slip rate and that this change in slip rate was not instantaneous, but instead evolved over some time scale (or alternatively a slip scale). Later, Andrew Ruina (1983, Slip instability and state variable friction laws, JGR, 88, 10,359-10370) introduced a state variable that could predict this behavior.

The Dieterich-Ruina friction law is briefly summarized in Figures 8.26 and 8.27, which I copied from a presentation by Chris Marone at Penn St. This friction law provided a new framework for studying fault dynamics. Most importantly, the definition of several constants $\left(A, B, V_0, D_c\right)$ accurately mimics the experimentally

observed slip rate as a function of time for a wide variety of materials. This friction law is often used as the basis of numerical modeling of fault slippage in a wide range of problems. Notice that the Dieterich-Ruina friction has a state variable, $\Theta(t)$, which is described by a 1st order differential equation in either time, or alternatively, slip. $\Theta(t)$ has units of time and it controls how the friction evolves after jumps in slip velocity (or alternatively, shear stress). The evolution after a jump evolves logarithmically (either time or slip) towards a new steady state. Clever experiments that allow one to observe the evolution of contact patches during slip show that the asperities are constantly evolving on a slipping surface. Slow slip velocities allow dislocation creep to increase the asperity area (indentation creep), thereby increasing the friction. In this case, the friction decreases with increasing sliding velocity (but only after the effect of the state variable has decreased. On the other hand, increasing the slip velocity can cause higher temperatures at the asperities, which allows them to relax faster and thereby increasing the friction when the sliding rate is increased. These two effects are included in the Dieterich Ruina friction law as the constants, A and B. If A > B, then the Dieterich Ruina law predicts that increasing the sliding rate results in lower friction. This is called rate-weakening friction and it produces dynamic instabilities in the sliding process. That is, if there is a perturbation that causes the sliding rate to increase, then this decreases the friction, which in turn, increases the sliding rate. This dynamic instability naturally results in earthquake nucleation.

On the other hand, if A < B, then faster sliding results in higher friction and is referred to as rate-strengthening friction. Perturbations to the sliding rate tend to damp out with time. I often see that faults that fail in earthquakes are interpreted as having A > B, whereas faults that experience creep are interpreted as having B > A.

It is worth noting however, that B > A seems to imply that a fault creeps steadily. It is difficult to observe spatial and temporal variations in creep (there are no radiated waves and the creep is deep in the Crust). However, when creep has been observed in detail, it's my experience that it mainly occurs during creep events. These creep events seem to be much like other earthquakes, but with a greatly reduced rupture velocity (usually less than 1 km/hr).

There is a very large literature that discusses the attributes of Dieterich-Ruina friction and these class notes are far too brief to really do this subject justice. I recommend reading the review paper by Rice, Lapusta, and Ranjith (Rate and state dependent friction and the stability of sliding between elastically deformable solids, J. of the Mech. And Phys. Of Solids, 2001, 49, 1865-1898).

One important cautionary note concerns the range of laboratory conditions that were observed when creating this friction law. In particular, the experiments limited slip rates to microns per s and observed characteristic slip distances, D_c, of microns. Furthermore, the changes in friction were small changes about average frictions of about 0.7. That is to say, the dynamic friction had similar values to traditional Byerlee friction, with the important addition of time evolution that provides a natural pathway to event nucleation. It's worth noting that when these experiments were sped up to earthquake slip velocities (> 1 m/s), then frictional melting always occurred. Once the material surfaces melts, these friction laws no longer apply. That is, while rate-weakening Dieterich-Ruina friction appears to be a reasonable model for the nucleation of earthquakes, it is inconsistent with observations of large earthquakes. Furthermore rate-strengthening friction provides a convenient model for creeping faults, but it fails to predict the common observation that creeping faults typically fail in slow earthquakes, which are often called creep events.

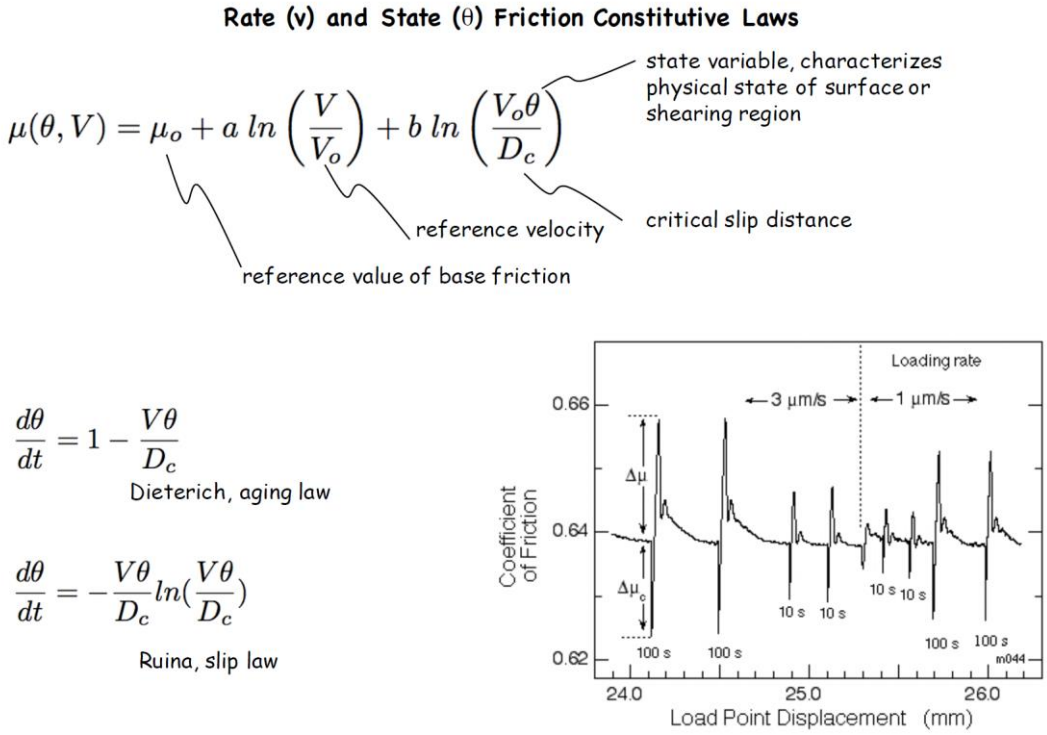


Figure 8.30. Copied from Chris Marone

Rate (v) and State (θ) Friction Constitutive Laws

1) $\mu(\theta, V) = \mu_o + a\, ln\left(\frac{V}{V_o}\right) + b\, ln\left(\frac{V_o\theta}{D_c}\right)$

2) $\frac{d\theta}{dt} = 1 - \frac{V\theta}{D_c}$

Convention is to use a, b for friction and A, B for Stress

$$\tau(\theta, v) = \tau_o + A\, ln\left(\frac{V}{V_o}\right) + B\, ln\left(\frac{V_o\theta}{D_c}\right)$$

$$A - B = \frac{\Delta\tau}{\Delta lnV}$$

Steady-state velocity strengthening if a-b > 0, velocity weakening if a-b < 0

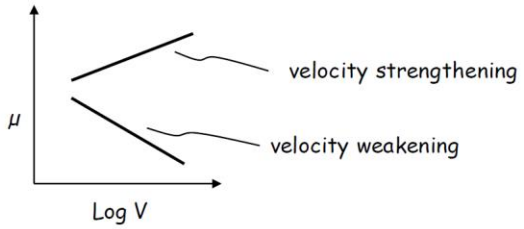


Figure 8.31. (copied from Chris Marone)

Frictional Heating

Assuming either instantaneous weakening friction, slip-weakening friction, or even Dieterich-Ruina friction, we can estimate the average long-term rate of frictional heating (or power per unit rupture area) on a fault to be

$$P_F = \overline{\dot{D}}\sigma_n\overline{\mu}_d \qquad 8.96$$

Where $\overline{\dot{D}}$ is the average geologic slip rate of a fault and $\overline{\mu}_d$ is the average dynamic coefficient of friction. It is a simple calculation to estimate the heat production on the San Andreas fault assuming that the dynamic coefficient of friction is about 0.6 (i.e., Byerlee friction). Assuming that the average dynamic friction on the San Andreas is about 200 MPa (see Figure 8.22), and that the slip rate is about 3 cm/yr (or about 10^{-9} m/s) gives $P_F \approx 200\ \mathrm{mW/m^2}$. This is actually a problematically large power; the average power of heat transferred through the Earth's continental crust (otherwise known as the heat flow) is 66 mW/m^2 (Davies and Davies, 2010). That is, the assumption of Byerlee friction implies that the average rate of frictional heat produced by the San Andreas fault is several times

the average heat flow in the Crust and thus we would expect to observe elevated heat flow in the vicinity (within 10's of km) of the San Andreas fault. This elevated heat flow is not observed and this fundamental problem is often called the **Heat-flow Paradox** (see for example, Scholz, C. H., The Mechanics of earthquakes and Faulting, Cambridge University Press, Cambridge England, 1990).

Actually, the heat flow paradox is far more severe than is indicated by the calculation above. In particular, a variety of observations indicate that average slip rates during earthquakes are at least 2 m/s. This means that, if dynamic friction is about 200 MPa during the slip event, then we expect frictional power of at least 400 MW/m^2 during the sliding. This is an extraordinarily large number. If the average slip is 1 m, then the frictional heat is the energy of 200 sticks of dynamite applied to every square meter of the fault. Or, for comparison, the power of a nuclear power reactor is about 1 GW. The specific heat of granite is 790 J/^{o}Ckg, and the density is 2.7x10^{3}kg/m^3. Assuming that a temperature change of 1,000^{o}C would cause melting (rocks have a very low latent heat of melting), then the heat $E_{1\text{mm melt}}$ required to raise the temperature of a 1mm thick section of the fault by 1,000^o C is

$$\begin{aligned} E_{1\text{mm melt}} &\approx 4\times10^8\left(\frac{\text{J}}{\text{s}}\right)T_S \\ &\approx 2.7\times10^3\left(\frac{\text{kg}}{\text{m}^3}\right)\times\left(\frac{\text{m}}{10^3\text{mm}}\right)\times 8.2\times10^2\left(\frac{\text{J}}{^0\text{kg}}\right)\times10^{3\ 0} \\ &\approx 2.2\times10^5\left(\frac{\text{J}}{\text{mm-m}^2}\right) \end{aligned} \qquad 8.97$$

Now the total frictional heat energy is

$$E_{Friction} \approx 4\times10^8\left(\frac{\text{J}}{\text{m}^2\text{s}}\right)T_S \qquad 8.98$$

Where $T_S \equiv$ duration of slip at a point. If $T_S = {}^1\!/_2\, s$, then the slip is about 1 m and there is enough frictional heat to raise a 2-m thick zone by 1,000^o. Clearly, any coefficient of sliding friction larger than 0.05 would cause instantaneous melting of the fault-zone materials. This fact was originally pointed out by Harold Jefferies, and a more detailed calculation was described by Paul Richards (197?).

Although there are geological observations of melts observed in fault zones that have been exhumed by tectonic uplift followed by erosion of overburden materials (see for example, the Punchbowl Fault described by Chester and Chester, 199?), the total volume of melt is very small compared with the amounts

expected from ordinary Byerlee friction. In fact, once a melt layer is developed, one expects the fault tractions to be controlled by the viscosity of the melt. Furthermore, we would expect to see that the final fault shear traction should drop to zero, which is often referred to as complete stress drop. A 200 MPa stress drop is a factor of one hundred larger than the average observed in major earthquakes. This means that the average slip should be a factor of 100 times larger than is typically observed.

Notice that the estimate of $200\,{}^{\mathrm{kJ}}\!/_{\mathrm{mm\text{-}m^2}}$ to melt a mm-thick layer is an order of magnitude smaller than estimates of the far-field radiated energy from a earthquake with a 1-m slip, about $2D\,{}^{\mathrm{MJ}}\!/_{\mathrm{m_{slip}\text{-}m^2}}$ (see 8.42). That is to say that we should expect to find extensive frictional melting in the vicinity of active faults. The big mystery is, why they are not observed.

At this point, I will say that the paucity of melted rocks in exhumed faults implies that the total frictional heat per unit rupture is less than about 200 kJ/m^2. This implies very low dynamic friction for meter-sized slips. At this time, such a low dynamic friction has not been observed in a laboratory experiment. Which is to say that although it's clear that dynamic friction in large earthquakes is very low, we don't know why.

Strong Rate-Weakening Friction

In my 1990 slip-pulse paper, I argued that strong rate-weakening friction could be the key to understanding the origin of slip pulses. I argued that it could also help to explain the low heat observed from tectonic deformation.

One simple realization of a rate weakening friction is to hypothesize that

$$\mu = \frac{\mu_0}{1+\left({}^{\dot{D}}\!/_{\dot{D}_0}\right)^{\beta}} \qquad 8.99$$

(see Cochard and Madariaga, 199?). In this friction law, the friction has a power law dependence on slip velocity. There is no inherent distance in this friction law, and hence this friction law cannot provide any inherent length scale to rupture problems that are described by this law. While none of the friction laws discussed to this point are capable of generating slip pulses, this friction law can. In order to estimate how much less heat can be produced by this friction law, we integrate the frictional force over the slip in a slip pulse, or

$$E_F = \int_0^{T_0} T_F(t) D(t)\, dt = \int_0^{T_p} \left(\frac{\mu_0}{1 + \left(\dot{D} / \dot{D}_0 \right)^{\beta}} \sigma_n D_p(t) \right) dt \qquad 8.100$$

Now assuming that the slip history of a pulse can be approximated as a truncated growing crack, then

$$D_p(\tau_p) \approx \begin{Bmatrix} D_0 \sqrt{\tau_p} & 0 < \tau_p < 1 \\ D_0 & \tau_p > 1 \end{Bmatrix} \qquad 8.101$$

where $\tau_p \equiv t / T_p$. And then $dt = T_p d\tau_p$, and

$$\dot{D}(\tau_p) \approx \frac{1}{2} \frac{D_0}{\sqrt{\tau_p}} \qquad 8.102$$

Substituting 8.101 and 8.102 into 8.100 gives

$$E_F = \sigma_n \mu_0 D_0 \int_0^1 \frac{1}{1 + \left(\dfrac{\dot{D}(\tau_p)}{\dot{D}_0} \right)^{\beta}} \sqrt{\tau_p}\, d\tau_p$$

$$= \sigma_n \mu_0 D_0 \int_0^1 \frac{1}{1 + \left(\dfrac{D_0}{\dot{D}_0} \dfrac{1}{2\sqrt{\tau_p}} \right)^{\beta}} \sqrt{\tau_p}\, d\tau_p \qquad 8.103$$

Or

$$E_F = \sigma_n \mu_0 D_0 F_{rw}(\beta) \qquad 8.104$$

Where F_{rw} is a factor that includes the effect of power-law rate-weakening and

$$F_{rw}(\beta) = \int_0^1 \frac{4\tau_p}{1 + \left(\dfrac{1}{\sqrt{\tau_p}} \right)^{\beta}} d\tau_p \qquad 8.105$$

Unfortunately, I have not been able to find an analytic solution to this integral. However, this definite integral can be numerically evaluated, and the answer is shown in table 8.2.

β	F_{RW}
0.1	0.98
0.5	0.94
1	0.88
2	0.77
4	0.61
8	0.43
16	0.27
32	0.15
64	0.081
128	0.042

Table 8.2

Table 8.2 seems to imply that the use of this rate-weakening law still results in substantial frictional heating. That is, unless the law assumes very strong rate weakening (i.e., $\beta > 20$). The paucity of melt observed in exhumed faults is strong evidence that dynamic friction drops very rapidly behind the rupture front. Perhaps it is a sudden transition from high static friction $\left(\mu_{static} \sim 0.6\right)$ to very low dynamic friction $\left(\mu_{dynamic} < 0.05\right)$. At the time of these notes, I am unaware of experimental observations of such low dynamic friction.

Dynamic Friction in Landslides sudden transition to low dynamic friction

One of the most direct ways to observe a rapid transition from high static friction to very low dynamic friction is to study the dynamics of landslides. For example, Fig. 8-32 is a photograph of Mt. St. Helens just seconds prior to its eruption in 1980. Mt. St. Helens is a typical composite volcano. These volcanoes are conical structures that form by the accumulation of pumice, ash, and volcanic bombs that eject from the top of the volcano. In essence, a composite volcano is simply a large pile of unconsolidated material that is sitting at its angle of repose. The angle of repose is the steepest slope that a pile of granular material can obtain in static equilibrium. The angle of repose is determined by the coefficient of friction between the granular materials. In the case of Mt. St. Helens, the volcanic cone was in static equilibrium with a slope of about $\theta \approx 29^{o}$. The inter-granular coefficient of friction of a marginally stable cone of material is $\mu_f \approx \tan\theta$, or in the case of Mt. St. Helens, $\mu_f \approx 0.54$. Steady inflation of the magma beneath

the volcano occurred for a period of several months prior to the eruption. This inflation (about a 100-m bulge) over-steepened the slope. About 10 seconds prior to the eruption, a M 5.1 earthquake shook the over-steepened slope, and an enormous landslide was triggered. Figure 8-33 is a photograph taken early in the landslide. A remarkable time lapse of photographs provides data to estimate the velocity of the sliding mass.
https://www.bing.com/videos/search?q=mt+st+helens+time+lapse+video&qpvt=mt+st+helens+time+lapse+video&view=detail&mid=58ACA9463AB2BDCC740F58ACA9463AB2BDCC740F&&FORM=VRDGAR

In the time lapse, it can be seen that approximately 3 km^3 of material slid downslope with a velocity of about 200 km/hr. The base of the slide appears to be on a steep slope of about $\beta \approx 50^o$. The acceleration of the landslip can be approximated as $\ddot{u} \approx g\left(1-\mu_{dynamic}\right)\sin\beta$. The time lapse indicates that the friction dropped almost instantaneously to near zero and that the landslide mass accelerated as if frictionless on the steep sliding plane. As the mass slid downslope, the pressure from the weight of the landslide mass was decreased in the magma chamber (very high viscosity alkaline magma) beneath the mountain. This caused rapid outgassing of volatiles that were dissolved in the previously confined magma. This rapid outgassing caused the lateral blast phase of the eruption.

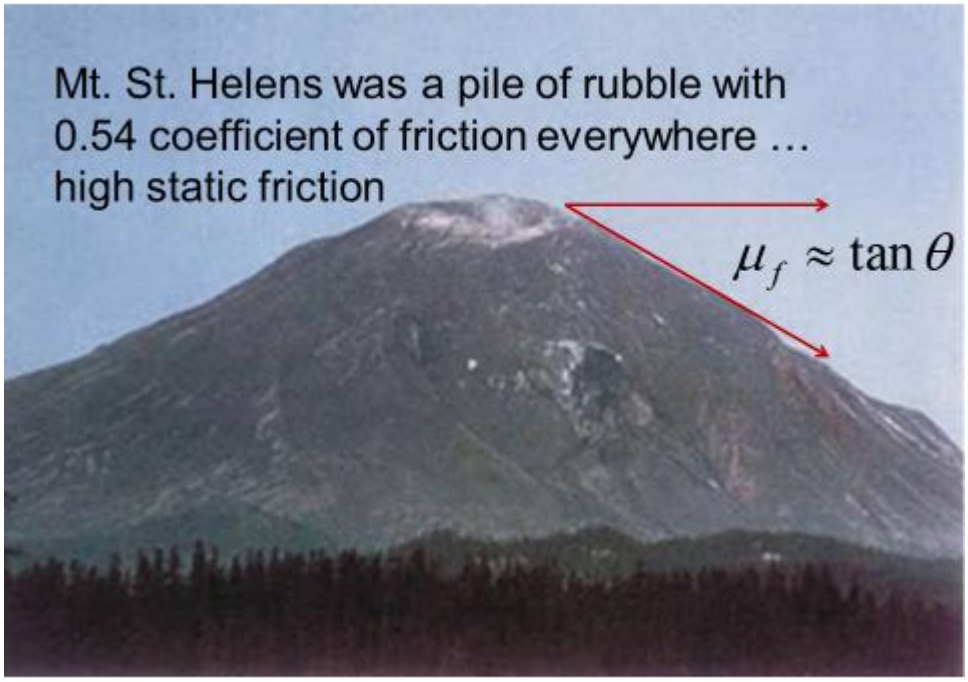


Figure 8.32. Mt. St. Helens volcano prior to its eruption in 1980.

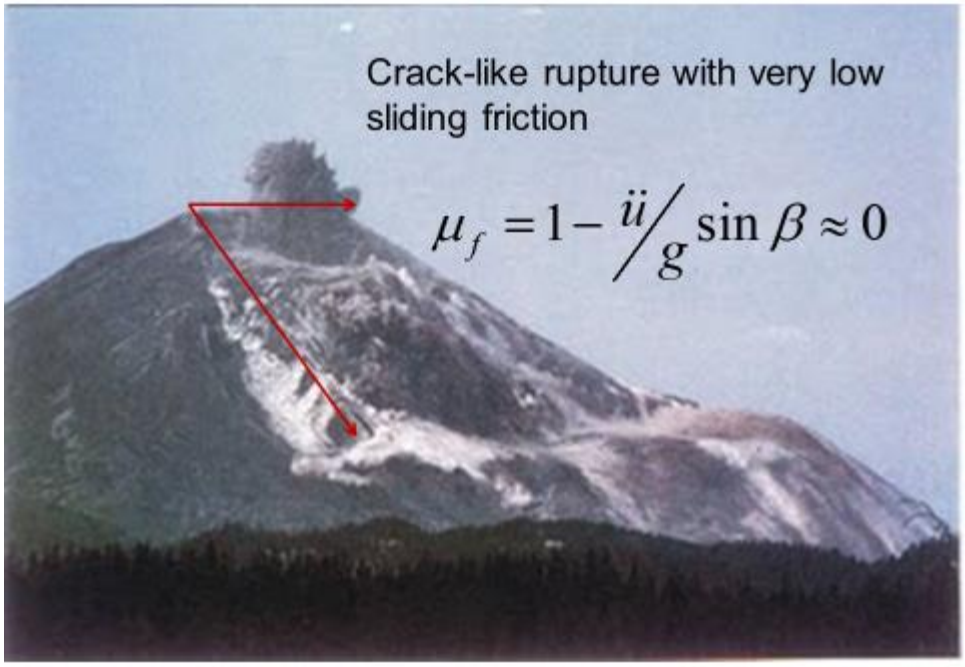


Figure 8.33. Mt. St. Helens during the massive landslide that depressurized the magma chamber.

Landslides are very similar to earthquakes; they are dynamic shear sliding on a slip plane. However, in earthquakes, the rupture is confined to a portion of the fault plane, while in landslides, there are no ends of the rupture plane. In essence, landslides are crack-like; the entire fault plane slips simultaneously. The downslope slips can become very large since gravity provides the driving mechanism. In an earthquake the fault-plane shear stress decreases as slip increases. In landslides, the stress does not decrease until the mass is at the bottom of the hill.

The most important point of Figure 8.33 is that the static friction at Mt. St. Helens was at least 0.54 prior to the M 5.1 earthquake. A failure plane with very low dynamic friction (less than 0.1) formed very rapidly during the massive landslide. If the rupture velocity was similar to earthquakes (V_R about 2 ½ km/s), then this plane formed in less than 1 s.

One simple way to parameterize the dynamic friction in a landslide is to estimate the ratio of the vertical displacement, *H*, to the horizontal displacement, *X* , of the center of mass. This is a measure of the average slope of the landslide, and the effective coefficient of dynamic friction is just H/X . Estimates of dynamic friction in landslides vary from almost 1.0 to as low as 0.06. There seems to be a clear indication that landslides with larger total volumes tend to

have smaller dynamic frictions. $\mu_{dynamic} \approx H/X \approx 1/\sqrt[8]{1+V}$ (where V is landslide volume in m^3) approximately fits widely scattered observations (Johnson, Campbell, and Melosh, 2016, The reduction of friction in long runout landslides as an emergent phenomenon, JGR, 121, 881-889, doi/10.1002/2015JF003751).

Force Chain Networks

I will now briefly describe the behavior of granular particle systems. This will help to introduce the concepts of self-organization, emergent systems, and fluidization.

Sand is an example of a material that has a rich variety of behaviors. It is comprised of simple grains, each of which is described by the dynamics of a rigid body. However, a collection of interacting sand grains can display behavior that is difficult to anticipate based solely on the rigid body dynamics of each particle. For example, sand in an hourglass can flow like a viscous liquid. Sand can also interact with wind to form sand dunes that evolve and appear to slowly propagate through space. Sand can also compact into a closely packed material that is well suited to form the foundation of structures (condensed matter physicists call this a jammed granular material).

In the case of sand foundations, it's important to understand the conditions that cause inelastic yielding of the sand. One approach to collections of sand is to consider that the particles are so small and numerous that they can be simulated using a continuum (i.e., there is no smallest scale). Numerous tests of yielding sand has demonstrated that it can be described as an elasto-plastic continuum. In the case of sand, the plastic yield shear stress is dependent on the confining pressure of the sand. This is usually called Drucker-Prager plasticity. In contrast, steel is described with Von Mises plasticity, where the yield stress is independent of confining pressure.

The idea that there is a maximum shear stress that can be supported by sand seems intuitively reasonable, and you might believe that that is all you need to understand about the deformation of an assemblage of sand particles. However, there has been new research that is revealing surprising mechanics behind the simple statement that a sand is elasto-plastic. In a sense, it should not be unexpected that there is something else at play; sand is not elastoplastic at the grain level (it's elastic-brittle). It's only a large assemblage of sand particles that behave plastically.

Experiments that reveal the mechanics of interacting sand grains show that most of the load carried by sand is actually borne by a relatively small percentage of the sand grains. Furthermore, even if the external tractions may produce bulk shear

of the sand assemblage, the actual loads are primarily axial stresses directed along collections of grains. These are called **force chains**. The details of how and when individual chains appear are very sensitive to particular shapes and orientations of individual grains. Although these systems were first studied using physical experiments, recent advances in computational dynamics have demonstrated that these types of systems can be numerically simulated (for example, see https://resolver.caltech.edu/CaltechTHESIS:12042014-104112714).

Fig. 8.34 is copied from the website of Robert Behrenger (Duke Univ. Physics Dept., deceased). The force chains carry axial forces that resist shearing force. The axial forces cause shear in the photosensitive beads. A video from this experiment shows how finite motions result in eventual failure of individual force chains. After failure, new force chains rapidly develop and the ensemble behavior of the system appears as if it's a plastic material.

The force chains shown in Fig. 8-34 are a complex network that forms almost instantaneously. This network is an example of a **self-organizing system**. The formation of the force chains is natural and fundamental to understanding the mechanics of sand. Even though prediction of the details of the network of force chains is very difficult, you can be certain that appropriate force chains will occur.

Although the sand system deforms in a sequence of small jumps, the effect of individual chains is small enough that a system of many particles behaves similarly to traditional sliding friction. The system shown in Fig. 8.34 is an example of a **jammed granular system**. If the confining pressure of the system is removed then the particles can move freely and the same sand grains can show deformation that is similar to a viscous fluid, which is typically referred to as a **granular fluid**. Granular fluids are important in many industrial processes since this is a way to mix and transport granular materials. The transition between jammed and fluidized states can occur very quickly.

Fig. 8-35 is a photo taken after a massive landslide that was triggered by the M 7 ½ 1959 Hebgen Lake Earthquake. A substantial fraction of a 2.3 km high mountain slid into the Madison River Valley at an estimated speed of 150 km/hr. The momentum of the slide was sufficient to cause the slide to climb several hundred m up the opposite valley wall. As soon as the slide's velocity dropped to zero as it climbed the valley wall, it froze into place. This seems to be direct evidence of sudden transitions in the apparent friction when 1) it was initially stable on the mountain slope, 2) it suddenly transitioned to very low dynamic friction allowing the slide to accelerate downhill at high speed, and 3) suddenly transitioned back to high friction when the sliding velocity dropped to zero.

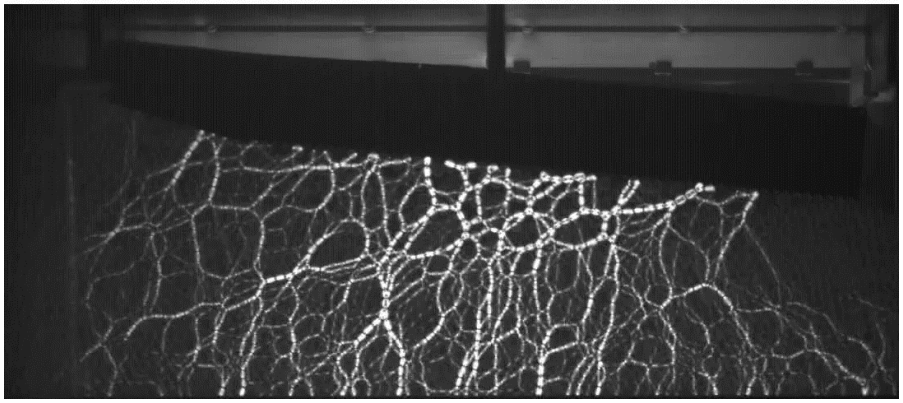


Fig. 8-34. Photograph of force chains that result from the shearing of an assemblage of photo-sensitive plastic beads that are confined to a rectangular bucket. The black wedge above is moving to the left. The polarized light is transmitted by beads experiencing shear strain that is the result of uniaxial stress carried by the force chains (from Robert Behrenger's web site).

.

Figure 8.35. Landslide across the Madison River in Montana triggered by the 1959 M 7 ½ Hebgen Lake earthquake. Note that the slide had sufficient momentum to climb the opposite valley wall. I seems that the friction transitioned back to high static friction when the sliding velocity dropped to zero.

Some of the most dramatic examples of large landslides are from the Hawaiian Islands. The dotted lines in Figure 8.36 map out the areal extent of landslide deposits that originated from the slopes of Hawaiian volcanos (Moore, J., D. Clague, R. Holcomb, P. Lipman, W. Normark, and M. Torresan, 1989, Prodigious submarine landslides on the Hawaiian Ridge: Journal of Geophysical Research, v. 94, p. 17,465–17,484). Several of these enormous landslides appear to have run out more than 100 km from their source region. Furthermore, there is evidence that giant tsunamis (run-up heights exceeding several hundred meters) carried beach rocks far up the flanks of existing volcanos. The most recent of these events has been dated at about 100,000 ybp (Megatsunami deposits on Kohala volcano, Hawaii, from flank collapse of Mauna Loa, McMurty, and others, 2004, Geology, 32, 741-744, doi: 10.1130/G20642.1).

Figure 8.37 shows a map of the head scarp of one of these landslide structures (P Lipman, T. Sisson, and J. Kimura - 2-4. Hilina slump area, 2001, Public Domain, https://commons.wikimedia.org/w/index.php?curid=69511814). This scarp is called the Hilina slump (or alternatively, the Hilina fault). Figure 8.38 is a photograph of the Hilina escarpment. Obviously, the landslide is currently in equilibrium, which means that the friction in the landslide exceeds the shear forces that result from the gravitational weight of the slide.

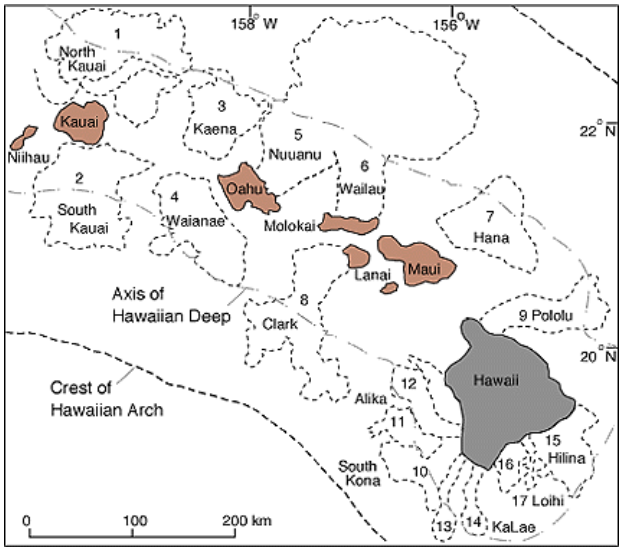


Fig. 8.36. Spatial extent of debris fields created during pre-historic collapse of the flanks of Hawaiian volcanoes. These enormous landslides sometimes ran more than 100 km along the ocean bottom. (Moore and others, 1989)

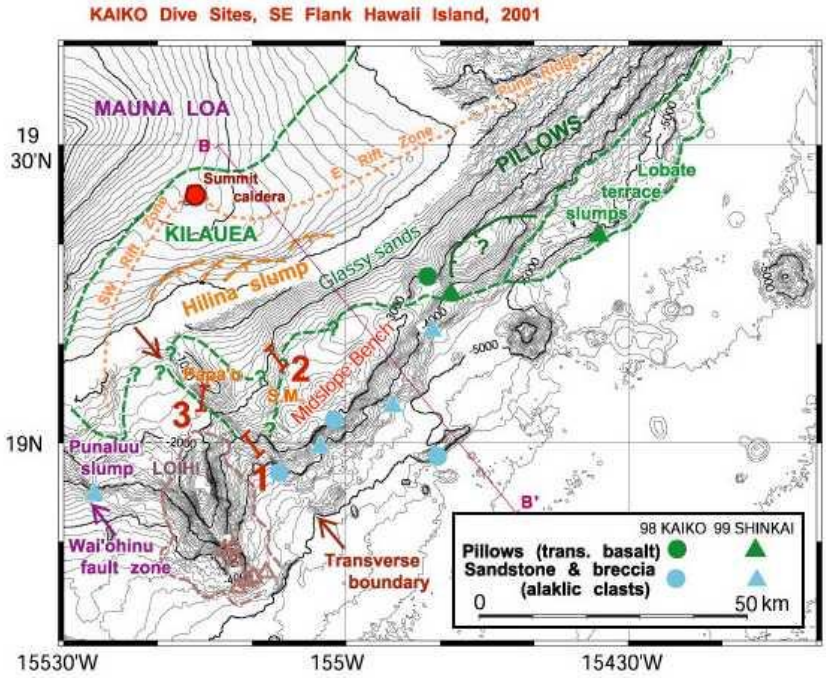


Figure 8.37 Map showing the extent of the Hilina slump which appears to be an incipient flank collapse structure.

Figure 8.38. Photograph of the escarpment of the Hilina fault (landslide).

Figure 8.39 shows a schematic cross section of the Hilina slump. The eruption of lavas built the broad swell of the Island of Hawaii. The old ocean floor sits at the base of volcanic deposits, which are more than 10 km thick in the center of the volcano. The weight of the volcanics has depressed the ocean floor so that the interface between the volcano and the ocean floor dips shallowly towards the volcano. This means that the toe of the landslide actually moves upwards due to gravitational collapse of the head of the landslide.

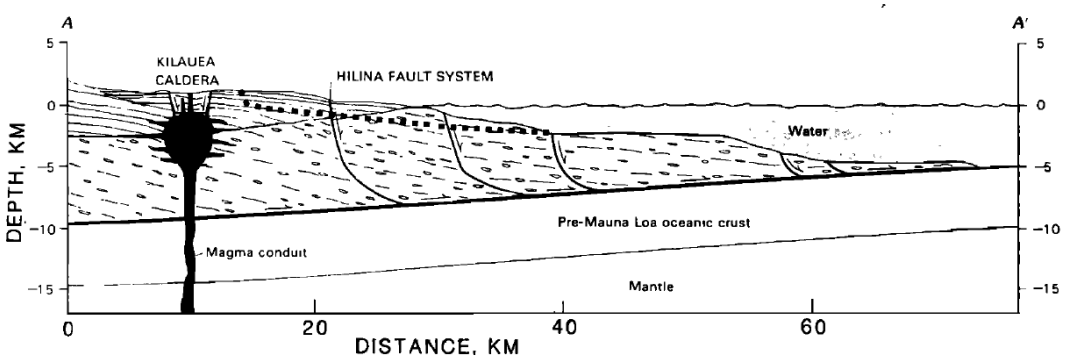


Fig. 11. Section through Hilina slump with no vertical exaggeration modified from *Lipman et al.* [1985]. Location of section shown in Figure 10. Dashed pattern forming base of volcano is subaerially erupted clastic rocks including phreatic explosion ash, littoral cone ash, flow foot breccias, minor pillowed flows, and pillow breccias, landslide debris, and turbidity flow deposits. Most of this material was probably erupted from Mauna Loa except for the uppermost few kilometers. Lined pattern forming upper part of volcano is subaerially erupted tholeiitic basalt, predominantly thin (2–10 m thick) pahoehoe and aa lava flows, but including ash layers. Lava from Mauna Loa predominates over that from Kilauea in north part of section. Dotted line indicates possible sole of future debris avalanche favored by steep regional slope.

Figure 8.39. A cross section showing the structure of the Hilina slump. Notice how this cross section resembles cross sections intended to show the shallowest section of subduction zones.

Landslides vs. Earthquakes or Cracks vs. Slip Pulses

Although the Hilina slump appears to be an example of an incipient long-run-out landslide, it has experienced other modes of failure. In particular, there have been several large earthquakes that seem to have occurred on the basal interface between the ocean floor and the volcanic deposits. The 1975 M 7.4 Kalapana earthquake and the 2018 M 6.9 Leilani Estates earthquake both appear to have occurred as shallow-dipping thrust earthquakes along the interface at the base of the Hilina slump. Figure 8.40 shows the surface projection of the contours of slip that were used to simulate teleseismic P-waves.

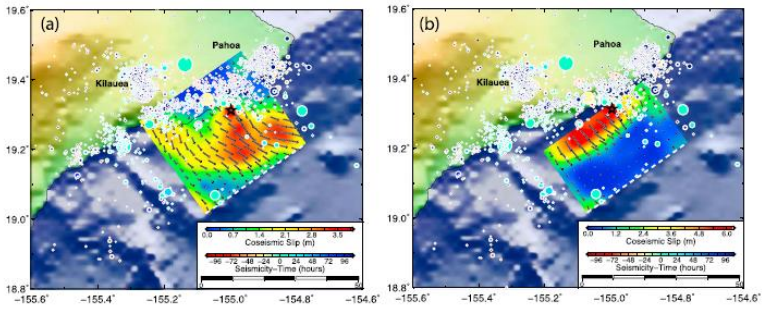


Figure 4. Finite-fault slip inversions using teleseismic broadband *P* and *SH* waves for the 4 May 2018 Hawaii Island earthquake for (a) $\phi = 235°$ and $\delta = 7.5°$, constrained to fit onshore GPS data and regional tsunami recordings (Bai et al., 2018), and (b) $\phi = 235°$ and $\delta = 3.0°$ constrained to fit the GPS data. The inversion in (a) gives $M_0 = 8.7 \times 10^{19}$ Nm (M_W 7.2), while that in (b) gives $M_0 = 7.5 \times 10^{19}$ Nm (M_W 7.2). U.S. Geological Survey catalog seismicity from 11 April to 14 May is shown by time-coded circles. Subfault slip direction and magnitude are indicated by arrows (direction of hanging wall movement). The black star indicates the 2018 event epicenter. The faults dip gently toward the northwest and are likely the basal décollement between the volcanic mass and the former Pacific seafloor. Inferred slip distribution and seismic moment depend on the dip angle estimate, so precise estimation is important.

Figure 8.40. Slip model of the 2018 M 6.9 Leilani Estates earthquake (Lay, T., Ye, L., Kanamori, H., & Satake, K.(2018). Constraining the dip of shallow, shallowly dipping thrust events using long-period Love wave radiation patterns: Applications to the 25 October 2010 Mentawai, Indonesia, and 4 May 2018 Hawaii Island earthquakes. Geophysical Research Letters, 45, 10,342–10,349. https://doi.org/10.1029/ 2018GL080042)

As is the case for almost all shallow crustal earthquakes, these moderate-sized earthquakes were probably propagating slip pulses. Furthermore, the dynamic friction during the slip pulses was probably very low (just like most shallow crustal earthquakes). The main difference between these earthquakes and a giant long-run-out landslide is that the landslides were certainly expanding cracks. Of course, a standard crack has a crack tip at at least one end (In many cases the crack extends to the surface). In the case of a landslide, both ends extend to the surface.

Presumably, there is a critical dimension for slip pulses. If the slip pulse becomes larger than this dimension, then it transitions into an expanding crack. At this point, I don't know what this critical size is, but I can make an educated guess. For example, the M 8.8 2010 Maule earthquake was well recorded by high-rate GPS stations located directly above the rupture. Figure 8.41 shows the displacements at several stations located above the rupture. Clearly this is an example of a slip pulse; the slipping had already stopped at stations in the south near the epicenter before rupture even began beneath stations in the north. Heaton, T., S. Minson, M Simons, 2011, S11A-2201 Characterization of the Slip Pulse for the 2010 M 8.8 Maule Earthquake, Fall AGU).

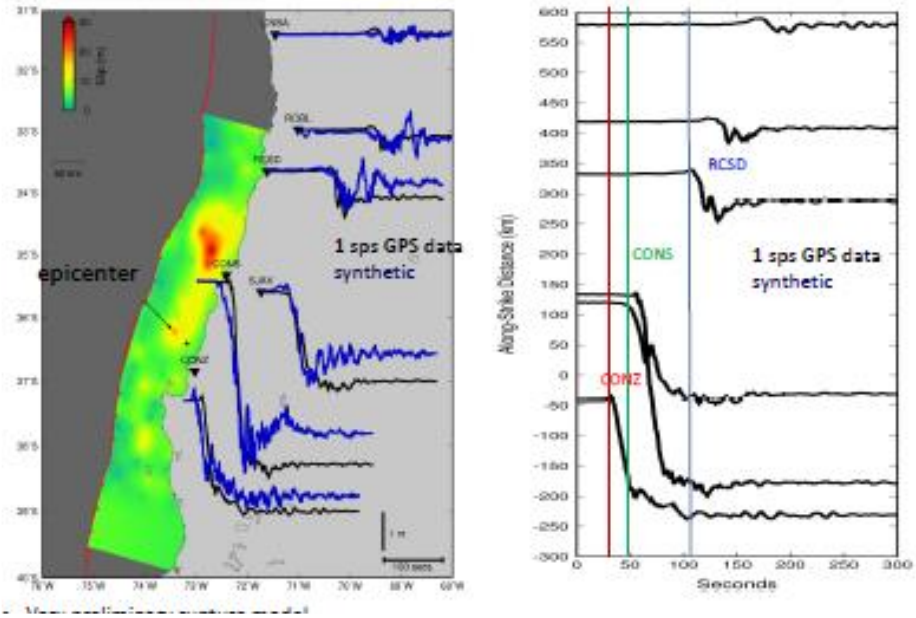


Fig. 8.41. GPS displacements recorded during the M 8.8 2010 Maule earthquake. This is clear evidence that the rupture consisted of a slip pulse; not all of the points on the rupture surface are slipping simultaneously.

While almost every reasonably well recorded earthquake seems to be best modeled with slip pulses, there are several examples that suggest that slip pulses grew so large that they turned into expanding cracks. I think that the best example of this behavior is from the 2011 M 9.2 Tohoku earthquake. Figure 8.42 shows the surface projection of contours of slip on the subduction interface (Simons, Minson, Sladen, Ortega, Science,). This slip model was constructed to simulate co-seismic GPS displacements combined with tsunami waveforms, which are primarily determined by the spatial pattern of coseismic uplift. Although the inferred slip is large (as much as 40 m in the hypocentral region), the inferred rupture history can be interpreted as a very large slip pulse. The Simons and others model appears to show that the rupture did not extend all of the way to the ocean floor. Unfortunately, the model resolution in the shallowest region (at the trench) is poor since the displacement data is only available from on-shore GPS stations. Furthermore, in the limiting case of zero depth, there are no radiated seismic waves for the case of a horizontal fault. That is, slip that occurs at the upper edge of a subduction earthquake is practically invisible for seismic studies.

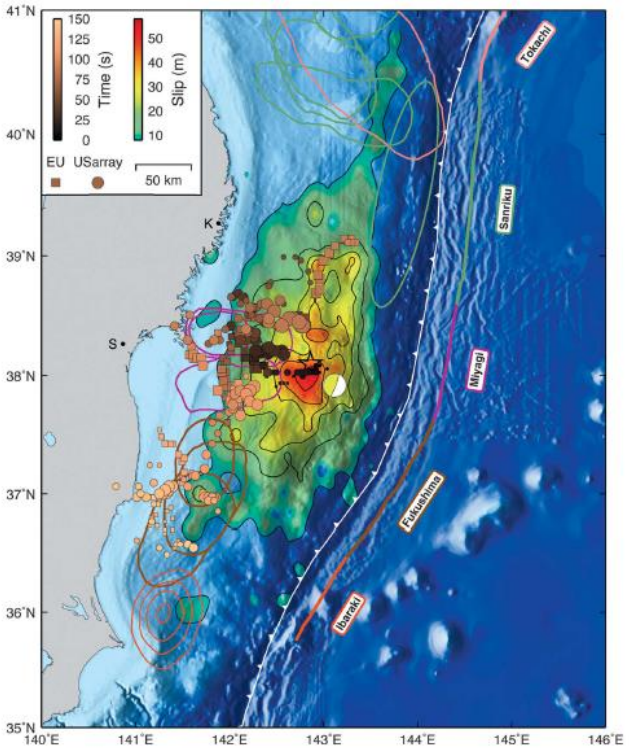


Figure 8.42

There was a lucky break for researchers studying the Tohoku earthquake. Ocean bottom observatories TJT1 and TJT2 had been installed just west of where the mega-thrust fault intersects the ocean floor (see Figure 8.43). Sonar positioning had been used to calculate the position of these stations on the ocean floor. Although the pressure records from the earthquake were of interest, the most amazing observation came from relocating the stations after the Tohoku earthquake. In particular, the change in position of TJT1 and TJT2 was determined to be 62 $\pm$ 20 m and 53 $\pm$ 20 m, respectively. In addition, the pressure gauges indicated that the stations had been uplifted by 5 m (Ito and others, 2011). It seems as though there was exceptionally large slip on the very shallowest part of the rupture. Ito and others showed that this large shallow slip was consistent with a large pulse in tsunami height that was recorded on two ocean bottom pressure sensors, TM1 and TM2 which are located closer to the Honshu coast (see Figure 8.43). Ito and others summarize their findings in Figure 8.44; they concluded that the slip in the 50-km wide shallow part of the rupture (average depth of 4 km from the ocean bottom) was about 80 m.

Remarkably, there was no evidence for this enormous slip to be found in the short-period accelerograms recorded on either Honshu or in the teleseismic body

waves. Of course, this is what one would expect from crack-like rupture; the high-frequency radiation is only from the propagating crack tip. Furthermore, in the limit of a horizontal fault that becomes very shallow, there are no radiated waves of any sort (including surface waves). That is, very shallow thrusting on very low-angle planes is essentially invisible to seismometry. Unfortunately, this faulting geometry can cause static uplift of the ocean bottom, which can cause large damaging tsunamis. Shallow crack-like slumping at an ocean trench seems to be a logical explanation for **tsunami earthquakes,** which are a class of nearly silent earthquakes (or slumps) that generate dangerous tsunamis (see McKenzie and Jackson, 2012, Tsunami earthquake generation by the release of gravitational potential energy, EPSL, 345-348, 1-8).

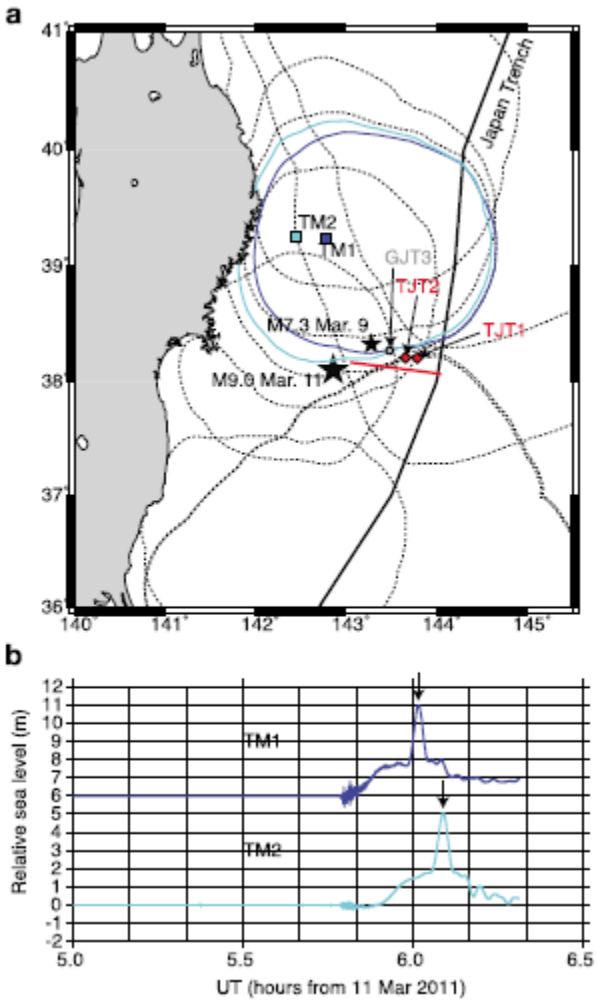


Figure 8-43. Locations of ocean-bottom observatories operating during the 2011 Tohoku earthquake (Ito, Y, T. Takeshi, Y. Osada, M. Kido, D. Inazu, Y. Hayashi,

H. Tsushima, R. Hino, and H. Fujimoto, 2011, Frontal wedge deformation near the source region of the 2011 Tohoku-Oki earthquake, Geophys. Res. Lett., 38, L00G05, doi:10.1029/2011GL048355).

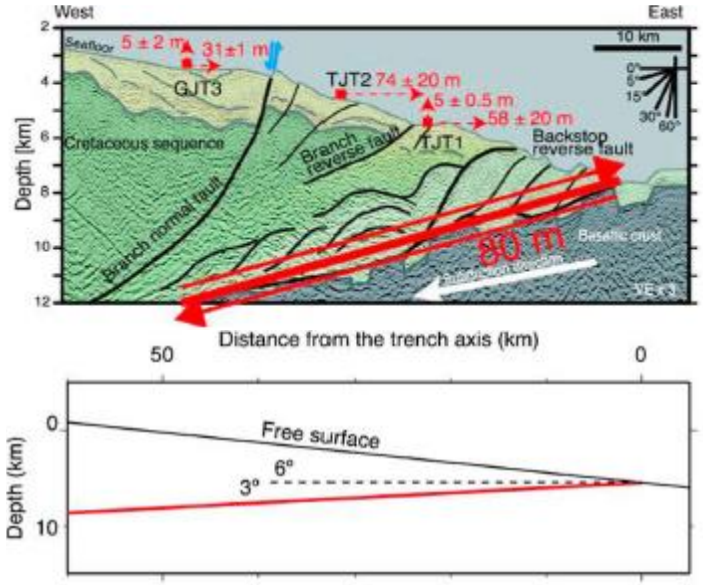


Figure 8-44. Schematic showing the region of inferred large slip near the Japan Trench for the 2011 Tohoku earthquake (Ito and others, 2011). The dip of the megathrust is so shallow that it appears that the shallowest part of the earthquake can be interpreted as a gravitational slump.

While it is uncommon to find direct evidence for crack-like rupture in earthquakes, crack-like rupture might also help to explain puzzling features of the enigmatic 1960 M 9.5 Chilean earthquake. Crack-like rupture produces very large slips that are spatially smooth; the ratio of short-period to long-period radiated energy is small for crack-like ruptures as compared to pulse-like ruptures of comparable potency.

Figure 8.45 shows the peak amplitude of broad-band (1 s to 90 s) teleseismic P-waves for the largest earthquakes of the 20th Century. M_W for these events is based on the potency (aka seismic moment) which was estimated from a variety of long-period measurements that include geodetic positioning, tsunami excitation, and excitation of free oscillations. Notice that, with the exception of the 1964 M 9 ¼ Alaskan earthquake, the largest earthquakes have long-period body waves that saturate with magnitude. This behavior is inconsistent with standard spectral scaling laws. It presents us with a mystery. Why are the potency rate functions of the largest earthquakes smoother than other earthquakes? Figure

8-46 shows potency-rate functions inferred for these largest events. These time functions were derived from teleseismic P-waves recorded by the long-period Benioff seismometer in Pasadena, which has a response that is flat to velocity between 1 s and 90 s (see chapter 3). In the case of the largest earthquakes (M > 9), it appears that there is a large portion of the potency rate function that is not visible in the 1 to 100 sec teleseismic P-waves.

One plausible explanation is that many of these largest events may have sections with crack-like ruptures, which are characterized by large, spatially-smooth slips.

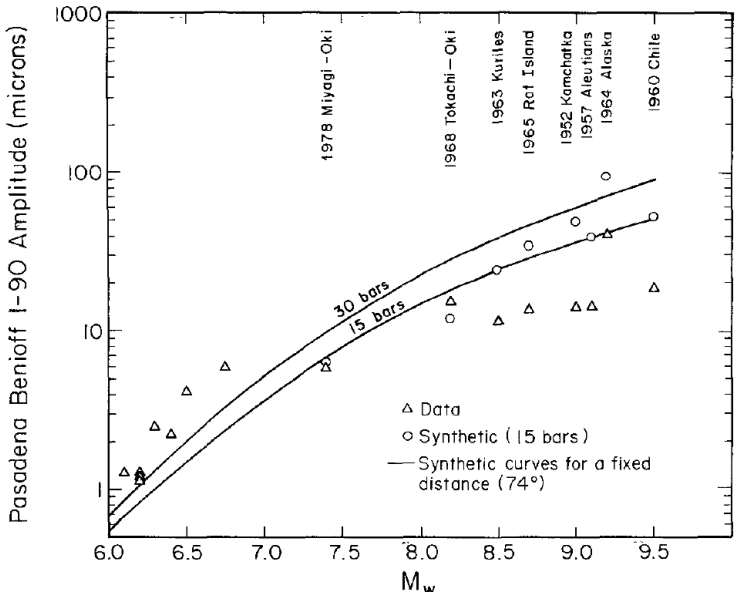


Figure 8-45. Maximum distance corrected P-wave amplitude on the Pasadena broad-band seismograph (aka, Benioff 1-90) for the largest earthquakes of the 20th Century. The triangles are observed data and the circles are predicted from standard f-2 spectral scaling. This scaling seems to match the observations of most earthquakes, except those with M > 9. Hartzell, S., and T. Heaton, 1988, Failure of self-similarity of large shallow subduction earthquakes, Bull. Seism. Soc. Am., 78, 478-488. http://resolver.caltech.edu/CaltechAUTHORS:20121121-115204715

Figure 8.46 shows Potency Rate functions that are 1) compatible with the observed 1-90 records, 2) have potencies that are compatible with estimates using free oscillation data, and 3) that have overall durations that are inferred from the source dimensions and typical rupture velocities (Hartzell, S, and T. Heaton, 1985, Teleseismic time functions for large shallow Subduction zone earthquakes, Bulletin of the Seismological Society of America , v. 75, pp. 965-1004. http://resolver.caltech.edu/CaltechAUTHORS:20130130-145105862).
The apparent saturation of intermediate-period body waves is also recently reported by Kanamori and Ross (Reviving m_B, *Geophys. J. Int.* ,2018, **216,** 1798–1816 doi: 10.1093/gji/ggy510).

The bottom line seems to be that there is evidence that slip pulses can transition into growing cracks if the pulse becomes large enough (perhaps slips > 20 m are predominantly crack-like).

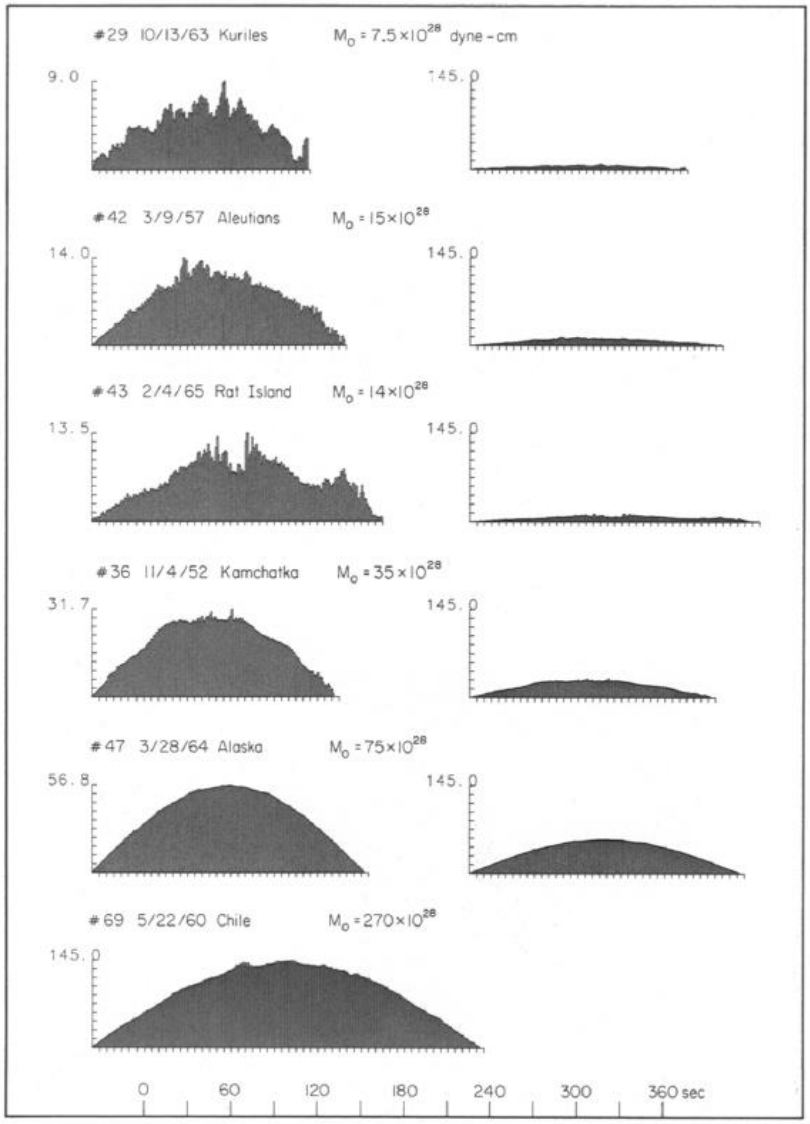


Figure 8. 46

Initiation of cracks

At the other end of the scale, earthquake nucleation is almost certainly a crack-like process. In order to have dynamic instability (an earthquake), it is necessary that the rate at which potential energy (usually strain energy) is removed from the system must exceed the rate of energy dissipation in inelastic processes (friction). For example, see equation 8.95, which estimates the minimum area that a crack-

like rupture can experience dynamic instability. When Dieterich-Ruina friction is assumed for slowly increasing shear stress (e.g., plate motion), then inelastic strains tend to concentrate in an area that eventually becomes the hypocenter of an expanding crack. This temporal evolution is a consequence of the evolving state variable, and the nucleation process may accelerate over days in a small patch. The important point is that ruptures cannot initiate as slip pulses; the rupture area of the slip pulse would be smaller than the critical length for instability.

Transition from cracks to pulses

Assuming that all earthquakes initiate as growing cracks, and that most significant shallow crustal earthquakes appear to be propagating pulses, leads to the conclusion that there must be some length where there's a transition from crack-like rupture to pulse-like rupture. Currently, finite fault modeling of seismograms is the methodology that allows us to image slip pulses. This detailed 3-d imaging is more difficult when the distance to the observation is large compared with the dimension of the rupture. That is, it has not been feasible to image the ruptures of earthquakes smaller than M 5 ½ with enough precision to resolve slip pulses.

Another way to investigate the transition of cracks to pulses is shown schematically in Figure 8.47 (from Meier, Heaton, and Clinton, 2016). This schematic shows several circular arcs on a rupture plane. The arcs are meant to show crack (or healing) fronts at different times. The shaded areas show the slipping region for several different times. For areas near the point marked as nucleation, all of the area is slipping in the moments after the rupture front has passed; it's an expanding crack. At later times, the rupture front has propagated further from the hypocenter and the slip is now confined to a pulse.

The schematic shows expected slip profiles at different times. When the earthquake is crack-like, we expect to see spatially smooth slip profiles, whereas when it's a pulse, we expect to see spatially heterogeneous slip; I'll come back to that to explain why pulses are inherently unsteady. Panel d is a schematic of the potency rate function that one would expect from such a conceptual model. Remember that the far-field displacement should look like the potency rate function.

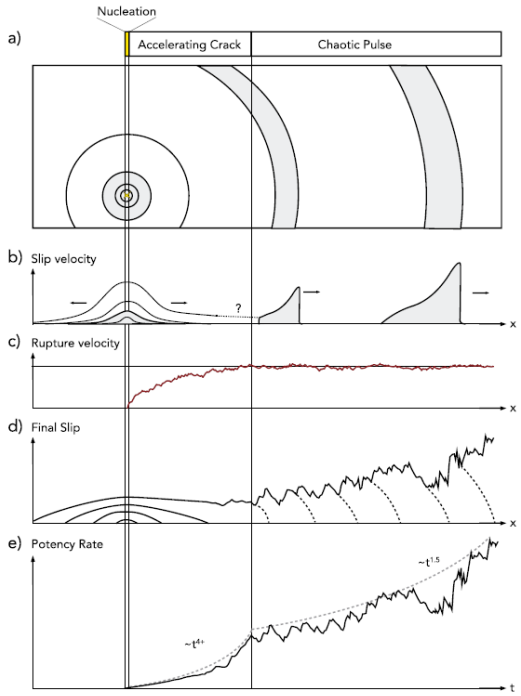


Fig. 8-47 Cartoon showing concept for transition between crack-like and pulse-like rupture (see text).

Meier, Heaton, and Clinton investigated a large set of near-source displacement records (doubly integrated from accelerograms). The records were all at hypocentral distances of less than 25 km and they were all processed in the same way. The records are sorted into magnitude bins and then the log of the median peak displacement for each magnitude bin is plotted as a function of log time (see Fig. 8-48). The use of a log-log plot allows us to look for power-law behavior. Time is defined with respect to the first P-wave. Pay close attention to the logarithmic time scale; the first decade is only one tenth of a second and the center of the plot is just 1 second. The S-wave pick times are displayed at the bottom of the plot as single points for each record. Notice that the S-waves do not begin to arrive until several seconds after the P-arrival. Also notice that for the first 0.2 s, the median motions are approximately the same, independent of the magnitude. Furthermore, the pgd's grow quickly and increase as approximately t^3. We interpret this initial rapid growth as crack-like rupture in the hypocentral region. There also appears to be a change in the pgd growth rate at about ¼ s. That is, pgd seems to grow as t^3 until it transitions to a growth rate of about $t^{1.3}$ for times greater than ¼ s. Of course, earthquakes smaller than M 4.5 seem to be all over by ¼ s, so the transition to $t^{1.3}$ growth is only observed for earthquakes larger than M 4.5.

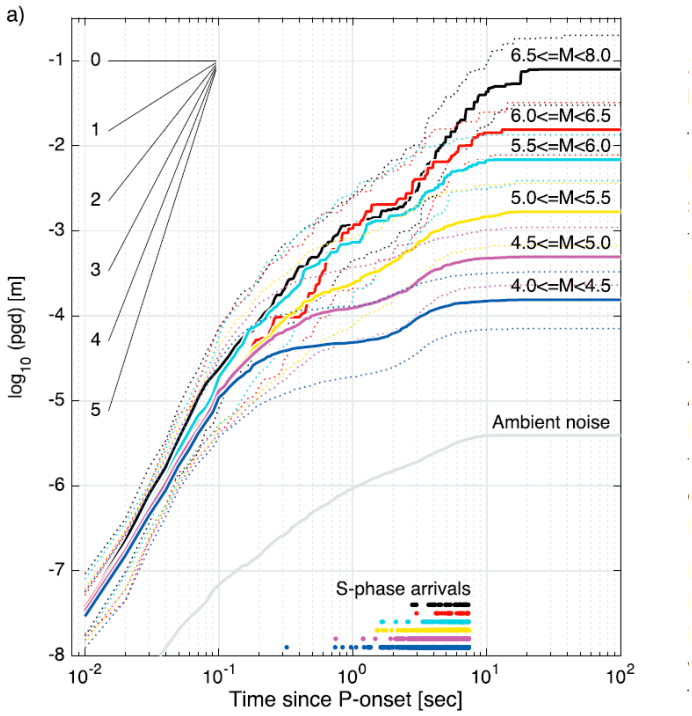


Fig. 8.48

I interpret this break in growth at 0.25 s (corresponds to M 4.5) as the transition from cracks to pulses. I admit that this evidence is suggestive and not compelling. If this is the transition from cracks to pulses, then we can estimate the rupture length at which the transition occurs. For crack-like rupture, the total event duration is about twice the time required for the rupture to spread, or $T_{event} \approx 2\sqrt{S}/V_R$. $L_{crack-pulse} \approx \frac{1}{2} V_R (0.25s)$. If we use a typical rupture velocity of 2.5 km/s, then the transition would occur at a rupture length on the order of 250 m. Unfortunately, the assumption of a length-scale independent rupture velocity is not at all obvious. That is, it seems plausible that the rupture velocity increases as the crack grows. That possibility of increasing rupture velocity seems difficult to assess and I will ignore it for the present. Assuming that the dimension of the rupture at which this break in scaling is 250 m, then the rupture area would be on the order of 6×10^4 m^2 (assuming that the rupture approximately square). We can use 8.22 to estimate that that corresponds to a magnitude 3.0 earthquake. This estimate is similar to the M 4.5 at which the scaling changes in Fig. 8.48. Using 8.30, we can estimate the total change in potential energy as

$$\begin{aligned}\Delta E_W &\approx SD\left(\sigma_0 - \frac{1}{2}\Delta\sigma\right) \\ &\approx 3\times10^{-5} S^{3/2}\left(\sigma_0 - \frac{1}{2}\Delta\sigma\right)\end{aligned} \qquad 8.106$$

Or,

$$\frac{\Delta E_W}{S} \approx 3\times10^{-5}\sqrt{S}\left(\sigma_0 - \frac{1}{2}\Delta\sigma\right) J/m^2 \qquad 8.107$$

If the rupture area is 6×10^4 m^2, then $\frac{\Delta E_W}{S} \approx 2.4\times10^{-2}\left(\sigma_0 - \frac{1}{2}\Delta\sigma\right) J/m^2$. If the prestress σ_0 is a typical Byerle stress of 200MPa, and if the stress drop is the average of 3 MPa. Then the change in potential energy of a M 3.0 is about 5 MJ/m^2. This enough energy to melt several mm's of fault material. This potential energy change can be compared with the radiated energy that is estimated using 8.57. The following is from an unpublished note from Hiroo Kanamori. I lightly modified his equations which were written in the familiar form of seismic moment and assuming a circular rupture. I will assume the standard f^{-2} model

$$\begin{aligned}E_R(f) &= 8\pi P^2 f_c^4\left(\frac{\mu^2}{15\rho\alpha^5} + \frac{\mu^2}{10\rho\beta^5}\right)\int_0^f \frac{f^2}{\left(f^2 + f_c^2\right)^2} df \\ &= 8\pi P^2 f_c^4\left(\frac{\mu^2}{15\rho\alpha^5} + \frac{\rho}{10\beta}\right)\int_0^f \frac{f^2}{\left(f^2 + f_c^2\right)^2} df \\ &= 8\pi P^2 f_c^4 \rho\left[\left(\frac{\beta}{\alpha}\right)^4 \frac{1}{15\alpha} + \frac{1}{10\beta}\right]\int_0^f \frac{f^2}{\left(f^2 + f_c^2\right)^2} df\end{aligned} \qquad 8.108$$

The S-wave energy is large compared to the P-wave, so neglect the P-wave.

$$\begin{aligned}E_R(f) &\approx 0.8\pi P^2 f_c^4 \rho/\beta \int_0^f \frac{f^2}{\left(f^2 + f_c^2\right)^2} df \\ &= 0.8\pi P^2 f_c^3 \rho/\beta\left[-\frac{f f_c}{2\left(f^2 + f_c^2\right)} + \frac{1}{2}\arctan\left(\frac{f}{f_c}\right)\right]\end{aligned} \qquad 8.109$$

If $f_1 \gg f_c$, then

$$E_R\left(f_1 < f < \infty\right) \approx \frac{2\pi\rho P^2 f_c^3}{5\beta}\frac{f_c}{f_1} \tag{8.110}$$

Using the standard Brune model,

$$f_c \approx 0.42\beta\left(\frac{\Delta\sigma}{\mu P}\right)^{1/3} \tag{8.111}$$

So

$$\begin{aligned} E_R\left(f_1 < f < \infty\right) &\approx \frac{2\pi\rho P^2 C^3\beta^3\left(\frac{\Delta\sigma}{\mu P}\right)}{5\beta}\frac{C\beta\left(\frac{\Delta\sigma}{\mu P}\right)^{1/3}}{f_1} \\ &= \frac{2\pi\rho P^2 C^4\beta^3}{5 f_1}\left(\frac{\Delta\sigma}{\mu}\right)^{4/3} \\ &\approx 0.04\rho P^2\beta^3\left(\frac{\Delta\sigma}{\mu}\right)^{4/3}\frac{1}{f_1} \end{aligned} \tag{8.112}$$

To compare the radiated energy per unit rupture area with the potential energy per unit rupture area, we need to divide 8.112 by the rupture area

$$S = \left(\frac{7\mu P}{16\Delta\sigma}\right)^{2/3} \tag{8.113}$$

And we obtain

$$\frac{E_R\left(f_1 < f < \infty\right)}{S} \approx 2\times 10^{-2}\frac{\Delta\sigma^2}{\rho\beta f_1} \tag{8.114}$$

If we assume that $\Delta\sigma = 3\text{MPa}$, $\rho = 3\times 10^3\ \text{kg/m}^3$, and $\beta = 3.5\ \text{km/s}$, then $\frac{E_R\left(f_1 < f < \infty\right)}{S} \approx 20\frac{1}{f_1}$ kJ/m^2. However, the 3 MPa stress drop is a number that is appropriate to simulate static stress drop. Our current derivation is designed to simulate near-source high-frequency motion. In this case, it seems more appropriate to use effective stress instead of static stress drop. Hanks and McGuire used near-source acceleration records to conclude that effective stress is about 10 MPa. Inserting 10 MPa into 8.114 gives an estimated radiated energy of

2MJ/m^2. Clearly, there is ambiguity in obtaining the radiated high-frequency energy estimate. However, I would say that it's in the range of 1 MJ/m^2.

In order to form a slip pulse, there must be enough potential energy change to sustain the high-frequency radiated energy in a slip pulse. If we say that the high-frequency energy (f > 2 Hz) radiated by a slip is described by a white noise acceleration spectrum (this is basically Brune's model), then we can use 8.114 to provide a very rough estimate of the radiated high frequency energy. Since the rupture area of a slip pulse is small compared to a crack, the estimate may be larger than that just provided (as large as 200 kJ/m^2??). Our earlier estimate of potential energy change of 5 MJ/m^2 for a M 3 indicates that there is abundant available energy to support the radiation of high frequency waves (greater than 2 Hz) that are observed in the near-source region of slip pulses.

The Sound of Slip Pulses

I have used the phrase, "white noise," without providing much definition or context. The *"colors of noise"* is a concept that originates in audio engineering. It has been observed that many forms of noise have random phase with a power spectrum that is approximately described by a power law $f^{2\alpha}$. Recall that the power spectrum is an energy spectrum. That is, if the power spectrum is $f^{2\alpha}$, then the amplitude spectrum is f^{α}. In the case of sound, the amplitude of sound is proportional to acoustic pressure, which is proportional to the particle velocity.

In audio engineering, the term, white noise, refers to a sound that has equal energy in all frequency bands ($f^{2\alpha=0}$, just like white light). In this case, the amplitude spectrum is also flat and white. A good example of the sound of white noise is the sound that an old-fashioned AM radio makes when it's tuned between stations (sometimes called shot noise, or sometimes it's called "static"). If the amplitude spectrum of the particle velocity is constant, then the amplitude spectrum of the particle displacement is f^{-1}, and the spectrum of the particle acceleration is f^{1}. Now the observation of the "sound of a slip pulse" is that it has constant particle acceleration amplitude, or the velocity amplitude spectrum is f^{-1}. This gives a power spectrum that is f^{-2}. Noise of this type is typically called red noise. This form of noise is quite common in physical systems. For instance, the time integral of randomly-timed force impacts (e.g, Brownian motion) produces f^{-1} red noise. Red noise sounds very much like the roar of a waterfall.

It seems that slip pules are accompanied by white-noise acceleration with a total radiated energy on the order of 1 MJ/m^2. It's as if slip pulses have a distinctive sound (the sound of a waterfall). Furthermore, this noise seems to be consistent with the notion that the slip pulse is characterized by Brownian-motion-like impacts from a fluidized granular material.

Modeling Slip Pulses

In my 1990 paper, I presented observational evidence that earthquake rupture is pulse-like and I then presented a conceptual model of dynamic friction that I claimed would produce slip pulses. In particular, I hypothesized that dynamic friction was inversely proportional to slip speed, which is a particular realization of the friction law in equation 8.99. At the time that I wrote the slip pulse paper, I searched for an analytic solution for a steady slip pulse. The only solution that I found is a solution by Freund (1970), although pulse-like ruptures were first introduced by Joffe. You might hear the terms, "Joffe pulse," or sometimes "Heaton pulse"; these are the same thing as a slip pulse

Fig. 8.50 shows Freund's solution for a mode III slip pulse of length l_p that is propagating at constant velocity V_R . The shear stress at infinity is uniformly equal to τ_0 and the shear stress on the slipping patch is τ_p. The dynamic stress drop of the pulse $\Delta\sigma_p$ is given by

$$\Delta\sigma_p \equiv \tau_0 - \tau_p = \frac{\mu}{\pi}\frac{D}{l_p}\sqrt{1-\left(\frac{V_R}{\beta}\right)^2} \qquad 8.115$$

Which implies that

$$D = \Delta\sigma_p l_p \frac{\pi}{\mu}\frac{\beta}{\sqrt{\beta^2 - V_R^2}} \qquad 8.116$$

Or alternatively,

$$l_p = \frac{D}{\Delta\sigma_p}\frac{\mu}{\pi}\frac{\sqrt{\beta^2 - V_R^2}}{\beta} \qquad 8.117$$

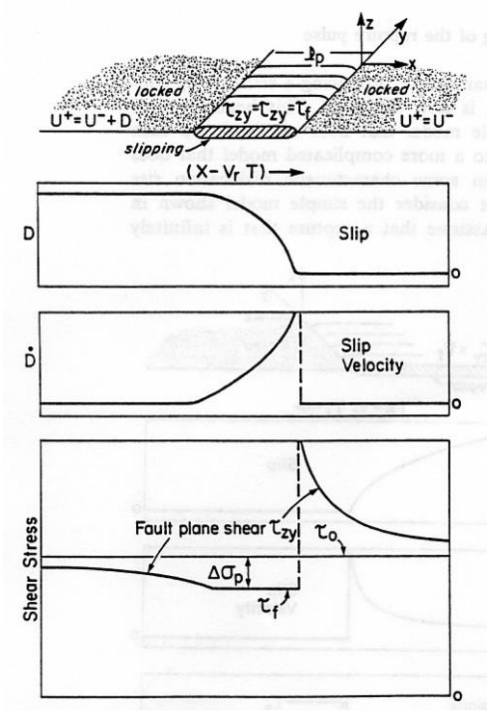


Figure 8-49. Idealized model of Freund (1970) in which a pulse of slip propagates steadily at velocity V_R. A uniform shear stress τ_0 is applied at infinity and the shear stress on the slipping portion of the fault is assumed to be to τ_f .

This solution is obtained by the appropriate transformation of the static solution of a mode III crack of length l_p . This solution has numerous quirks that limit its usefulness. In particular, if the rupture velocity approaches the shear velocity, then, for a given slip, the length of the slip pulse tends to zero. Or alternatively, given the length of the slip pulse and the slip, then the dynamic stress drop of the slip pulse tends to zero as the rupture velocity approaches the shear wave speed. These features are the result of singular stress at the crack tip that results from instantaneous stress drop (no fracture energy). Presumably, the stress at the leading edge of the slip pulse is approximately the shear stress that is required to overcome static friction (i.e., Byerlee friction).

As was mentioned earlier, there is no inelastic work at the crack front (fracture energy) and the crack front has singular stresses that cannot be handled numerically. The second issue is that this friction law has no stable steady-state solutions. In particular, a steady-state pulse should propagate at a constant velocity and with a slip pulse of constant slip

Another strange aspect of this solution is that the static stress drop is zero. That is, the pulse propagates indefinitely, so the total rupture length is infinite while the slip is a constant. Finally, since this solution is obtained from a transformed static solution, there are no radiated waves generated by this slip pulse. This turns out to be an important **feature of a slip pulse that propagates steadily. It is invisible to seismology. A slip pulse only radiates far-field waves when 1) either it changes its amplitude as it propagates, or 2) it changes its rupture velocity.**

You can obtain further insight in this by reading the analysis associated with Figures 7.12 and 7.13. Interestingly, this invisibility of steady-state slip pulses means that we **cannot resolve the upper limit of slip velocity from seismic data.** That is, you can only observe the details of a slip pulse if you are so close to the pulse that you can observe the near-field terms (basically the static part of the solution). Since seismic measurements are only available from the Earth's surface, it is practically impossible to observe the true nature of slip pulses that propagate at depth greater than a couple of km. However, it is possible to put a lower bound on the average slip velocity. In particular, the duration of pulse-like ground velocities in the forward directivity direction provides a lower bound on slip velocity. Since directivity pulses are only observed for mode II ruptures (they are on the fault normal component of stations located near a fault with the provision that the rupture is propagating parallel to the slip direction (i.e., mode II). Modeling of records of this type indicates that the length of slip pulses is typically less than 10% of total rupture length.

It appears that the average slip velocity in slip pulses is greater than 2 m/s. Since the elastic strain in shear waves is on the order of the particle velocity (1/2 the slip velocity) divided by the wave speed (see Chapter 3), we can conclude that a particle velocity of 1 m/s in a material with a shear wave speed of 3 km/s implies a shear strain of about 3×10^{-4}. Therefore, a slip velocity of 20 m/s implies a shear strain of about 3×10^{-3}, which is about the maximum shear strain that rocks can sustain at the confining pressure of earthquakes. While 20 m/s slip velocities (at 10 km depth) is an order of magnitude faster than is typically assumed for earthquakes**, it is not possible to exclude the possibility of slip velocities as high as 20 m/s**. Again, high slip velocity in a slip pulse that is steadily propagating at depth does not necessarily produce large high-frequency radiation.

It is important to recognize that the solution given by 8.115 is not a "self-healing" pulse. It is only a slip pulse because I constrained the fault rupture to be pulse-like. In my 1990 slip-pulse paper, I hypothesized that friction that scales with inverse slip velocity should produce slip pulses. As it turns out, it has not been possible to simulate the behavior of two half spaces that are governed by a friction law that varies inversely with slip speed.

While a solution like that shown in Fig. 8.24 can match boundary conditions, it is not a stable solution. In order to achieve a stable solution, then it must be that slight perturbations from homogeneity in the prestress should have little effect on the solution. Assume that you have found a slip pulse that will propagate steadily at a particular homogeneous prestress. Now assume that some slight perturbation occurs as the pulse runs (e.g. a tiny patch with higher prestress). When the pulse hits the perturbation, then the pulse temporarily begins to grow larger, which means that it has slightly larger slip speed. However, larger slip speed means that the friction drops, which causes the pulse to grow even larger. That is, if you put a tiny positive perturbation to a steadily running slip pulse, then it grows spontaneously without bound. Similarly, if you perturb the solution such that it temporarily becomes smaller, then it quickly shrinks to zero.

Rice and Lapusta recognized this difficulty and they concluded that a friction that is purely rate weakening would produce slip pulses with unstable solutions (Rice, J. R., Spatio-temporal complexity of slip on a fault, J. Geophys. Res., 98(B6), 9,885–9,907, 1993). They concluded that pure rate weakening friction is a pathological mechanics problem. They said that this problem is ill posed. By that they meant that there was insufficient information to construct a solution to the problem. In particular, there is no length scale to rate weakening friction; while a stable slip pulse has a length scale associated with it (see Chapter 5 of Ahmed Elbanna's PhD dissertation for further discussion).

Aagaard and Heaton (2008, Constraining fault constitutive behavior with slip and stress heterogeneity, J. Geophys. Res., 113, B04301, doi:10.1029/2006JB004793) studied this problem using 3-d finite element models of sliding faults with a friction law that had both slip weakening (fracture energy) and rate weakening. They were unable to simulate rupture using pure rate weakening (i.e. zero fracture energy) since the solutions became singular at the crack tips; it was not possible to attain numerical stability. Based on this knowledge, they hypothesized that, if there is a solution to pure rate-weakening friction, then the solution would have to be described by a fractal, since fractals and homogeneous solutions are the only classes of solution that have no inherent scale.

Aagaard and Heaton were searching for ways to produce earthquakes that had self-sustaining heterogeneity. That is, they observed that most finite fault source inversions resulted in slip models that 1) were characterized by slip pulses, and 2) the slip models were spatially complex. Aagaard and Heaton attempted to find ways to produce self-sustaining spatial heterogeneity that would persist through many millennia of earthquakes. Although they were able to find spatial distributions of frictional properties that would produce spatially heterogeneous slip in response to plate motions, they found that the residual stress that was left after the heterogeneous event was then spatially correlated with the heterogenous friction. The result was that the next event in the sequence was spatially

homogeneous. Aagaard and Heaton wanted to show that strong rate weakening could produce earthquakes that had 1) self-sustaining heterogeneous slip, 2) slip pulses, and 3) that required very little inelastic dissipation energy (friction and fracture energy). Unfortunately, achieving numerical convergence with low fracture energy and with strong rate weakening was not possible with even the largest super computers. When they recognized that the number of computations required for this calculation grows as $1/\Delta x^4$, they realized that finite element simulations could not solve this problem. Worse, the simulation of strong rate weakening and low dissipation requires that the model should evolve into a prestress that is spatially complex. In particular, the prestress must be fractal (I'll get back to that later). This makes the numerical simulation problem even more overwhelming.

Aagaard and Heaton found that they could produce more spontaneous slip heterogeneity by increasing the strength of the rate-weakening in their friction law. A general observation that comes from numerical finite element modeling of rupture dynamics is, if you assume relatively homogeneous prestress, then the change in potential energy per unit rupture area grows as the rupture length grows. The only parameter that is available to absorb all the potential energy is the fracture energy; the fracture energy must exceed the radiated energy if you want to control the dynamics of an earthquake. I have seen numerous dynamic models with enormous fracture energies. My opinion is that these reported fracture energies are simply introduced as a way to avoid numerical instability in the calculation of stress at the rupture fronts. I will shortly argue that prestress is extremely heterogeneous and that this prestress heterogeneity serves to limit the growth of dynamic ruptures.

Following the realization that there are fundamental difficulties with 3-d finite-element modeling dynamic ruptures assuming strong rate-weakening friction, Ahmed Elbanna and I took a different approach. We decided to study the impact of strong rate weakening friction on a 1-d spring-block-slider model (sbm). The sbm (shown in Fig. 8.50) consists of a number $(i=1,...,m)$ of identical point masses that are connected with identical leaf springs (stiffness k_l) to rigid blocks (above and below) that move with a steady differential velocity v. Additionally, the blocks are connected to each other by identical coil springs (stiffness k_c). Finally, there is uniform rate-weakening friction between the masses and the lower rigid block ($F_{fi} = \frac{1}{1+b\dot{u}_i}$). Dynamic friction is specified to occur at any block that exceeds a specified maximum force. While you might think of this maximum stress in a block is the "strength" of this system, it seems that changing the value of this maximum has very little effect on the overall behavior of this system.

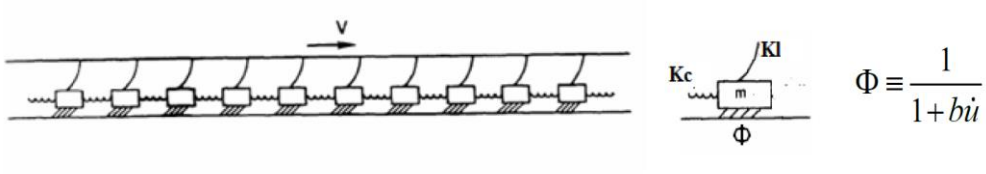


Figure 8.50

The sbm has several tunable parameters that control the system behavior. In particular, if the coil springs are stiff, then there is strong coupling between the masses and when failure occurs, it tends to involve all of the blocks. In fact, events tend to be crack-like in the sense that all bocks are moving during an event. As I have already mentioned this friction law provides a positive feedback in the dynamics of this problem. That is, the faster the sliding, the lower the friction which produces even faster sliding. Positive feedback dynamics is often a characteristic of chaotic systems. For example, the stock market is a highly complex system. When stock prices are rising, people want to buy so they can get in on the profits. Conversely, when stock prices are falling, people want to sell before they lose their investment. Wild fires are another example of a positive feedback system. The larger a fire, the faster it spreads. Epidemics are another example of a positive feedback system. All of these systems have complex behavior that is difficult to predict. In fact, often the behavior is so complex that we resort to random variables to describe them. However, it is important to recognize that complexity does not necessarily imply random. In fact, surprising structure can spontaneously appear as the result of chaotic dynamics.

Elbanna and I chose to investigate the sbm model since it is simple enough to explore with modern computers, while it still has characteristics that are of interest for earthquake dynamics. In particular, it is possible to explore the consequences of strong rate weakening friction using a sbm model. One reason that sbm's are amenable to computer simulations is that they are nearest-neighbor models. That is, motion of a single block only affects the adjacent blocks. This contrasts with an elastic continuum; slip on a fault segment causes stress changes throughout the entire medium. The fact that the motions at a finite-element node are affected by the motions of all other nodes is responsible for the enormous size of 3-d continuum calculations. By their very nature, slip pulses are short-range phenomena, and it's critical to have a numerical scheme that concentrates on short-range phenomena if there is any hope of studying them.

There are two distinct stiffnesses in the sbm, the coil springs and the leaf springs. If the friction is dropped to zero, then the coupled blocks vibrate through traveling waves. These waves are dispersive because of the two stiffnesses. If the system is frictionless and composed of only coil springs, then the system is

nondispersive and the sound wave velocity is, $c_s = \sqrt{k_c / m}$. If leaf springs are added to the frictionless system, and if it is excited by harmonic forcing of frequency ω, then the wave speed is dispersive and is given by

$$c_p = c_s \sqrt{1 - \frac{k_l}{(\omega c_s)^2}} \qquad 8.118$$

Fig. 8.51 shows the spatio-temporal characteristics of a typical slip pule in an sbm with velocity weakening friction. Notice that the pulse is very localized as it runs (pulse width is 5% of the rupture length). Furthermore, the pulse seems to be nondispersive and running at the sound velocity (since the springs are linear, there is no way for disturbances to jump to supersonic velocities). The apparent constant rupture velocity is typical of most events in this system and it is a key feature that allowed us to develop the slip-pulse energy equation that allows us to simulate this system with a single equation (more later on the pulse-energy equation).

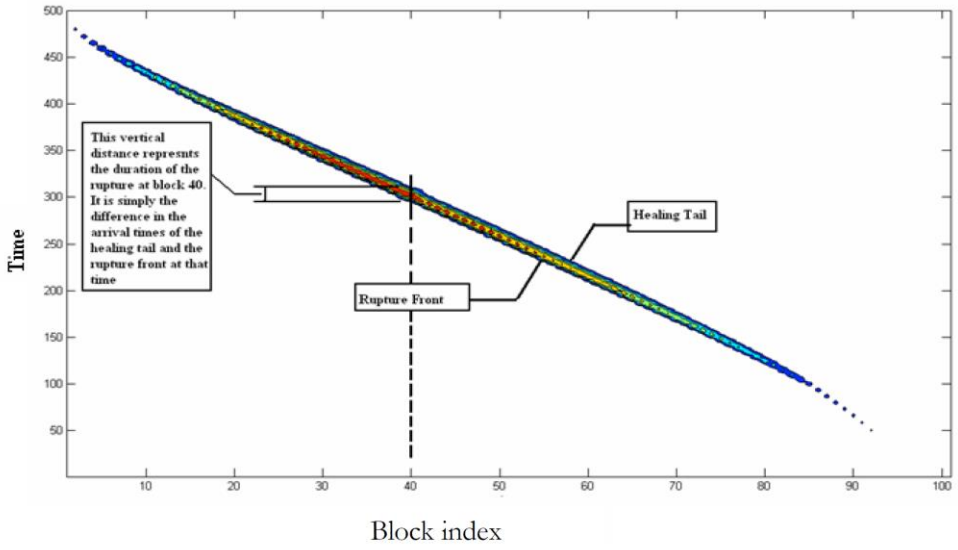


Figure 8.51

Even though all of the springs, masses, and friction laws are identical for every block in our model, the individual events are quite complex. Typically, we initiate the system by assuming a higher pre-stress in some group of blocks. Although individual events can become quite large, they do not grow without bound since the energy in the leaf springs that drives the system has an upper bound. After a large number of events have occurred, the system attains an unusual kind of steady state. That is, the statistical properties of the events and the prestress no

longer change as more events occur. That is, the systems evolve towards a **strange attractor**. In dynamics, an attractor is a state that a system tends to evolve to independent of initial conditions. An attractor is called strange if is described by a fractal. An example of the nature of the prestress that has evolved after tens of thousands of events is shown in fig. 8.52.

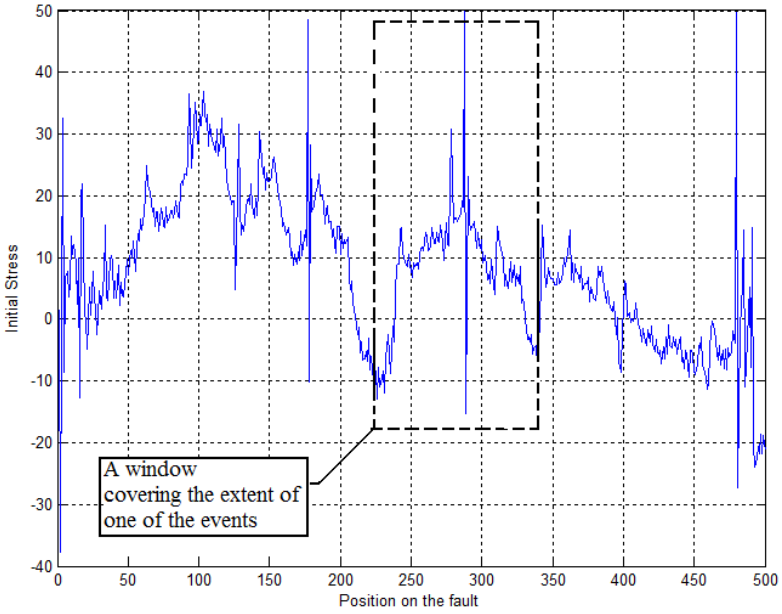


Figure 8.52

The model is unitless and the maximum shear before dynamic failure is 50 units. In this example, the next event starts as an expanding crack at the narrow spike at block 290 and it then quickly becomes two slip pulses that propagate bilaterally until they arrest at block 225 on the left and block 345 on the right. Notice that the pulses propagate at the relatively low prestress of 10 units and that they arrest after propagating across regions of **negative shear stress**. Many colleagues have been surprised that a frictional system that is being forced with right-lateral shear can evolve patches of left-lateral shear stress. In the case of slip pulses in an sbm, the answer is quite simple. As a slip pulse propagates, it can accumulate large kinetic energy (i.e., large slips). Because of the strong feedback in the friction law, slip pulses can also arrest very abruptly. When this happens, there can be a patch of very large slip that is frozen in when the event is over. The slip may be locally large enough that it causes shear stress in the opposite direction of the motion of the rigid plates driving the system.

Even though there are patches of negative shear in this self-organized prestress, there are never any events that have slip in the opposite direction of forcing. That is because events only nucleate when the stress exceeds a threshold and this only happens when the prestress has the same polarity as the driving stress. This

same principle applies to the Earth. If there are patches of negative prestress (e.g., left-lateral stress on the San Andreas fault), then they are mostly invisible to seismology since such patches will not nucleate future events. For more discussion on this subject, see later in this chapter (Smith and Heaton.)

Figures 8.53 shows an example of the types of events that occur after the system has evolved into a strange attractor. We call this example Case A. In each of these plots, a sequence of thousands of consecutive events are plotted. Each event (event index is the horizontal axis) shows the blocks that moved during that event. Fig. 8.53 shows events number 62,000 through number 66,000. Most of these 4,000 events involve only short ruptures, although there are many long ruptures as well. If you stare at this sequence, it would be easy to convince yourself that there are likely to be some blocks with different properties that control the locations of events. Of course, this is not the case; everything is uniform in the model.

There are also persistent local knots that experience a large number of mostly short ruptures. These knots sometimes persist even after a long rupture extends through the knot. Inevitably, though, these knots are transitory, and some future large event erases them and causes new knots to appear. It's important to remember that sbm's only have nearest neighbor interactions. This means that once a knot forms, conditions at the knot are totally independent of what is happening at more distant nodes in the system.

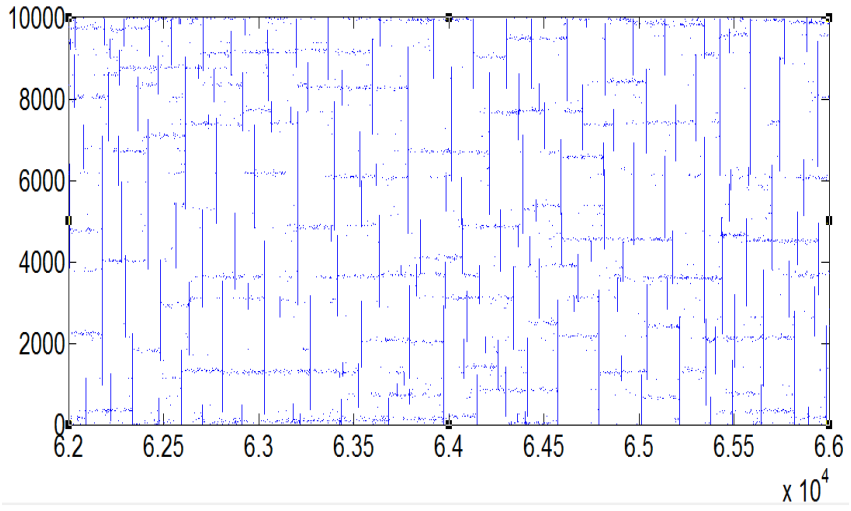


Figure 8.53

Although I claim that this system is chaotic, there is actually quite a bit of structure in the spatio-temporal sequence of the larger events. Complex does not mean that it's random. In nonlinear dynamics, a system is called chaotic if its long-term solution is sensitive to perturbations in initial conditions. If a small

change in the prestress at some time causes the future events to diverge from the unperturbed system, then the system is chaotic, which is the case for this system.

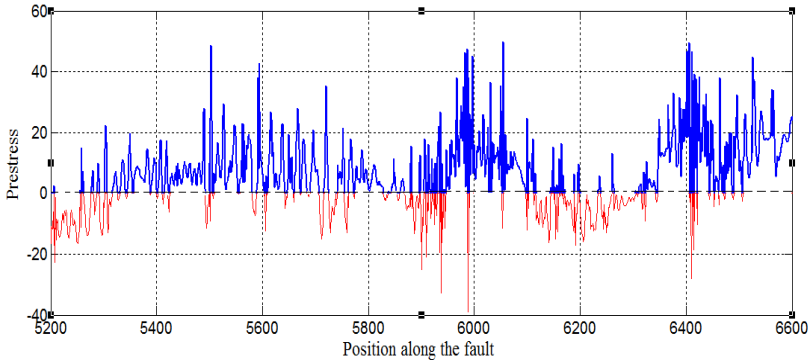


Figure 8.54

Fig. 8.54 shows a snapshot of the shear stress that is associated with Figure 8.53 (case A). Only 1,400 of the 10,000 total blocks are shown to better display the spatial complexity. Areas where the prestress is negative (opposite to the driving stress) are shown in red. Even though the next event on this system is entirely controlled by this prestress distribution, it is virtually impossible to guess what the next event will look like. You've got to run the computer to calculate the complex evolution of the next slip pulse as it propagates through the system. Later, I will show an equation that does in fact do this calculation. An important feature of the system is that two events with virtually identical prestresses can be very different if the events nucleate at different places.

Figure 8.55 shows the relationship between average slip in an event and the rupture length of the event for Case A. Although there is large scatter, there is a clear trend that shows that events with larger rupture lengths have larger average slips. In fact, the ratio of the average slip divided by the rupture length is approximately independent of the overall size of the event (as measured by the rupture length). Figure 8.48b shows the same events, but in this case, the vertical axis is the average change in stress (stress drop) and the horizontal axis is rupture length. This implies that the average stress drop in this system is independent of the event size. This is very similar to the observations of natural earthquakes. In the case of chaotic slip pulses, the **scale invariance of stress drop is a consequence of self-organization of the prestress**; it's not caused by spatial heterogeneity of the friction law.

Although the stress drop is scale invariant in this model, there is still a scatter of about a factor of 100 between individual events. Similar scatter is seen in natural earthquakes. I have the impression that many researchers believe that the scatter in observed stress drops is primarily caused by erroneous measurements. Personally, I believe that most of this scatter is real and that it comes from the

natural variation arising from rupture complexity of slip pulses. In fact, the average stress drop of earthquakes is almost certainly a measure of the heterogeneity of the spatial distribution of slip. That is, if the amplitude of a slip pulse changes slowly as it propagates, then the probability that the pulse will drop to zero is small compared to a similar sized pulse whose amplitude changes rapidly as it propagates. This topic is discussed in a paper by Liu-Zeng, Heaton, and DiCaprio, 2005, The effect of slip variability on earthquake slip-length scaling, Geophys. J. Intl., 162 (3), 841-849. http://resolver.caltech.edu/CaltechAUTHORS:20130305-102043001. It seems that the more heterogeneous the slip, the higher the average stress drop.

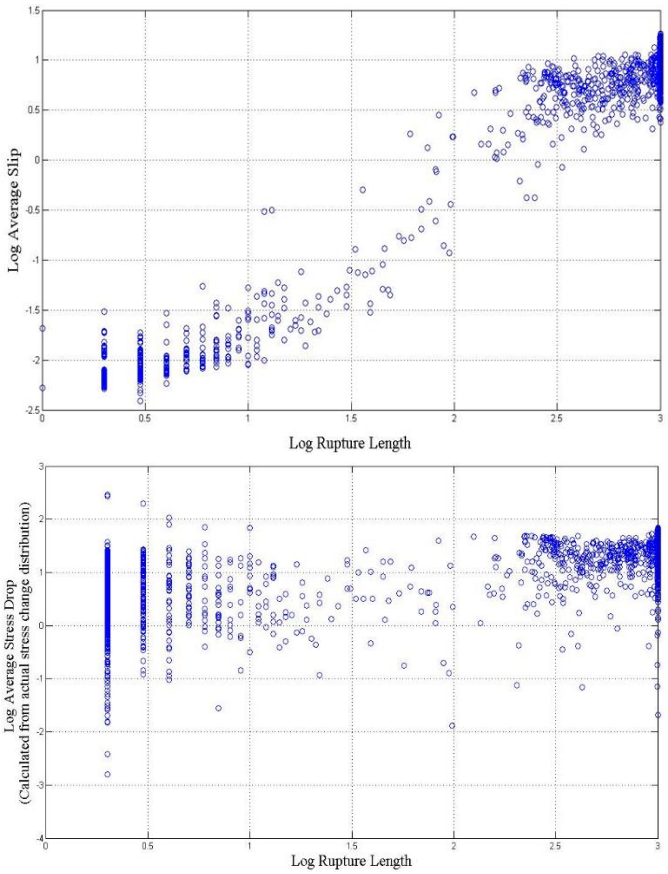


Figure 8.55. CaseA

There are several parameters to play with in the sbm models. The first and most important is the ratio between the coil spring stiffness and the leaf spring stiffness. If the coil springs are very stiff, then the blocks tend to move in unison and then there are more large events and fewer small events (Case A, Fig. 8.53). That is, stiffer coil springs produces events with less spatial heterogeneity and it also results in suites of events that have lower b-values (Gutenberg-Richter).

Figure 8.56 shows the events produced by a system with more compliant coil springs (Case B). Notice that there are more moderate sized events than Case A with stiffer coil springs. That is, this system has a lower b-value. Although the average stress drops for this system are higher than for case A, they are still scale invariant. The scale invariance seems to come from the fractal nature of the prestress. Figure 8.57 shows that the prestress for case B is more heterogeneous than for case A.

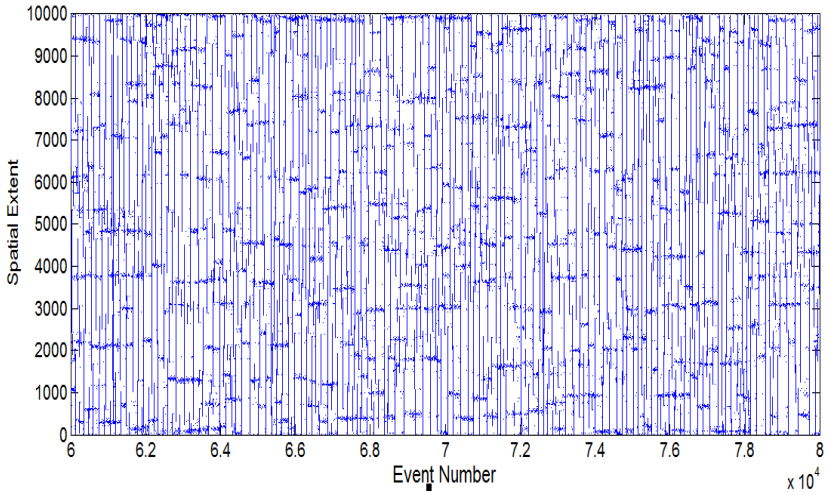


Figure 8.56. CaseB

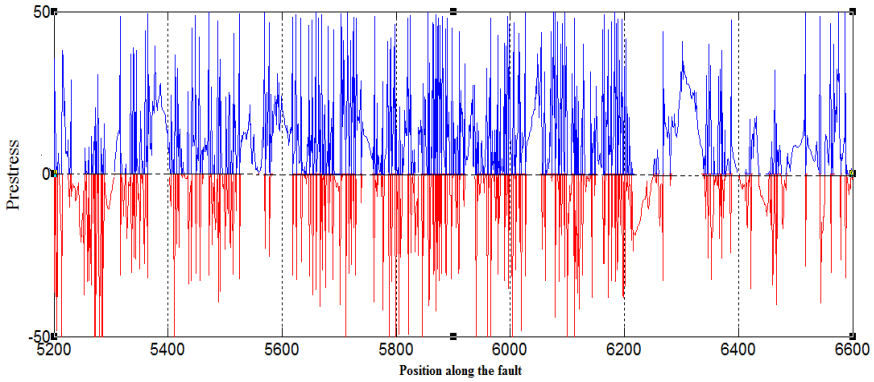


Fig. 8.57. Case B

How to interpret the chaotic sbm

The sbm system has the advantage that slip-pulse dynamics can be studied with modern computers, whereas 3-d elastic systems cannot, even if the events are confined to a single planar fault. Nevertheless, it's important to acknowledge that there are important differences between the Earth and an sbm. For example, we

can tune a sbm to produce power-law frequency-magnitude statistics (similar to the Gutenberg-Richter law). But the sbm confines all of the events to a single line (it's a 1-d model). In contrast, it seems clear that real earthquakes are distributed within a volume. As modern seismology is providing ever greater fidelity to resolve spatial patterns in seismicity, we are seeing ever more complex structures. That is, seismicity seems to not be amorphous clouds of events. Instead, seismicity seems to occur on complex assemblages of planar faults. The implication is that earthquakes occur on fractal networks, while at the same time, the dynamics of the failure is fundamentally chaotic and described by fractals.

At this point, it seems hopeless to attempt to model all this geometric and dynamic fractal complexity. Perhaps a more constructive question is to ask what we would do with our models if we were capable of realistically simulating earthquake phenomena. My opinion is that these new dynamic models provide us with a deeper understanding of failure processes, and they hint at new frameworks for characterizing mechanical properties.

The Earthquake Cycle

Several decades ago, Keitti Aki proposed that all earthquakes shared simple common characteristics (i.e., stress drop, rupture velocity, geometric aspect ratios). Even though I have argued that his insight was fundamentally flawed, Aki's conjecture has been a useful framework to understand the characteristics of earthquakes. Actually, Aki recognized that spatio-temporal complexity was missing from his conjecture, and he later proposed models such as the **barrier model** (patches of high mechanical strength) to introduce complexity. Kanamori and Anderson's seminal paper on earthquake scaling was an extension of Aki's conjecture. Kanamori later introduced the **asperity model** (patches of high prestress) to add spatio-temporal complexity into the model. Although barriers and asperities are fundamentally different, they serve the same purpose of adding complexity.

Whether it's barriers or asperities, both models lead to crack-like dynamics; neither of these models are chaotic. Their behavior may be complex, but it is fundamentally predictable; it's ultimately controlled by spatial variations in earth materials that define friction. Based on the behavior of these crack-like models, researchers have proposed that earthquakes are fundamentally dominated by repeating large earthquakes. A cartoon of this view of earthquakes as large events that repeat regularly (controlled by the average slip per event and the fault loading rate) is often referred to as the **seismic cycle** (see Fig. 8.58).

Geologic evidence of slip in past earthquakes has suggested that complex slip patterns repeat from one slip event to the next one. This has motivated the notion that the large earthquakes are **characteristic earthquakes**. I often observe colleagues who opine about where we are in the seismic cycle. For example, "the big one is overdue." While this type of earthquake cycle logic is

natural if earthquakes are crack-like ruptures, it seems nonsensical from the viewpoint of chaotic slip pulses. Slip pulses arise from extreme spatio-temporal variations in dynamic friction. Dramatically low dynamic friction only occurs in the core of a propagating slip pulse.

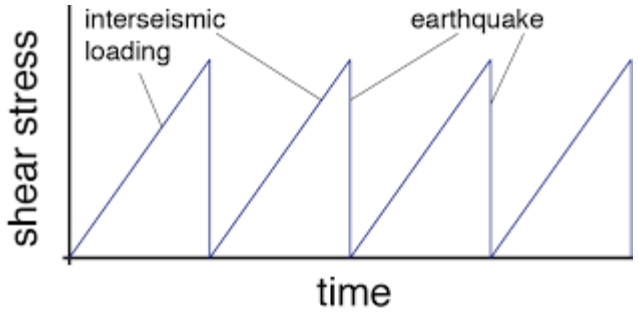


Figure 8.58. Cartoon of the simple Earthquake Cycle

One way to get a feel for importance of the crustal strength is to estimate the total strain energy in the Crust of southern California and to then estimate how many earthquakes would be required to remove all of this energy. This is kind of a whimsical calculation, but it can help to build intuition.

I begin by assuming that the shear stress in rocks is close to the limit of Byerly friction (I have heard several colleagues make this claim). That is, I will assume a coefficient of friction of 0.6 and that between the free surface and a depth of 15 km, the shear stress is approximately increasing linearly at a rate of 20 MPa/km. Now the strain energy density is given by $\frac{1}{2}\sigma\varepsilon = \frac{1}{2\mu}\sigma^2$. Therefore,

$$
\begin{aligned}
E &= \iiint \frac{1}{2}\sigma\varepsilon \mathrm{dV} \\
&= \frac{1}{2\mu}\left(\text{Area of s.CA.}\right)\int_{0km}^{15km}\sigma^2 \mathrm{dz} \\
&= \frac{1}{2\mu}\left(\text{area of s.Cal.}\right)\int_{0km}^{15km}\left(\frac{z\cdot 20\mathrm{MPa}}{\mathrm{km}}\right)^2 \mathrm{dz} \\
&= \frac{4\cdot 10^2\,\mathrm{MPa}^2}{2\cdot 4\cdot 10^4\,\mathrm{MPa}}\left(\text{area of s.Cal.}\right)A\int_{0km}^{15km}\left(\frac{z}{\mathrm{km}}\right)^2 \mathrm{dz} \\
&= \frac{\mathrm{MPa}}{2\cdot 10^2}\left(\text{area of s.Cal.}\right)\int_{0km}^{15km}\left(\frac{z}{\mathrm{km}}\right)^2 \mathrm{dz} \\
&= \frac{\mathrm{MPa}}{2\cdot 10^2}\left(\text{area of s.Cal.}\right)\frac{1}{3}\frac{\left(15km\right)^3}{km^2} \\
&= 10\mathrm{MPa}\cdot km\cdot\left(\text{area of s.Cal.}\right) \\
&= 562\cdot MJ\cdot\left(\text{area of s.Cal.}\right)
\end{aligned}
\tag{8.119}
$$

Now let's take the approximate area of southern California to be 300 km by 300 km or $10^{11}m^2$. Then

$$E \approx 5\times 10^{20}\,J \tag{8.120}$$

Now the conversion to energy magnitude is

$$M_W = \frac{\log E - 4.8}{1.5} \tag{8.121}$$

So there is enough energy for a M= 10.6. Or alternatively, if the crust is uniformly at the Byerly limit, then there is enough strain energy for 7,500 M 8 earthquakes. In this whimsical world, the next earthquake nucleation could set off a chain reaction of 7,500 M 8 earthquakes. Clearly the assumption that the stress is near the Byerly limit is incorrect.

Strength of the chaotic spring-block slider model

While figures similar to 8.58 are common in earthquake studies, they typically avoid putting numbers on the vertical axis to describe shear stress. Nevertheless, there is an implied maximum shear stress that many researchers describe as the strength of the system. As I have discussed earlier, there are long-standing debates about what is the appropriate strength of the crust. Before, I discuss the strength of the crust, I want to discuss the strength of the sbm just presented. This system provides some important insight into how to view the strength of the crust.

Strength based on average stress

If the prestress in the sbm was uniform, then defining the strength of the system would simply be the stress at which failures occurred. However, the prestress is very complex. Given this complexity, there are several different ways that we could think of to measure the strength of the system.

Let's begin by looking at the prestress in the event shown in Fig. 8.52. We could define the strength to be the maximum shear force that can occur on a block, which in this case, is 50. However, the choice of maximum force of 50 has little effect on the overall behavior of this system. A more meaningful choice for strength would be to measure the total shear force on all the blocks that moved in the event that occurred as a result of the force distribution in Fig. 8.52. In a laboratory experiment to measure strength, we typically measure a load that results in failure, and then we divide by the cross-sectional area of the sample to obtain the strength of the material (described as a stress, or force per area). I will define this to be the average stress-based strength, or

$$\Gamma_{\bar{\sigma}} \equiv F/L = \int_o^L \sigma(x)\,dx \Big/ L = \bar{\sigma}_0 \qquad 8.122$$

In the case of the event for fig. 8.52, the average stress is less than 10, which is much less than the maximum shear stress of 50.

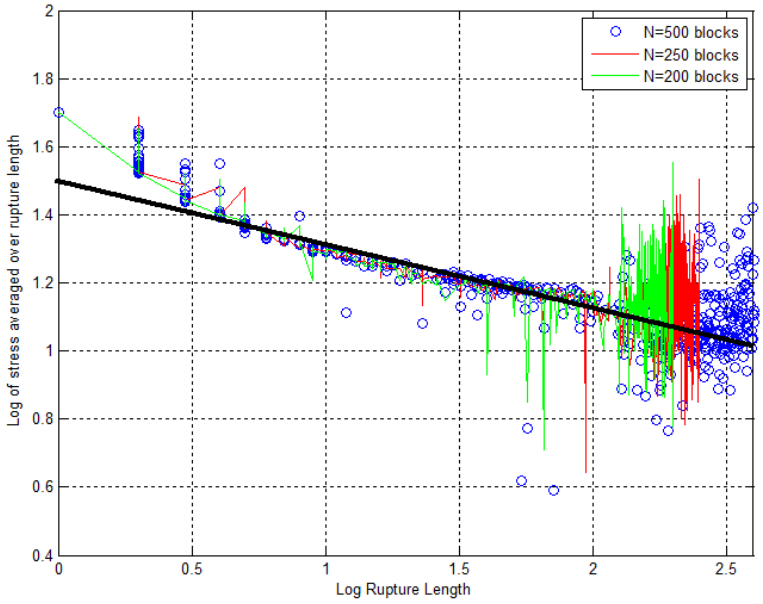


Fig. 8.59. Case A $\Gamma_{\bar{\sigma}} \approx 32 L_R^{-0.2}$

We can immediately see that if the Earth experiences slip-pulse failure, then the strength of the Crust is significantly lower than the maximum shear stress in the system (presumably the shear stress at the hypocenter). That is, laboratory experiments on the frictional strength of materials may not tell us much about the strength of the crust. My opinion is that the "Christmas tree" model (see Fig. 8.26) does not represent the strength of the crust. Instead, it provides an estimate of the maximum localized stress.

In the rate-weakening sbm, faster sliding is associated with lower friction. Faster sliding is more common in longer ruptures that have larger slip pulses. These physics are behind the correlation between strength of the sbm and the length of the ruptures. That is, the average stress decreases with the rupture length as is shown in Fig. 8.59. The heavy line is a log-log fit that indicates that average stress based strength scales with rupture length as, $\Gamma_{\bar{\sigma}} \sim L_R^{-0.2}$. Obviously, larger dimension solids (e.g., the Crust) support larger events; the larger the event, the lower the strength.

This may, at first, seem counter-intuitive. Perhaps it's easier to see if you realize that the definition of strength that I am using involves finding the average of the self-organized prestress at different lengths. It seems that the longer the averaging length of prestress, the lower the average. I will get back to that shortly when I discuss power laws.

Strength based on work

In the laboratory, it is straightforward to calculate the average stress in a sample. However, this is not a feasible measurement in the Earth since we don't know the total load on the failure surface. In principle, we could measure the stress tensor at enough locations that we could obtain a realistic average. In fact, that is what would be required if we want to take a spatial integral to estimate total load. If stress is heterogeneous, then it is not feasible to measure it at enough points to spatially integrate it.

Perhaps you have already observed that I have previously discussed the strength of the crust using arguments about heat energy, fracture energy, and radiated energy (see eq. 8.89). In the sbm, there is no radiated energy or fracture energy, frictional heating is the only dissipation to absorb the potential energy released by the springs. Accordingly, I define **strength based on work** as

$$\Gamma_W \equiv \Delta E_{dissipation} \Big/ L\bar{D} \qquad 8.123$$

You can think of Γ_W as being analogous to a plastic yield stress. It provides an estimate of how much work is required to change the material from an initial configuration to a final configuration. Since energy is an integrated quantity, it is not necessary to worry about the spatial irregularities associated with fractal stress.

Figure 8.60 shows the length scale dependence of Γ_W, and it should be compared with Fig. 8.59. There is a considerably stronger weakening of this strength with increasing length scale. In this example $\Gamma_W \approx 63 L_R^{-0.46}$ (it's case A, see Figures 8.54 and 8.55). The fact that longer ruptures are associated with higher slip velocities is the reason for this stronger length dependence of the energy based strength.

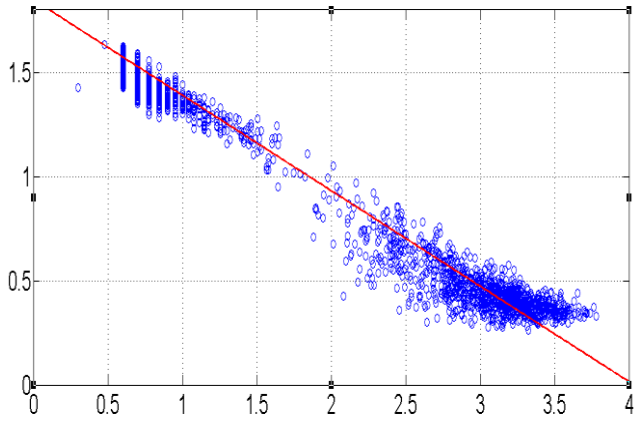


Figure 8.60. log-log Plot of energy-based strength, Γ_W, vs rupture length for events from case A of the sbm. Notice that there is a stronger dependence of energy-based strength on rupture length than there is for stress-based strength (see Fig. 8.59).

This is an important lesson from the sbm. That is, **when considering materials that experience slip pulses, determining the strength of the system using analyses based on energy considerations is likely to provide smaller estimates than is obtained by estimating the total load on the region that fails**.

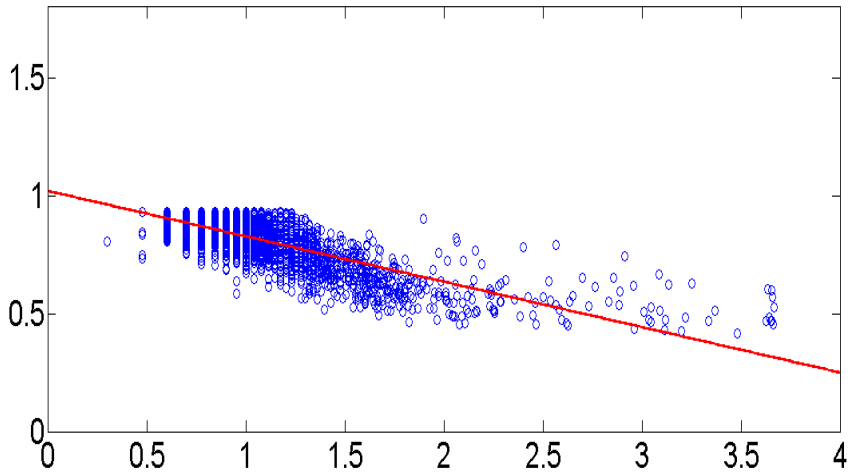


Figure 8.61. Case B

$\Gamma_W \approx 10 L_R^{-0.19}$

Power-Law Scaling

While crack-like rupture dynamics can be quite complex, the far-field radiation from these models can be described by some scaling relations about amplitude and duration. Fundamental theorems concerning the principle of "equivalent width" can be used to infer the shape of Fourier amplitude spectra for crack-like models. Crack-like models with constant stress drop typically produce amplitude spectra that can be characterized with simple power laws. In particular, these power laws have simple exponents (e.g., $f^{-2}, f^{-3/2}$). These laws have exponents that are rational numbers because of some simple geometries (e.g., rupture area scales with L^2, stress drop scales with D/L, derivatives and integrals are equivalent to multiplying or dividing f or f^{-1}, etc.).

However, in chaotic self-organized systems, there are also typically power laws, but in this case, the exponents can be any real number (Although this is a speculation, I suspect that these exponents are typically an irrational number). The power law behavior is the result of self-organization of the system into a self-similar fractal. By definition, fractals are scale free.

I believe that self-organization of chaotic dynamics is the key to understanding Aki's observation that earthquakes seem similar over a wide band of length scales. Ironically, it became popular to erroneously use the term self-similar for Aki's conjecture (his model is similar, but not self similar). But in the end, the more appropriate model of earthquakes is a system with chaotic dynamics that self organizes into a self-similar state (scale free).

It may seem that we are at an impasse. Models like the sbm provide the insight that we are dealing with a chaotic self-organizing system, but computational limitations prevent us from meaningful simulations of 3-d solids that yield during strong rate-weakening friction.

One approach to this impasse is to develop new equations that provide us with meaningful calculations, while at the same time requiring fewer calculations. The need to allow the system to self-organize while simultaneously honoring the complexity that comes from positive feedback dynamics is why this is so difficult. It's important to recognize that if the solution is described by a fractal, then there is little hope of using conventional continuum mechanics (Navier's equation). That is, Navier's equation is a 2nd order partial differential equation, whereas our solutions likely look like fractals. A fundamental difficulty with fractals is that the conventional definition of a derivative doesn't make sense for fractals. That is, derivatives are defined to be the change in a function's value as the step size shrinks to zero. Fractals are irregular at all scales, even when the step size shrinks to zero. This means that **this class of problem can only be investigated with discrete models**. In effect, we anticipate that we will be unable to derive exact solutions to problems.

It is important to recognize that the Fourier spectra of parameters describing individual events that occur as a result of self-organized chaotic dynamics do not necessarily follow a strict power law. Instead, it is **the expected value of** these parameters that **obeys a power law**.

Although I have argued that the amplitude spectrum has a power-law expected value, I have avoided discussion of the phase spectrum. This is a tricky issue, since chaotic behavior is definitely not the same as random phenomena. That is, there is structure (often hidden) that results from these systems. This structure is fundamental to the identification of **strange attractors.** In the case of our problem, we can anticipate that the phenomena are **statistically stationary**. That is, the expected values of our parameters are independent of the initial conditions of the system.

Slip v Length (stress drop) scaling

One example of a stochastic power-law model comes from a study of slip vs length scaling by Liu-Zheng, Heaton, and DiCaprio. We considered a simple 1-d model of slip as a function of linear distance x. In particular, we assumed that slip is a stochastic function of *x*. We hypothesized that the amplitude of a slip pulse as it propagates along the fault is described as a **random walk**. Consider that I have discretized distance along the rupture as x_i. I will assume that

$$D(x_{i+1}) = D(x_i) + R_i \qquad 8.124$$

Where R_i is a number chosen from a bin of random numbers. This is known as a 1-d random walk and this process has been used to simulate the **Brownian motion** of particles that are randomly impacted with fluid molecules. 1-d Brownian motion is the same as taking the integral of random white noise over all past steps. The random sequence $R(x)$ has a Fourier transform that is also composed of random numbers. That is, the spectrum looks like random numbers that do not change their expected values as a function of wavenumber. Therefore, the amplitude spectrum is flat and this is called **white noise**. In contrast, the stochastic slip is the spatial integral of this random function and thus it has a $1/k$ spectrum.

I can generalize this model by assuming that the slip is a fractal function of space, or

$$\tilde{D}(k) = \tilde{D}_0 \frac{R(k)}{k^{\alpha}} \qquad 8.125$$

In standard Brownian motion, $\alpha = 1.$ However, since we are dealing with self-organizing fractals, we anticipate that α may be an irrational number; in this case the process is called a fractional random walk. Unfortunately, current measurements of slip variations are inadequate to obtain a stable estimate for the amplitude spectrum of slip of real earthquakes.

Liu-Zheng, Heaton, and DiCaprio (2005, GJI, doi:10.1111/j.1365-246X.2005.02679.x) used this fractional random walk model to reproduce measured values of the ratio of average slip divided by the rupture length. In particular, measurement of Potency and rupture length are available from many earthquakes. As I will now show, the ratio of Potency to rupture length depends on the value of α .

We considered this to be an example of a class of problems known as **gambler's ruin**. The focus of this problem is to estimate how many games a gambler can play before she loses all of her money. That is, slip as a function of space is transformed into a sequence of games of chance. Once the slip crosses the x-axis, the earthquake is considered to be over. That is, the rule of this game is that earthquakes are comprised of spatially continuous regions of slip. Figure 8.62 shows examples of spatially contiguous slip profiles for different values of α . All of the walks start at zero slip and end at zero slip; most of the examples involve only several steps until the slip returns to zero. The three examples in Fig. 8-62

are chosen to show the spatial characteristics of the larger events in this numerical experiment. The spatially averaged slip for each event is shown as D1, D2, and D3. Notice that the red curve $(\alpha = 1.0)$ is quite rough. This is true Brownian motion. The length of this red rupture is 72 km and the average slip is 4.5 m, which gives a slip to length ratio of 6×10^{-5}, which is comparable to the average value of earthquakes (see 8.33).

The next thing to notice in Figure 8-62 is although the different events have comparable rupture lengths, the average slip decreases as the roughness is filtered out. This trait is clearly displayed in Figure 8-63, which shows how increasing smoothness, α, results in systematically longer rupture for a given average slip. That is, the rougher the fractional random walk, the better the chance that the slip will cross the x-axis (termination of the event).

You should recognize that the ratio of average slip to rupture length is a measure of the average stress drop for these events. Thus, this model predicts that **rougher slip distributions produce higher average stress drops**. Also notice that, although the red curve $(\alpha = 1.0)$ has a rupture length of 73 km, it very nearly terminated at the 67-km mark, which would have been a 44-km rupture. That is, it was mere chance that the red event did not have a significantly higher stress drop. In this model, two events could have identical slip distributions up until one of them stopped rupturing. If the second event continued to rupture it, might end up with a much lower stress drop, even though the radiation from the first part of the ruptures was identical. This **example shows that average stress drop may not tell us much about the dynamics** of a rupture.

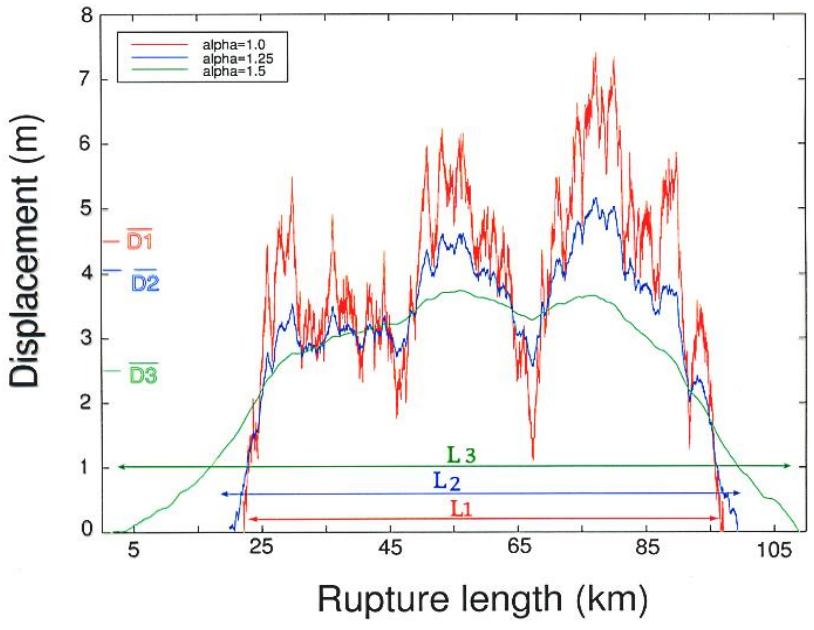


Figure 8-62. Stochastic simulations of slip as a function of distance on a 1-d fault. Random numbers are filtered with a low-pass power law filter, after which, stretches of positive only are selected to simulate heterogeneous slip on ruptures. Alpha=1 is a full integral of random white noise, alpha = 1.5 is the 3/2 integral of random noise. All chosen events have the same magnitude (i.e., potency). Notice that the ratio of the average slip to the rupture length (aka, stress drop) increases as the slip is more heterogeneous (smaller alpha).

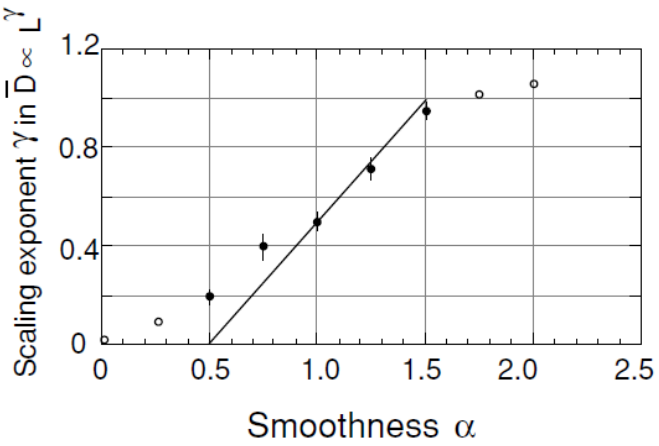


Figure 8-63. Similar to Fig. 8.62, but the vertical axis is the dependence of the slip to length scaling on heterogeneity. Notice that scale-free stress drop (D ~ L) implies that alpha=1.5.

Now the observation that average stress drop is approximately independent of scale is direct evidence that average slip scales linearly with rupture dimension. Notice that in standard 1-d random walk (aka, Brownian motion $\alpha = 1.0$), the expected rupture length scales as the square root of the average slip. However, when the smoothness increases to $\alpha = 1.5$ (one and one half integrals of random white noise), then the expected rupture length scales linearly with average slip. Curiously, $\alpha = 1.5$ seems to be some type of critical point. That is, if $\alpha > 1.5$, then the random walks tend to diverge with distance (the gambler always wins).

While a 1-d fractional random walk model can adequately explain stress drop scale invariance, it does a poor job of simulating the Gutenberg-Richter relationship. Figure 8-64 shows the slip/length ratios of many simulated events And for several values of σ. It appears that a smooth distribution , $\alpha = 1.5$, produces just as many long ruptures as it does short ones. That is the b-value is approaching 0. In contrast, a standard random walk ($\alpha = 1.0$) produces many more short ruptures than long ones.

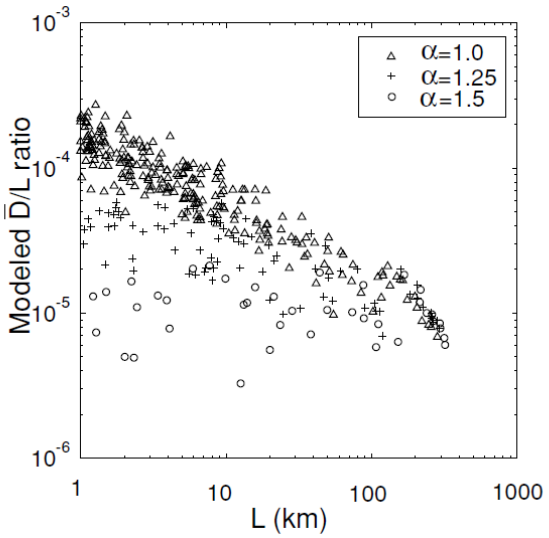


Figure 8-64. This figure follows from Fig. 8-63. Stress drops (aka, D/L) of individual stochastic simulations are shown as a function of rupture length (remember, it's a gambler's ruin problem) for three different roughnesses of slip.

This observation raises some interesting questions. Are large strike-slip ruptures with very long ruptures (340 km for 1857 Ft Tejon, or 450 km for 1906 San Francisco) compatible with the Gutenberg-Richter law? If they are to be explained by a fractional random walk, then the slip must vary slowly along the rupture. I admit that I don't know how to explain these very long ruptures. However, if you examine the slip to length ratio of major historic crustal earthquakes (see Wells and Coppersmith, BSSA, 84, 974-1002), you will see the distribution shown in Fig. 8-65a. In comparison, Fig. 8-65b shows the result of a combination of numerical simulations of fractional random walks with $\alpha = 1.25$ and $\alpha = 1.5$. This particular combination matches the observation that the variation in the slip to length ratio is larger for shorter ruptures than it is for longer ones (notice the trend in the upper limit in both plots).

The observed similarity between measure slip/length ratios and fractional 1-d random walks is evidence that the observed scale invariance of stress drop may be closely related to the heterogeneity of ruptures, and not a simple characteristic of friction. The fact that, for a given magnitude, stress drops can vary by more than two orders of magnitude has often been attributed to modeling error. Personally, I prefer to interpret this large range of reported stress drops as being a real phenomenon that just reflects variations in random walks.

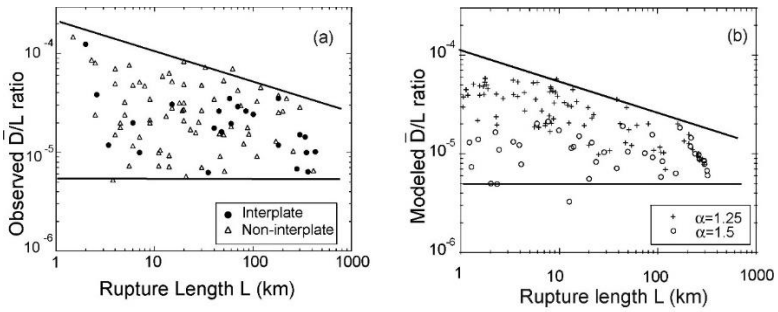


Figure 8-65. Comparison of observed slip v length ratios v stochastic simulations of slip v length

It is important to recognize that the models of spatial complexity that I have just described all come from 1-d models; there are no complicated fault systems similar to what is observed in the real earth. In fact, there is great variety in the geometric complexity of fault systems. Some faults are remarkably simple planar unconformities. In fact, it is possible to view the Punchbowl fault in great detail by a pleasant trip to Punchbowl Regional Park. The Punchbowl fault is a major right-lateral strike-slip fault that runs sub-parallel to the San Andreas fault just north of the San Gabriel Mtns. This fault is interpreted to be an early version of the San Andreas and it appears that the total offset on this fault is about 45 km. Luckily, erosion over millions of years has exposed extensive sections of the fault for detailed investigation. There is a very good field guide by Fred and Judy Chester (they have many excellent publications describing many aspects of this fault). A visit to the Punchbowl allows you to view the fault along its strike and it's plain to see that it is geometrically flat (despite the large total offset). It's even more amazing to view the fault up close. In most areas, the total thickness of the fault zone is less than a meter and there is evidence that most of the sliding was confined to a zone only several mm wide.

The San Gabriel Fault is another early version of the San Andreas that has been exposed by erosion (again studied by Chester and Chester). If you hike along this fault zone, you will find some places where the fault is a simple thin zone (like the Punchbowl fault), but you will also find other places where the fault becomes very complex (a complex zone of pulverized rock more than several hundred meters wide (for example, visit Red Box Gap just north of Mt. Wilson).

Recent advances in remote imaging have provided a new perspective on the complexity of individual earthquake ruptures. For example, comparison of before and after satellite-based photographs of the desert floor show that the 2019 Ridgecrest rupture occurred on a complex network of fault structures (Figure 8-66). Given the observed geometric complexity of observed faults, it's easy to think that the dynamic complexity of earthquakes is entirely due to this geometric

complexity. This is an important unsolved problem in earthquake physics. That is, we suspect that 3-d complexity of faults is important, but we can't even numerically simulate the dynamic complexity that occurs on simple planar faults that experience pulse-like ruptures.

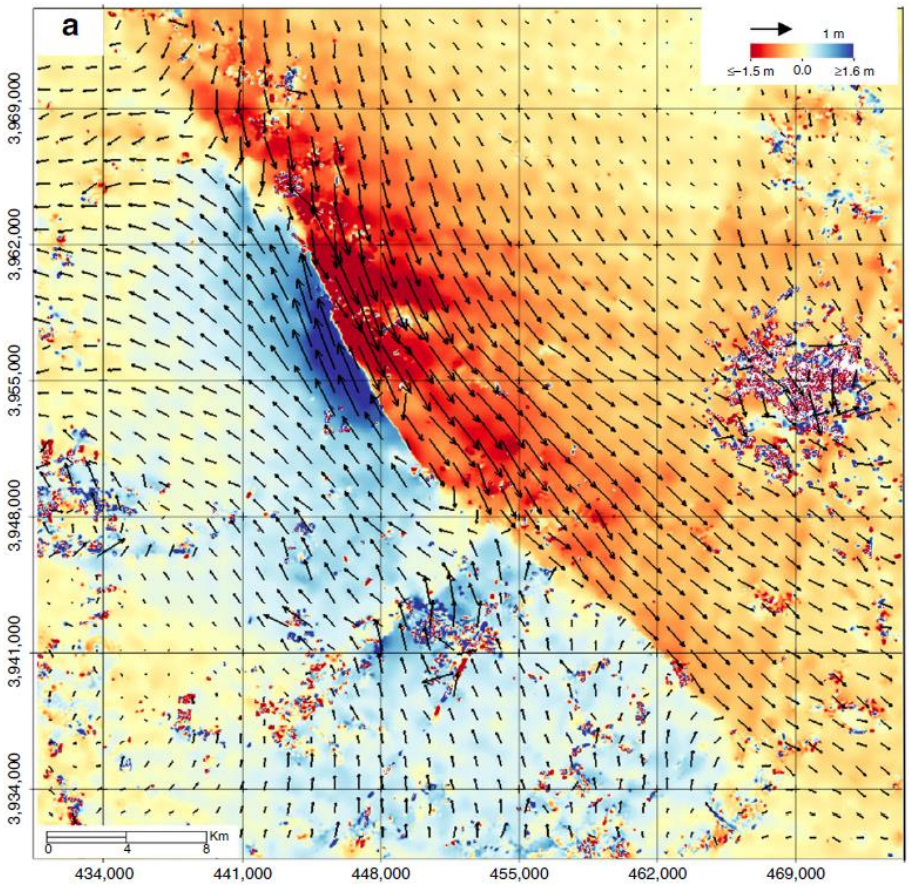


Figure 8-66. Deformation of the ground surface in the vicinity of the 2019 Ridgecrest earthquake sequence. (Chen, Avouac, Aati, Milliner, Sheng, and Shi, https://doi.org/10.1038/s41467-019-13750-w).

The type of complexity shown in Fig. 8-66 is typical of other well-observed crustal ruptures (e.g., 1991 Landers, 2001 Hector, 2011 Cucapah-El Major, 2017 Kumamoto, 2018, and 2016 Kaikoura). In contrast, there are long sections of the San Andreas fault that appear to be relatively simple and planar (especially the Carizzo Plain). The very long ruptures in 1857 and 1906 appear to have been on relatively straight sections of the San Andreas.

Although we observe geometric complexity in fault systems, it is difficult to simulate and interpret. In particular, any kind steadily rupturing fault that has localized jumps in fault orientation (e.g., a kink) will produce singular stress

changes at geometric corners. Some modelers have argued that geometric corners are the likely places to initiate and terminate rupture (often referred to as **fault segmentation**). Although this seems like a plausible assumption, there are clear examples where fault ruptures seem to ignore these segment boundaries.

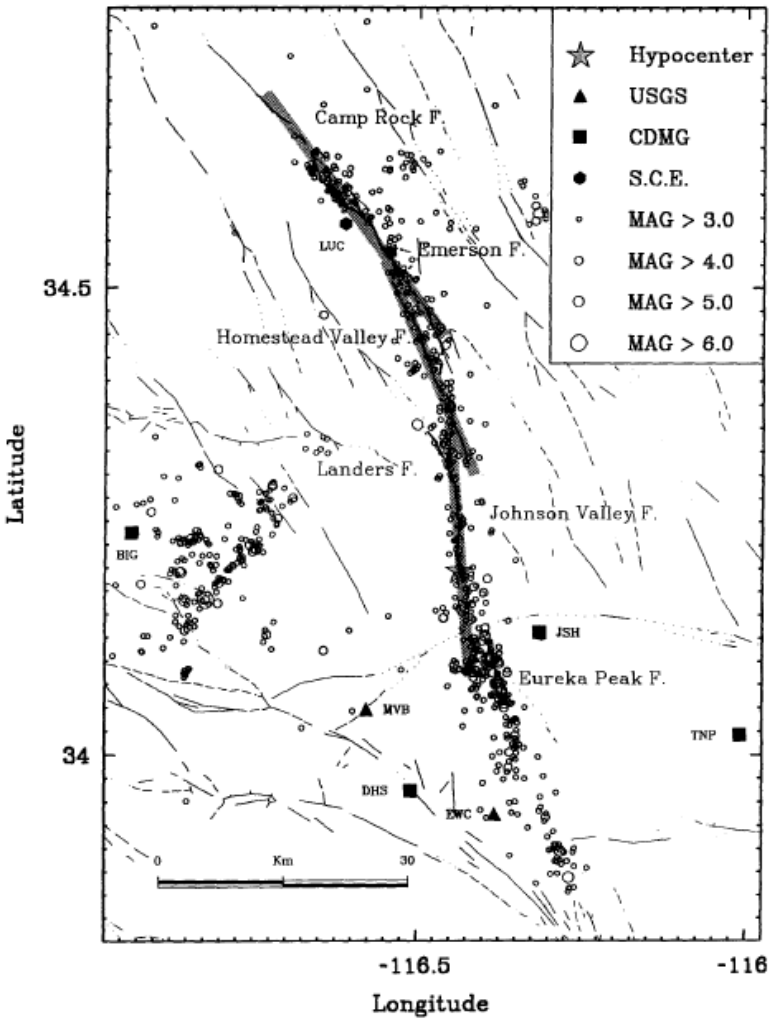


Figure 8-67. Geometric complexity of the 1991 Landers earthquake.

The 1991 Landers rupture is an example that shows a failure of the fault segmentation hypothesis. That is the rupture progressed northward along the Johnson Valley Fault, and then it continued to rupture when it intersected the Homestead Valley Fault. It then ruptured right through another segment boundary between the Homestead Valley Fault and Emerson/Camp Rock fault. The rupture finally terminated in the middle of a fault segment (perhaps the straightest fault segment in the whole complex). While it may be that fault complexity and segmentation has a first-order effect on rupture dynamics, the

Landers earthquake clearly shows that simple models do a poor job of predicting reality.

Heterogeneous Stress in the shallow Crust

There are numerous places in this chapter that I have argued that stress is likely to be spatially complex (look at the discussion of chaotic slip pules). In the following, I discuss the statistical characteristics of this heterogeneous stress. Unfortunately, this discussion may be difficult to follow. Furthermore, the hypothesized model is very different from conventional models of crustal stress (for example, see Zoback and Zoback, or Scholz). The model was developed by Deborah Smith and me and it is designed to show that the effect of heterogeneous stress depends strongly on the length scale that is appropriate for different problems.

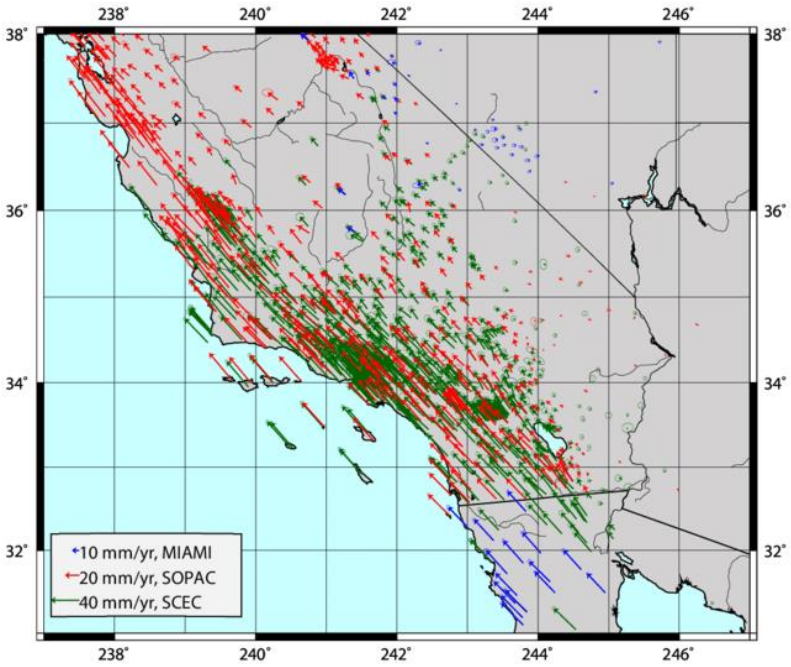


Figure 8-68. Map of Southern California with velocity arrows used in this study in a fixed North America reference frame. Red arrows: velocity provided by the Scripps Orbit and Permanent Array Center (SOPAC); blue arrows: data from the University of Miami; green arrows: data from the crustal motion map (version 3) by the Southern California Earthquake Center (SCEC). Hackl, M. & Malservisi, Rocco & Wdowinski, Shimon. 2009, Strain patterns from dense GPS networks, Natural Hazards and Earth System Sciences, 9. 10.5194/nhess-9-1177-2009.

The measurement of stress in the Crust is actually very difficult. Testing in deep boreholes is probably the most direct approach to this problem. Unfortunately, deep boreholes are extraordinarily expensive and they are typically limited to the upper 5 km of the Crust. Since direct measurements are so sparse, many earth scientists use simple models to infer the stress in the crust. One of the most successful methodologies to inform these models comes from observations of the spatio-temporal distribution of the changes in positions of geodetic monuments over time. The introduction of satellite-based surveying (Global Positioning Satellite system, GPS) in the 1990's has provided the data to track the steady deformation that drives plate tectonics. Figure 8-68 shows the velocities of stations in southern California relative to the average velocity of North America. These velocity vectors are very coherent; they all point to the northwest and their amplitudes steadily increase as for stations located further from stable North America.

The impressively coherent, long-term, point velocities in Fig. 8-68 might mislead you into the impression that the crustal stress in comparably coherent. In contrast, Fig. 8-69 shows the orientation of the principal compression axes measured in a number of boreholes in southern California. These orientations are primarily obtained by observing the orientation of **borehole breakouts**, which are spontaneous fractures in the walls of boreholes (primarily from the oil industry). Borehole breakouts typically occur in the faces of the borehole that are perpendicular to the least compressive principle stress axis (aka, the Sh axis). SH is the direction of the largest compressive principal stress axis. Because the Earth's surface is traction free, Sh and SH are generally horizontal and therefore, they are perpendicular.

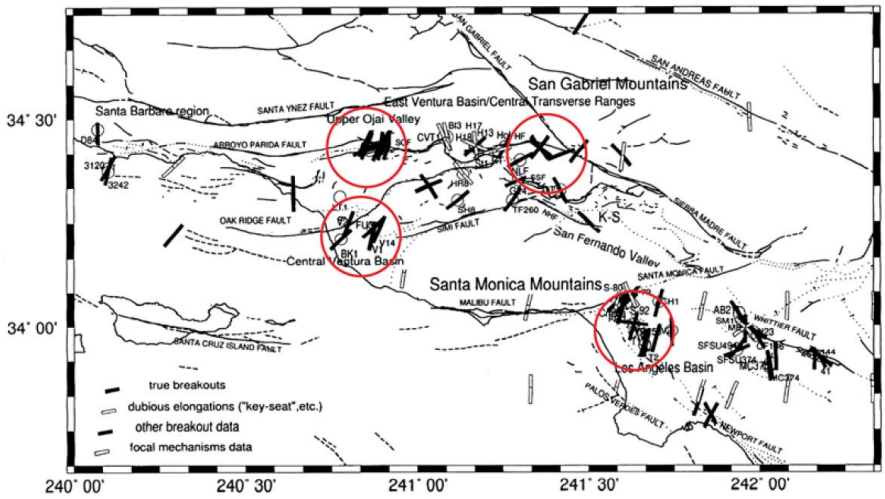


Figure 8-69. Wilde and Stock [1997, *J.Geophys. Res., 102,* 4969-4983] plotted inferred maximum horizontal compressive stress, SH , orientations from borehole breakouts in Southern California. There are a variety of orientations for borehole breakouts from the same borehole or from boreholes spatially close to one another. This suggests short-wavelength spatial stress heterogeneity. In this modified plot, we have used circles to point out a few of the locations studied by Wilde and Stock that show evidence for SH orientation heterogeneity. Modified by Smith and Heaton, 2011, Bull. Seism. Soc. Am., 101, 1396–1421, doi: 10.1785/0120100058

Notice that there is considerable scatter in the SH orientations. Although this scatter is sometimes described as "noise," observation of breakouts is relatively straightforward. That is, direct measurements of stress orientation provide evidence that the actual stress field may be far more complex than you might imagine from inspection of particle velocity field shown in Fig. 8-68. It's not noise.

The statistical analysis of the orientations of focal mechanisms is currently the most popular method to determine the spatial orientation of stress in the crust (see foundational papers by Angelier, J., 1984, J. Geophys. Res., 89, 5835–5848, and by Michael, A., ,1984, J. Geophys. Res., 89, 11517–11526 and also Michael, A., 1987, J. Geophys. Res., 92, 357–368. These studies make the following **key assumptions**: 1) stress is approximately uniform in space, 2) the crust has a large suite of pre-existing fault planes with a large variety of orientations, 3) these faults have a large variety of yield stresses, 4) seismicity in a region occurs on a diverse suit of planes with many orientations 5) (most importantly) the slip vectors of all

of the earthquakes are aligned with the direction of maximum shear stress in the region. According to this technique, the orientation of the principle stresses in a region is that orientation that minimizes the misfit of one of two plausible slip vectors (there are two conjugate planes for every focal mechanism) for the suite of focal mechanisms.

Focal-mechanism-based stress inversions are widely used as the basis to describe the stress distribution in the Earth's Crust. Unfortunately, there is a **fundamental inconsistency in the underlying assumptions**. Specifically, it seems impossible to have spatially uniform stress (assumption 1) if there is an assortment of randomly oriented fault planes with a very wide range of yield strengths (see Rivera and Kanamori, 2002, Geophys. Res. Lett. 29, art. no. 1088).

A particularly egregious violation of the assumption of homogeneous stress comes from a study of the aftershocks of the 1989 M 6.9 Loma Prieto earthquake by Zoback and Beroza (1993, Evidence for near-frictionless faulting in the 1989 M 6.9 Loma Prieta, California, earthquake and its aftershocks, *Geology, 21,* 181-185. Figure 8-70 shows the focal mechanisms of aftershocks of the 1989 Loma Prieta earthquake projected onto the main fault plane. Slip in this earthquake was oblique (a steeply dipping plane with a slip vector that was primarily right-lateral strike slip with some thrusting). Rather mysteriously, a significant percentage of the aftershocks had slip vectors in the opposite direction to the mainshock. That is, there were left-lateral aftershocks on the obviously right-lateral San Andreas fault. Zoback and Beroza explained this by hypothesizing that the static coefficient of friction was nearly zero on the San Andreas. This would have made the maximum principal stress axis nearly perpendicular to the San Andreas. Zoback and Beroza argued that there were small variations in fault strike and that the shear stress resolved to different polarities as the strike varied.

I find this explanation to be implausible for several reasons. 1) the amplitude of the slip (see the shaded contours) varies by more than a meter within several km. This implies large local strain changes, which implies large local stress changes. This observation seems incompatible with an assumption of uniformly low static stress on the mainshock fault plane. 2) Such a small static friction (less than 0.1) has never been observed in testing and it seems incompatible with the shear stresses that are necessary to support the gravitational load of the Santa Cruz mountains (directly above the rupture).

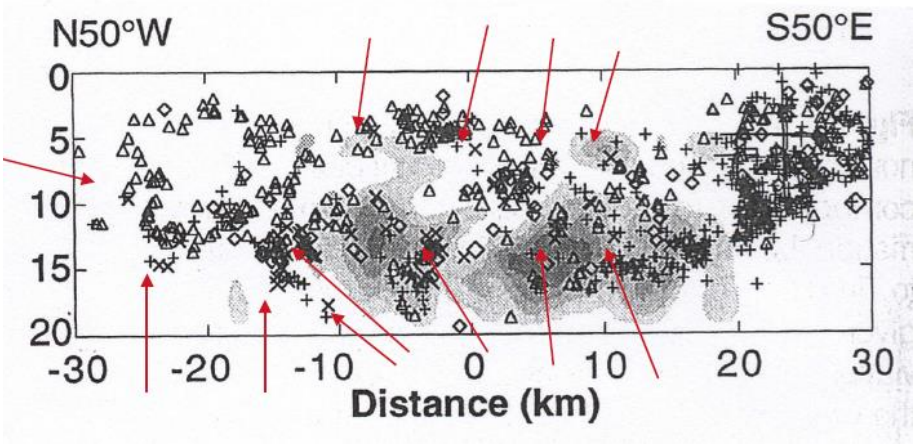


Figure 8-70. Map view of the aftershocks of the 1989 Loma Prieta earthquake (from Beroza and Zoback). The slip distribution derived from a finite-fault slip inversion is also shown as the shaded contours. The vertical axis is depth. (+) represents right-lateral focal mechanisms, (x) represents left-lateral, (Δ) reverse thrusting, and (◊) is normal. Each focal mechanism has two conjugate (perpendicular) planes and the plane that is most parallel with the mainshock is displayed. The Red Arrows (x) highlight aftershocks with slip vectors that are in the opposite direction from the mainshock.

Another explanation is that the stress is highly heterogeneous (similar to the stress in Fig. 8-52) and that there are actually patches of large left-lateral stress that exist on the San Andreas fault. These patches of negative shear stress would be the remnant of previous chaotic ruptures; the dynamic shear stress in the vicinity of a slip pulse are more than an order of magnitude greater than the average stress change. Since steadily increasing tectonic shear stress is in the direction of plate motions, regions of negative shear stress will not nucleate earthquakes. That is, only regions in which the local shear stress is large and also aligned with tectonic stress rates will have future hypocenters. The only time that we expect earthquakes in the opposite direction from plate motion is during an aftershock sequence. That is, the spatially complex slip in the mainshock means that the stress changes from the mainshock are even more complex than slip pattern. This interpretation seems to explain seismicity near the Loma Prieta earthquake; events with negative slips have not been observed in the background seismicity, they are only observed in the aftershocks.

Another curious feature of seismicity (either background or aftershocks) is that it ends to be clustered in space; I sometimes call it clumpy. That is, there are typically spatial knots of persistently high seismicity (just look at any high-resolution seismicity map). Although it may be tempting to assign different material properties to these spatial knots, I suspect that these high-seismicity

knots are ephemeral (perhaps on the time scale of decades to centuries). In particular, I suspect that patches of high seismicity are patches of high stress

Deborah Smith and I created a stochastic model of the stress tensor in an attempt to simulate patchy seismicity that has the statistical characteristics of observed focal mechanism catalogs. We hypothesized that, within some region, the deviatoric stress tensor, $\boldsymbol{\sigma}'(\mathbf{x},t)$, can be approximately separated into tectonic term that is steadily increasing and approximately spatially uniform, $\dot{\boldsymbol{\sigma}}'^{T}(t-t_0)$, and a term that is spatially heterogeneous and independent of time, $\boldsymbol{\sigma}'^{H}(\mathbf{x})$. This model assumes that at the time of the last nearby earthquake, t_0, the heterogeneous stress and the average background stress, $\boldsymbol{\sigma}'^{B}$, were reset to new values by the heterogeneous stress changes caused by chaotic rupture. The deviatoric stress in a region is written as

$$\boldsymbol{\sigma}'(\mathbf{x},t)=\boldsymbol{\sigma}'^{B}+\dot{\boldsymbol{\sigma}}'^{T}(t-t_0)+\boldsymbol{\sigma}'^{H}(\mathbf{x}) \qquad 8.126$$

We next hypothesize that events nucleate in the region whenever the shear stress exceeds some pre-determined value, $I_2'^{Y}$, where I_2' is the 2nd deviatoric stress invariant. This failure criterion is referred to as **Henky-Mises plastic yield criterion** and it is commonly used in the ductile yielding of steel. That is, we assume that the shear stress in the crust cannot exceed this limit; once the yield stress is achieved, the material experiences inelastic shearing along conjugate planes of maximum shear stress.

To construct this model, we begin with a 3-d Cartesian grid of discrete points, $x_{i,j,k}$. At each grid point, we use a Gaussian, mean-zero, random number generator to select the values of the components of the stress tensor. We assume that the stress tensor is symmetric, $\sigma_{ij}=\sigma_{ji}$, which ensures that angular momentum is conserved (see Chapter 3). The standard deviation of the diagonal components is chosen to be 1.0, while the off diagonal terms have a standard deviation of $1/\sqrt{2}$. This choice ensures that the **random stress is isotropic**; this is the only ratio that provides stresses that are uniformly distributed in orientation space.

Once we have assigned random numbers to each of the six independent stress components, we apply a **power-law spatial filter** to each stress component. Because our model is isotropic, we can uniquely specify this filter by defining the spectral properties along any line, $r_i=|\mathbf{x}_i|$, passing through the grid. In order to

construct the heterogeneous stress, we take the 3-d Fourier transform of each tensor component. We then multiply these spectra by

$$F\left(k_r\right)=\left(1+\left|k_r\right|\right)^{-\alpha} \qquad 8.127$$

After which, we take the inverse 2-d transform to obtain the stochastic, power-law, stress tensor in the spatial domain. This process creates a fractal-like distribution. While the expected value of the average components is zero, the mean of a finite number of random numbers is not zero. We remove the mean of our random numbers and then we reintroduce finite mean to our stress through the specification of $\boldsymbol{\sigma}'^{B}$. That is, the stress is the sum of a zero-mean stochastic part with a specified "background" stress.

Figure 8-71 shows the stochastic model of heterogeneous stresses projected onto a plane that intersects the 3-d grid (100x100x100 points).

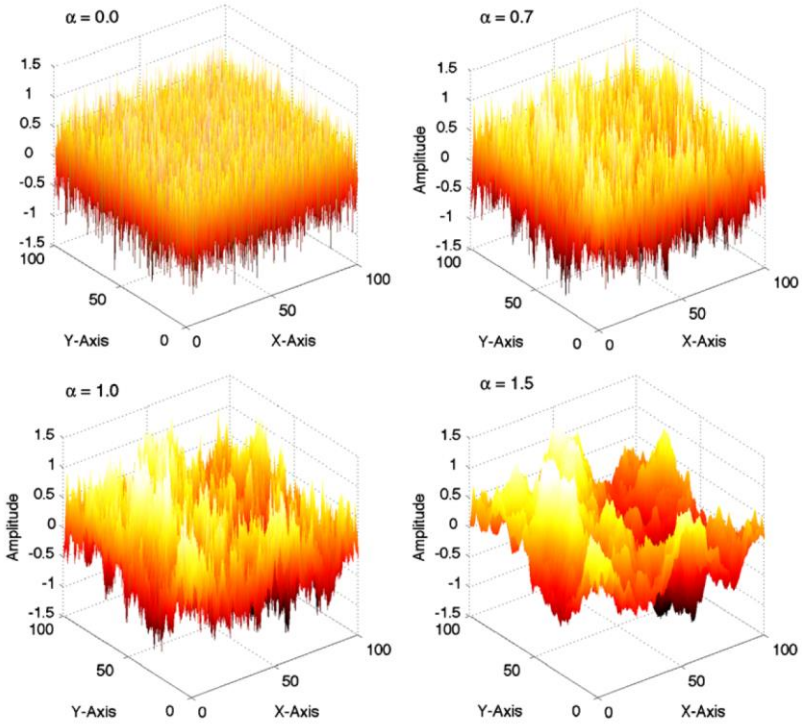


Figure 8-71. Examples of one component of the stochastic power-law heterogeneous stress for different powers, $\left(1/1+k_r\right)^{\alpha}$, of wavenumber.

In order to have a 3-d model with realistic values of stress, it is necessary to find a meaningful scale. In particular, we chose to relate the variance of the heterogeneous stress to the size of the uniform background stress $\boldsymbol{\sigma}'^{B}$. In particular, we define the **Heterogeneity Ratio, HR**, as

$$HR \equiv \sqrt{\frac{\bar{I}_2'^{H}}{I_2'^{B}}} \qquad 8.128$$

Where $\bar{I}_2'^{H}$ is the spatial average of the 2nd invariant of the heterogeneous deviatoric stress. $\bar{I}_2'^{H}$ is also a measure of the shear-strain energy density (see Housner and Vreeland, 1965). As it turns out, $\bar{I}_2'^{H}$ is also the sum of the variances of the components of $\boldsymbol{\sigma}'\left(\mathbf{x}, t = t_0\right)$ (see Smith and Heaton).

For this parameterization of the relative size of the uniform background stress and the stochastic heterogeneous stress, we want to ensure that the modeled stress is independent of the number of grid points used to describe a region. To accomplish this, we use the same outer scale for different grids, independent of the number of grid points. Fig. 8-72 shows examples of the amplitudes of maximum shear stress (heterogeneous plus background) at points along a line intersecting the grid. Random white Gaussian noise is used for the left column and the preferred power-law, $\alpha = 0.8$, is used in the right column. The heterogeneity ratio, HR, increases from top to bottom; the model in the lower right (HR=2.375 and $\alpha = 0.8$) best fits southern California focal mechanism data.

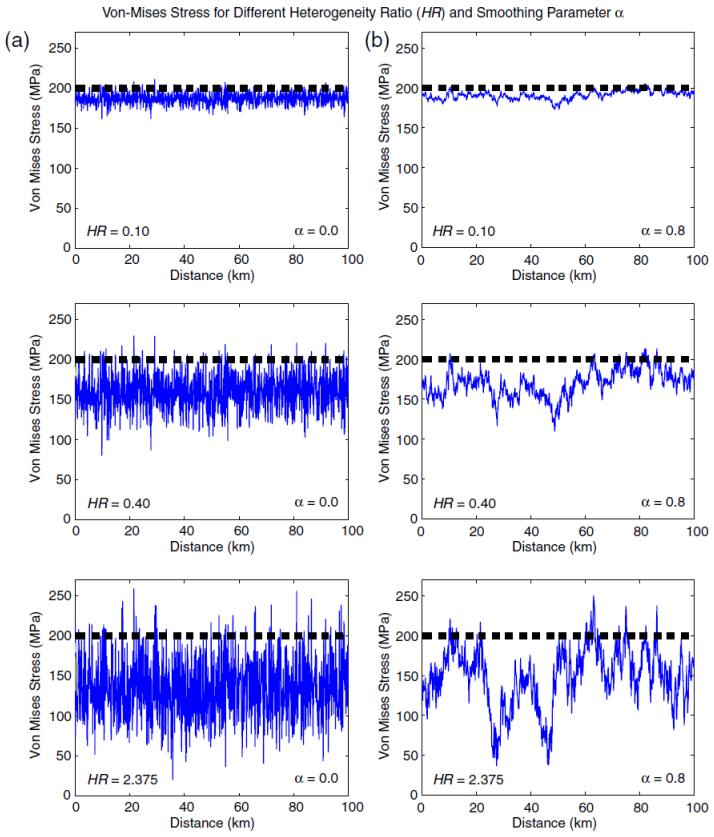


Figure 8.72. Maximum shear stress (aka, Von Mises stress) plotted along lines that intersect the 3-d spatial grid of fractal stress. The horizontal dotted lines are taken to be the local yield stress of the material (200 MPa). The overall heterogeneity increases from top to bottom. The heterogeneous stress is white-noise random in the left column, whereas the smoother preferred model is shown in the right column.

Notice that there is a horizontal dotted line at a shear stress of 200 MPa. This signifies the yield stress at about 8.0 km (assumes Byerly-like friction). Grid points at which the shear stress already exceeds the yield stress are removed from the model. After that, time increases and the spatially uniform tectonic stress, $\dot{\boldsymbol{\sigma}}'^{T}\left(t-t_0\right)$, steadily increases at a rate of 10 kPa/yr. This stress rate is consistent with the strain rate inferred for the high-strain rate regions of Figure 8-68.

The amplitude of stress grows with time at most of the grid points in Fig.8-71. The **rate at which this stress amplitude grows is determined by the relative orientation of the tectonic stress and the initial stress**. If these two stresses are aligned, then the growth rate of the amplitude is large. If, on the other hand the initial stress is opposite to the direction of the tectonic stress rate, then that

point evolves ever further from the yield stress. Only grid points at which the stress amplitude exceeds the yield stress experience hypocenters. Once a grid point is declared to be a hypocenter, its location, origin time, and focal mechanism (orientation of maximum shear stress) are included in a synthetic catalog. Furthermore, once an event is generated, the location of that grid point is removed from the model. There is no concept of magnitude in this simulation; all events are simple grid points and there is no stress transfer to adjacent grid points when an event occurs. This means that there are no aftershocks in this model.

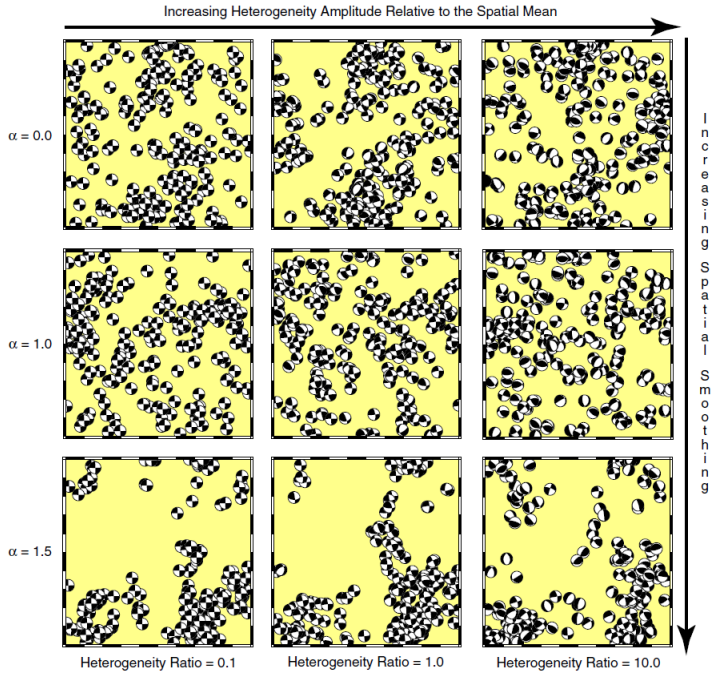


Figure 8-73. Simulated seismicity maps based on the fractal stress model of Smith and Heaton. Models increase in overall heterogeneity from left to right. Models increase in spatial roughness from bottom to top. Notice that spatial clustering increases as the models become smoother; the top row is random white noise plus a spatially uniform stress.

The most important concept in this model is that locations of hypocenters are **spatially biased; the seismicity is highest in places where the stochastic heterogeneous stress is aligned with the tectonic stress rate**. That is, one cannot use focal mechanism catalogs to infer the average stress orientation. Since focal mechanisms that are aligned with the stress rate tensor are favored, the inversion procedure tens to derive the orientation of the stress rate tensor. More

importantly, **the strong spatial heterogeneity in the stress is invisible to focal mechanism catalogs** (except for aftershocks). Figure 8-73 shows examples of focal mechanism catalogs that are generated by the procedure just described. Maps of the seismicity produced by different roughnesses, α , and heterogeneity ratios, HR, are shown. Notice that as α increases (i.e., the stochastic variation becomes smoother) the seismicity becomes more spatially clustered.

Figure 8-74 shows a comparison of the spatial coherence of focal mechanisms as measured by Hardebeck (2006) and a prediction of this value using the Smith and Heaton stochastic model. The horizontal axis is the distance between pairs of events, and the vertical axis is the average angular difference between the focal mechanisms of the pair. $\alpha = 0.75$ and $HR = 2.38$ produced simulations that were most similar to Hardebeck's data analysis. The Smith and Heaton study is difficult to describe in the simplified context of class notes. This is an important topic and I suggest that you attempt to read the entire manuscript (including Hardebeck's published comment and our reply).

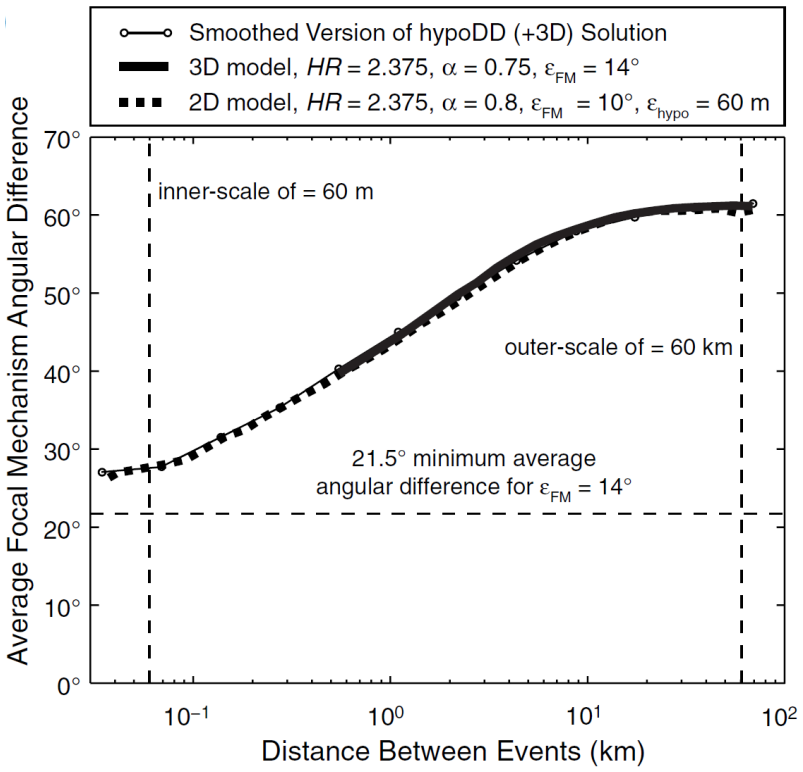


Figure 8-74. Comparison of the Smith and Heaton model for focal mechanisms in a stochastic stress field compared with the observation of Jeanne Hardebeck of a catalog of focal mechanisms in southern California (J. Hardebeck, 2006, Homogeneity of small-scale earthquake faulting, stress and fault strength, *Bulletin of the*

Seismological Society of America, *96*, 1675-1688). The best fit to Hardebeck's analysis is with $\alpha = 0.8$ and $HR = 2.4$.

In order to get a better intuitive understanding of our best fit stochastic model, we generated an example of one component of the shear stress on a 100-km grid and at a spacing of 1 cm. Since there are too many grid points to observe on a single plot, we plotted a line through the grid at a variety of scales in Figure 8-75. In particular, panel (a) shows 10,000 points on a line through the entire 100-km volume, panel (b) shows the same stress on a 10-km length that run from 15 km to 25 km, which is a region of higher stochastic stress. Panel (c) focuses in on 17.5 km to 18.5 km, which is the largest stress patch in the previous panel. Finally, panel (d) focuses in on a 100-m patch of the highest stress in panel (c). The high shear stress at 18.01 km will cause an earthquake nucleation at that point at some time in the future.

Average shear stress at Different Lengths

We can use Figure 8-75 to investigate the stress-based strength of the crust based on this model (see Figure 8-52 and equation 8.122). In particular, the spatially averaged shear stress is shown for each panel in Fig. 8-75. This average shear stress is about 190 MPa at 1 m, 174 MPa at 100m, 155 MPa at 1 km, 104 MPa at 10 km, and 60 MPa at 100 km. It is clear that, for this stress distribution, the average stress near potential nucleation points decreases with the length over which the stress is averaged.

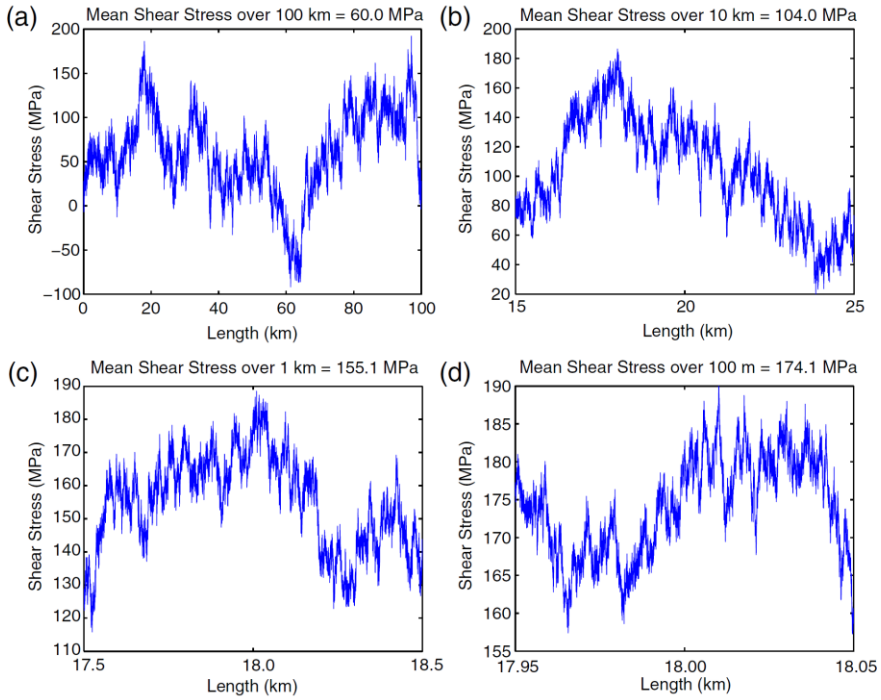


Figure 8-75 shows how the stress near a nucleation point (200 MPa yield) looks depending on the length scale. Clearly, the average stress in the longest length scale (panel a) in the 3-d fractal grid is lower than the average stress viewed at the shortest length scale (panel d).

Deborah Smith and I explored how this length scale dependence of average stress varies with the power-law smoothing parameter, α, using our stochastic stress model. The results of these numerical simulations are displayed in Figure 8-76. In this case, we assumed that

$$\overline{\sigma_L} \approx \left(\frac{1}{1+L} \right)^{\gamma} \qquad 8.129$$

where L is the scale length of the averaging and γ is a number determined from the stress generated in the model. Assuming that $\alpha \approx 0.8$ then indicates that the **average shear stress decreases as about the inverse fourth root of averaging length**. That is, changing from a length scale of 1 meter to a length scale of 10 km would decrease the average stress by a factor of ten.

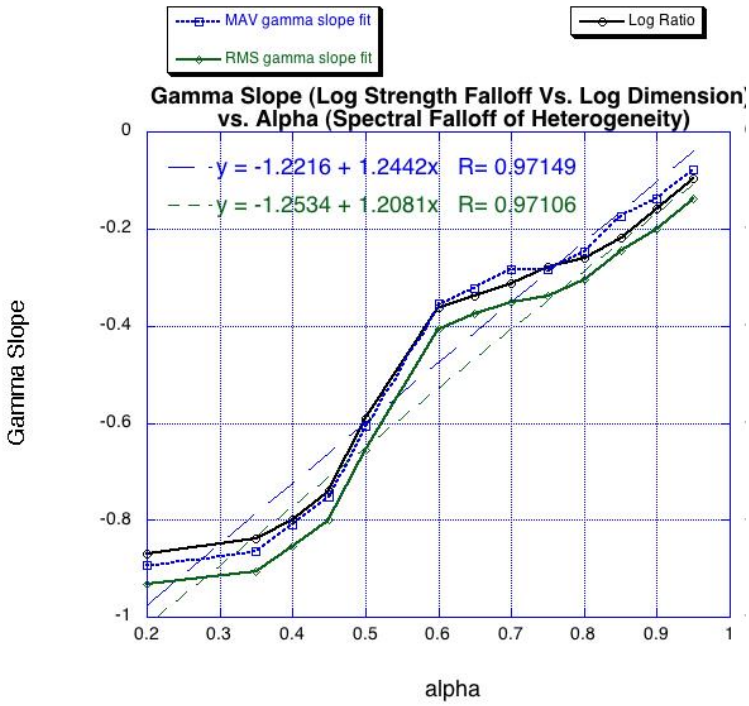


Figure 8-76. Based on simulations like that shown in Fig. 8-75, the dependence on stress-based strength on the length scale $\Gamma_\sigma \sim L^\gamma$ is shown for different values of fractal roughness.

There are numerous bold assumptions underlying this calculation and, unfortunately, it's almost impossible to directly test this model since the measurement of true stress is exceedingly difficult (i.e., expensive). The Cajon Pass Borehole experiment is one example of such a measurement. This was "the" major earthquake experiment of the late 1980's. It consisted of a 3.5 km deep borehole that was drilled about 3 km north of the San Andreas fault in Cajon Pass. With the view that rocks would be too damaged to measure if they were in the fault zone, the scientists managing the project chose to drill adjacent to the fault. A 300-m section of the downhole log of the orientation of borehole breakouts is shown in Figure 8-77. Panel (a) shows the observations. Note that at any given depth, there are two data points that are 180° different (breakouts on opposite sides of the borehole). Panel (b) shows the orientation of the principal compression axis along a 300-m line intersecting our preferred southern California stochastic stress model. Although the model and the observation are different, they do have similar variations over this length scale.

One of the most embarrassing observations of this expensive experiment was the fact that the spatial average of the orientation of the principal compression axis for the Cajon Pass borehole indicated that the San Andreas has an average shear stress that is left lateral (at least in the Cajon Pass). This is embarrassing since the

San Andreas fault is obviously a right-lateral fault. Furthermore, the geodetically measure strain rates are also right-lateral (see Figure 8-68). Barton and Zoback interpreted this observation to mean that the principal compression axis is almost perpendicular to the fault, which implies a very small shear stress on the fault. In our stochastic stress model, most regions are right-lateral. However, there are also significant regions of left-lateral stress.

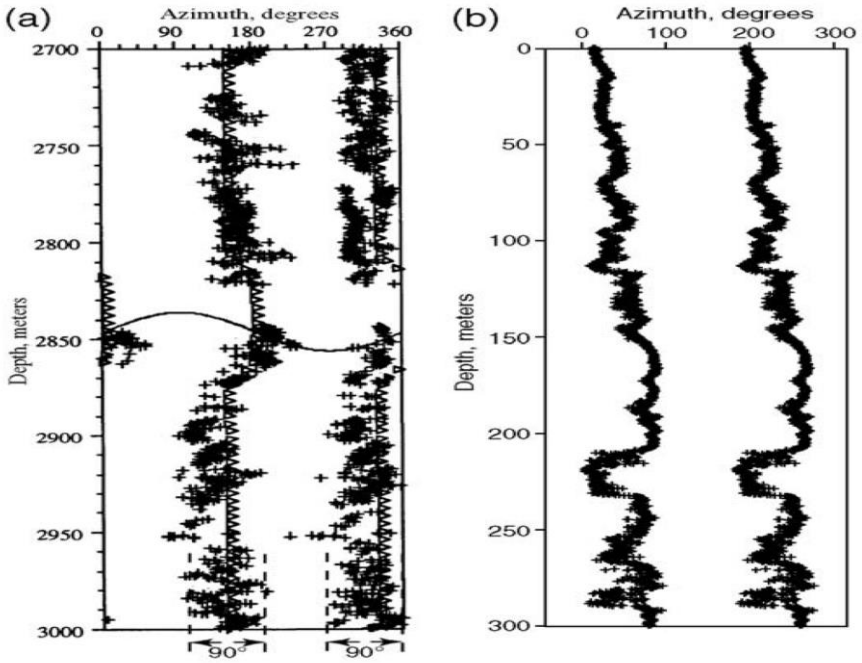


Figure 8-77. (a) orientation of borehole breakouts observed in a 300-m stretch of the Cajon Pass Borehole located 4 km from the San Andreas fault. Tensile stresses are maximum at opposite sides of a borehole and they are aligned with the maximum principal stress. (b) orientation of the maximum principal stress along a 300-m line transecting the preferred fractal model of Smith and Heaton

The stochastic stress model of Smith and Heaton is radically different from other models of stress that are derived from assumptions that the stress variations are small compared to the average. Most researchers that I have encountered do not seriously entertain the possibility that shear stress could actually be opposite to the direction of tectonic strain accumulation. I must admit that the stochastic power-law model we proposed only makes sense if rupture dynamics are truly chaotic. While I personally believe that the evidence points us in that direction, the reader

of these notes should always be cautious when discussing these issues (*caveat emptor*).

Failure Model that has low frictional heat, low fracture energy, and low radiated energy

Based on the insight obtained from the chaotic spring-block-slider model, the stochastic stress model of Smith and Heaton, and the gambler's ruin analysis of fault slip, I am now able to re-examine the stress paradox. That is, **is there a model of dynamic rupture that has the following critical features?**

1) high nucleation shear stress of about $0.6\sigma_n$

2) small frictional heating (less than 2 MJ/m^2)

3) small fracture energy (less than 500 kJ/m^2

4) radiated far-field seismic energy of about $D \times 2.0 MJ/m^2$.

The key to satisfying these elements is to recognize that 1) friction depends very strongly on slip speed; low-speed sliding (mm/s) produces friction similar to the Byerly estimate, while high-speed sliding (> m/s) is almost frictionless. The transition between these two types of friction seems to occur when about 500 kJ/m^2 of energy is available on the sliding surface. 2) This low dynamic friction produces unsteady slip pulses that can propagate through regions of much lower average stress (even localized patches of negative stress). 3) the chaotic system evolves to produce fractal-like prestress that is large enough to allow ruptures of any length (that is, the system is in a critical state).

The average stress drop of events depends on the distance that slip pulses propagate (heterogeneous slip is more likely to have smaller rupture lengths, and highly heterogeneous ruptures have average higher stress drops). Slip pulses propagate until they encounter low-stress regions. The Gutenberg-Richter relation is controlled by the spatial properties of the prestress. The prestress is the result of self-organization of a dynamic system that does not have inherent length scales.

One useful measure of this system is **seismic efficiency**, η_R, which is defined as the fraction of total energy that ends up as radiated seismic energy, or

$$\eta_R \equiv \frac{E_R}{\Delta W} = \frac{\Delta W - E_D}{\Delta W} = 1 - \frac{E_D}{\Delta W} \qquad 8.130$$

where E_R and E_D are radiated and dissipated energy, respectively. The observed low dissipation for large earthquakes implies that the seismic efficiency is high (approaching 1) for large earthquakes. That is, for large earthquakes,

$$\Delta W \Rightarrow E_{Radiation} = \sigma_{eff} P \approx 2MPa \times P \qquad 8.131$$

We know that the change in potential energy is given by

$$\Delta W = \int_{-W/2}^{W/2} \int_{-L/2}^{L/2} \frac{1}{2}\left(\sigma_0(x,y) + \sigma_1(x,y)\right) D(x,y)\, dxdy$$

$$= \int_{-W/2}^{W/2} \int_{-L/2}^{L/2} \sigma_0(x,y) D(x,y)\, dxdy - \frac{1}{2} \int_{-W/2}^{W/2} \int_{-L/2}^{L/2} \Delta\sigma(x,y) D(x,y)\, dxdy \qquad 8.132$$

To get some idea of how things work, I will simplify things by assuming that the rupture is approximately a line source. That is, $D(x,y) \approx D(x)\delta(y)$. Note that if the rupture physics is isotropic, then x can be in any direction. I can now evaluate the change in potential energy as

$$\Delta W = \int_{-L/2}^{L/2} \sigma_0(x) D(x)\, dx - \frac{1}{2} \int_{-L/2}^{L/2} \Delta\sigma(x) D(x)\, dx \qquad 8.133$$

The first integral is a difficult problem since $D(x)$ is a highly nonlinear function of $\sigma_0(x)$. In contrast, the second integral is entirely determined by $D(x)$ since $\Delta\sigma(x)$ is a linear function of $D(x)$. In particular, it is convenient to write this in terms of a Green's function.

$$\Delta\sigma(x) = D(x) * G_{\Delta\sigma}(x) \qquad 8.134$$

Where $G_{\Delta\sigma}(x)$ is a Green's function for a spatial impulse of slip. Uenishi and Rice (2003, JGR, doi:10.1029/2001JB001681) say that

$$\begin{aligned}\Delta\sigma(x) &= -\frac{\mu^*}{2\pi}\int_{-L/2}^{L/2}\frac{\frac{\partial D(\xi)}{\partial\xi}}{x-\xi}d\xi \\ &= \frac{\mu^*}{2\pi}\int_{-\infty}^{\infty}\frac{\frac{\partial D(\xi)}{\partial\xi}}{\xi-x}d\xi \\ &= \frac{\mu^*}{2\pi}\frac{\partial D(x)}{\partial x}*\frac{1}{x} \\ &= -\frac{\mu^*}{2}F_{Hi}\left[\frac{\partial D(x)}{\partial x}\right]\end{aligned} \tag{8.135}$$

Where $\mu^* = \mu$ for mode III and $\mu^* = \mu/(1-\nu)$ for modes I and II. F_{Hi} signifies a Hilbert transform. This can also be written

$$\Delta\sigma(x) = \frac{\mu^*}{2}\frac{\partial}{\partial x}F_{Hi}\left(D(x)\right) \tag{8.136}$$

Therefore,

$$\Delta\sigma D = -D(x)\frac{\mu^*}{2\pi}F_{Hi}\left(D'(x)\right) \tag{8.137}$$

I can take the Fourier transform of 8.137 to obtain

$$\begin{aligned}FT(\Delta\sigma D) &= \Delta\tilde{\sigma}(k)*\tilde{D}(k) \\ &= -\frac{\mu^*}{2\pi}\tilde{D}(k)*F_{Hi}\left(ik\tilde{D}\right) \\ &= -\frac{\mu^*}{2\pi}\tilde{D}*\left[-i^2k\,\mathrm{sgn}(k)\tilde{D}\right] \\ &= \frac{\mu^*}{2\pi}\left(k\,\mathrm{sgn}(k)\tilde{D}\right)*\tilde{D}\end{aligned} \tag{8.138}$$

And then, using Parseval's theorem

$$\int_{-L/2}^{L/2} \Delta\sigma D dx = \frac{\mu}{2\pi} \int_{-\infty}^{\infty} D(x) F_{Hi}\left(D'(x)\right) dx$$

$$= \frac{\mu}{2\pi} \int_{-\infty}^{\infty} \left[k \operatorname{sgn}(k) \tilde{D} * \tilde{D} \right] dk \qquad 8.139$$

$$=^{?} \frac{\mu}{2\pi} \int_{-\infty}^{\infty} \left[k \tilde{D}^2 \right] dk$$

I now need to estimate $\tilde{D}^2(k)$. I begin with a statistical description of $D(x)$.

I will assume that at each point along the rupture, the slip is chosen from a random number generator with a pdf described by a mean-zero Normal distribution with a variance of σ^2. Meier, Ampuero and Heaton investigated the Potency rate functions that resulted from finite source inversions of large earthquakes. They found that

$$\dot{P}(t) \approx \left[\Pi\left(\frac{t}{2L/V_R} \right) * \Pi\left(\frac{t}{2L/V_R} \right) \right] \left(1 - R(x)\right) \qquad 8.140$$

Where $R(x, \sigma = 0.38)$ is a number chosen from a group of random numbers whose frequency (probability density function, or pdf) is described by a Gaussian function, $P = \frac{e^{-R^2/2\sigma^2}}{\sigma\sqrt{2\pi}}$. That is, the random numbers are from a Normal distribution with a standard deviation $\sigma \approx 0.38$. 8.140 is an isosceles triangle (the expected value) multiplied by a random number that has a normal pdf with a mean of 1. Almost all of the random numbers are less than 1, so this multiplicative factor is almost entirely positive). It's important to realize that this functional description of spatially varying slip is based on fitting observed data (a rather complex observation); it's not clear that this functional form is derivable from physics.

Now recall (from Chapter 7) that, for a slip pulse propagating along a long narrow rupture (i.e., a line source), the Potency rate maps directly into the slip distribution. That is, $D(x) \approx \dot{P}(V_R t)$. From this, I infer that

$$D(x) \approx 6 \times 10^{-5} L \left[\Pi\left(\frac{x}{2L} \right) * \Pi\left(\frac{x}{2L} \right) \right] \left(1 - R(x)\right) \qquad 8.141$$

The scaling factor, $6\times10^{-5}L$, is from the observed ratios of average slip to rupture length (see 8.31). I can now attempt to calculate the Fourier transform of 8.141, which gives

8.142

Unfortunately, I am unable to take the Fourier Transform of 1. I note that I can write that

$$D^2(x) \approx E\left(R^2(x)\right)\left\{6\times10^{-5}L\left[\Pi\left(\frac{x}{2L}\right)*\Pi\left(\frac{x}{2L}\right)\right]\right\}^2 \qquad 8.143$$

Now the power of the random variable term is just the energy density of the spatially filtered random numbers. The final density is modulated in space by the triangle function. I think I can do my integral in the space domain by integrating the modulated power as a function of x.

$$\Delta W \approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\tilde{\tilde{\sigma}}_0\left(k_x,k_y\right)-\frac{1}{2}\Delta\tilde{\tilde{\sigma}}\left(k_x,k_y\right)\right)*\tilde{\tilde{D}}^*\left(k_x,k_y\right)*\Pi\left(Lk_x\right)*\Pi\left(Wk_y\right)dk_xdk_y$$

8.144

$$\Delta W = \int_{-\infty}^{\infty}\left(\tilde{\sigma}_0\left(k_x\right)-\frac{1}{2}\tilde{D}\left(k_x\right)\tilde{G}_{\Delta\sigma}\left(k_x\right)\right)*\tilde{D}^*\left(k_x\right)*\Pi\left(Lk_x\right)dk_x \qquad 8.145$$

Now the stress drop is a linear function of the slip distribution. That is, it can be calculated using a linear Green's function, $G_{\Delta\sigma}(x)$ or

$$\Delta\sigma(x) = D(x)*G_{\Delta\sigma}(x) \qquad 8.146$$

This convolution can also be defined in the wavenumber domain as

$$\Delta\tilde{\sigma}\left(k_x\right) = \tilde{D}\left(k_x\right)\tilde{G}_{\Delta\sigma}\left(k_x\right) \qquad 8.147$$

Unfortunately, I do not know what $G_{\Delta\sigma}(x)$ is, but for now,

The convolution with a rectangle is identical to performing a running-mean lo-pass filter (width $1/L$) on the integrand. Unfortunately, 8.145 is a rather cumbersome expression.

$$\Delta W = \int_{-\infty}^{\infty}\left[\tilde{\sigma}_0(k_x) - \frac{1}{2}\tilde{D}(k_x)\tilde{G}_{\Delta\sigma}(k_x)\right] * \tilde{D}^*(k_x) * \mathrm{sinc}(Lk_x)\,dk_x$$

$$\approx \int_{-\infty}^{\infty}\left[\tilde{\sigma}_0(k_x) * \mathrm{sinc}(Lk_x) * \tilde{D}^*(k_x) - \frac{1}{2}\tilde{D}(k_x)\tilde{G}_{\Delta\sigma}(k_x) * \tilde{D}^*(k_x) * \mathrm{sinc}(Lk_x)\right]dk_x$$

8.148

I can now take the Fourier transform of **Error! Reference source not found.**.

$$\tilde{G}_{\Delta\sigma}(k_x) = i\mu k_x\left(1+\sqrt{k_x}\right) \qquad 8.149$$

Where I used the fact that $FT\left(\sqrt{x}\right) = \sqrt{k_x}$. I am tired of explicitly writing everything as a function of k_x. I can now rewrite 8.148 as

$$\Delta W \approx \int_{-\infty}^{\infty}\left[\tilde{\sigma}_0 * \Pi(Lk_x) * \tilde{D}^* - \frac{1}{2}\mu i k_x\left(1+\sqrt{k_x}\right)\tilde{D} * \tilde{D}^* * \Pi(Lk_x)\right]dk_x$$

$$= \int_{-\infty}^{\infty}\left[\tilde{\sigma}_0 * \Pi(Lk_x) * \tilde{D}^* - \frac{1}{2}\mu i k_x\left(1+\sqrt{k_x}\right)\tilde{D}^2 * \Pi(Lk_x)\right]dk_x$$

8.150

At this point, I need to emphasize that we have little hope of obtaining accurate measurements of the prestress or the slip distribution, especially since we strongly expect that these are complicated functions of space. To deal with this, I will assume that these parameters can by described by their probability density functions. For the following section, you should understand that when I write prestress and slip, I actually mean that there are pdf's that describes the parameters for each value of x. I will then describe the distribution with an expected value and some pdf about the expected value. Now I will make the critical assumption that prestress and slip are random stationary functions of space. That is, I assume that statistical properties are invariant with respect to position. Furthermore, when I say the "change in potential energy," I really mean the *expected value of the change in potential energy*.

I need to be careful with 8.150, since it has terms that are the product of stochastic variables, that is, $\tilde{\sigma}_0(k_x) * \tilde{D}^*(k_x)$ and also $\tilde{D}^*(k_x)\tilde{D}^*(k_x)$. These products are also stochastic variables, with the following properties

$$E\left[\tilde{\sigma}_0(k_x)\tilde{D}(k_x)\right] = E\left[\tilde{\sigma}_0(k_x)\right]E\left[\tilde{D}(k_x)\right] + \mathrm{cov}\left[\tilde{\sigma}_0(k_x), \tilde{D}(k_x)\right]$$

8.151

And

$$E\left[\tilde{D}^*(k_x)\tilde{D}(k_x)\right] = E\left[\tilde{D}^*(k_x)\right]E\left[\tilde{D}(k_x)\right] + \mathrm{cov}\left[\tilde{D}^*(k_x),\tilde{D}(k_x)\right]$$
$$= E^2\left[\tilde{D}^*(k_x)\right] + \mathrm{var}\left[\tilde{D}(k_x)\right]$$

8.152

Where cov is the covariance function, which is defined as

$$\mathrm{cov}\left[\tilde{\sigma}_0(k_x),\tilde{D}(k_x)\right] \equiv E\left[\left(\tilde{\sigma}_0(k_x) - E\left[\tilde{\sigma}_0(k_x)\right]\right)\left(\tilde{D}(k_x) - E\left[\tilde{D}(k_x)\right]\right)\right]$$

8.153

Covariance is a measure of the correlation between random variables; when the covariance is positive, then two variables tend to vary together (positive variations in one variable tend to occur when there are positive variations in the other variable. When the variables are statistically independent, then their covariance is zero. I am getting tired of explicitly writing that my random variables are a function of k_x, so I will drop that formality. I can obtain the expected change in strain energy by substituting 8.151 and 8.152 into 8.150.

$$E[\Delta W] \approx \int_{1/L}^{\infty}\left[2E[\tilde{\sigma}_0]E\left[\tilde{D}^*\right] + 2\,\mathrm{cov}\left(\tilde{\sigma}_0,\tilde{D}^*\right) - \left\{E\left[\tilde{D}^2\right] + \mathrm{var}\left(\tilde{D}^*\right)\right\}\mu i k_x\left(1+\sqrt{k_x}\right)\right]dk_x$$

8.154

At this point, I will assume that my random variables, σ_0 and D, are a *zero-mean, Gaussian, stationary process*. That is, at each value of x, I assign a value by picking a random number, where the probability of the number is given by the Gaussian function. Basically, I start with Gaussian white noise and I then apply a spatial filter power-law that is a power law in wavenumber space (I spatially smooth the random white noise). I now note that the Fourier transform of Gaussian white noise is Gausian white noise. Furthermore, it is easy to show that

$$E\left(\tilde{D}^2\right) = \mathrm{var}\left(\tilde{D}\right) + E^2\left(\tilde{D}\right) \qquad 8.155$$

Since I assumed that $D(x)$ and $\sigma_0(x)$ are mean zero, then it follows that $\tilde{D}(k)$ and $\tilde{\sigma}_0(k_x)$ are also mean zero. Therefore 8.154 becomes

$$E[\Delta W] \approx 2\int_{1/L}^{\infty}\left[\mathrm{cov}\left(\tilde{\sigma}_0,\tilde{D}^*\right) - \mathrm{var}\left(\tilde{D}^*\right)\mu i k_x\left(1+\sqrt{k_x}\right)\right]dk_x \qquad 8.156$$

Now the covariance term can be written as

$$\operatorname{cov}\left(\tilde{\sigma}_0, \tilde{D}\right) = \operatorname{corr}\left(\tilde{\sigma}_0, \tilde{D}\right)\sqrt{\operatorname{var}\left(\tilde{\sigma}_0\right)\operatorname{var}\left(\tilde{D}\right)} \qquad 8.157$$

Where corr is the correlation coefficient between prestess and the slip. A correlation of +1 means that variables are perfectly correlated, -1 if they are oppositely correlated, and 0 if they are statistically independent. So 8.156 can be written

$$E\left[\Delta W\right] \approx 2\int_{1/L}^{\infty}\left[\operatorname{corr}\left(\tilde{\sigma}_0, \tilde{D}\right)\sqrt{\operatorname{var}\left(\tilde{\sigma}_0\right)\operatorname{var}\left(\tilde{D}\right)} - \operatorname{var}\left(\tilde{D}^*\right)\mu i k_x\left(1+\sqrt{k_x}\right)\right]dk_x$$

8.158

At this point, I will assume that my variables are wave-number filtered versions of Gaussian white noise with a variance of 1. That is, I assume that

$$\operatorname{var}\left(\tilde{\sigma}_0\right) \approx \left(\frac{1}{\left(1+k_x\right)^{\beta}}\right)^2 \qquad 8.159$$

And

$$\operatorname{var}\left(\tilde{D}\right) \approx \left(\frac{1}{\left(1+k_x\right)^{\alpha}}\right)^2 \qquad 8.160$$

Substituting these into 8.158

$$E\left[\Delta W\right] \approx 2\int_{1/L}^{\infty}\left[\operatorname{corr}\left(\tilde{\sigma}_0, \tilde{D}\right)\frac{1}{\left(1+k_x\right)^{\beta}} - \mu k_x\left(1+\sqrt{k_x}\right)\right]\frac{1}{\left(1+k_x\right)^{\alpha}}dk_x$$

$$= 2\operatorname{corr}\left(\tilde{\sigma}_0, \tilde{D}\right)\int_{1/L}^{\infty}\frac{1}{\left(1+k_x\right)^{\beta}}dk_x - \mu\int_{1/L}^{\infty}\frac{k_x + k_x^{3/2}}{\left(1+k_x\right)^{\alpha}}dk_x$$

$$= 2\operatorname{corr}\left(\tilde{\sigma}_0, \tilde{D}\right)\left|\frac{\left(1+k_x\right)^{\beta+1}}{\beta+1}\right|_{1/L}^{\infty} - \mu\left|\frac{\left(1+k_x\right)^{1-\alpha}\left[\left(\alpha-1\right)k_x+1\right]}{\left(\alpha-2\right)\left(\alpha-1\right)} + \frac{k_x^{5/2-\alpha}}{5/2-\alpha}\right|_{1/L}^{\infty}$$

8.161

can simplify it by assuming that the prestress, slip, and stress drop are isotropic on the fault plane. That is, I will assume that

$$D(x,y) \approx D_r(r) \tag{8.162}$$

Where $r = \sqrt{x^2 + y^2}$. At this point, I need to modify my 2-d Fourier transform such that it only depends on r. In particular,

$$\begin{aligned}
\tilde{\tilde{D}}(k_x, k_y) &\equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D(x,y) e^{-i2\pi(xk_x + yk_y)} dxdy \\
&= \int_0^{\infty} D_r(r) \left[\int_0^{2\pi} e^{-i2\pi(rk_r \cos\theta)} d\theta \right] rdr \\
&= 2\pi \int_0^{\infty} D_r(r) J_0(2\pi r k_r) rdr \\
&\equiv \hat{D}_r(k_r)
\end{aligned} \tag{8.163}$$

That is, I will greatly simplify the problem by assuming that the Fourier spectrum of slip is approximately given by

$$\tilde{\tilde{D}}(k_x, k_y) \approx \hat{D}_r(k_r) \tag{8.164}$$

Where the **hat connotes Hankel transform of zero order**, J_0 is a Bessel function of the first kind, and $k_r \equiv \sqrt{k_x^2 + k_y^2}$. A Hankel transform is very similar to our familiar Fourier transform, except that the kernel function is a Bessel function instead of a sinusoid. Similar to a sinusoid, $J_0(r)$ oscillates with increasing r. However a Bessel function's amplitude decays with distance and the asymptotic expansion is

$$J_0(r) \approx \sqrt{\frac{2}{\pi r}} \cos\left(r - \frac{\pi}{4} \right) \tag{8.165}$$

The Hankel transform has the following properties

$$\hat{D}(k_r) \equiv 2\pi \int_0^{\infty} D_r(r) J_0(2\pi r k_r) dr \tag{8.166}$$

$$D_r(r) = 2\pi \int_0^{\infty} \hat{D}_r(k_r) J_0(2\pi r k_r) k_r dk_r \tag{8.167}$$

This is an unrealistic assumption that allows me to simplify the math. In particular, 8.145 becomes

$$\Delta W = \int_{-\infty}^{\infty}\int_{0}^{2\pi}\left(\tilde{\sigma}_0(k_r) - \frac{1}{2}\tilde{D}(k_r)\tilde{G}(k_r)\right)\tilde{D}(k_r)\,\mathrm{sinc}(Rk_r)\left(k_r^{-1}d\theta\right)dk_r$$

$$= 2\pi\int_{-\infty}^{\infty}\left(\tilde{\sigma}_0(k_r) - \frac{1}{2}\tilde{D}(k_r)\tilde{G}(k_r)\right)k_r^{-1}\tilde{D}(k_r)\,\mathrm{sinc}(Rk_r)\,dk_r$$

8.168

Now I will note that a sinc function with scale width of R can be crudely approximated by a Rectangle function with a scale width of R. 8.168 can be approximated as

$$\Delta W = 2\pi\int_{-\infty}^{\infty}\left(\tilde{\sigma}_0(k_r) - \frac{1}{2}\tilde{D}(k_r)\tilde{G}(k_r)\right)k_r^{-1}\tilde{D}(k_r)\,\Pi(Rk_r)\,dk_r$$

$$= 2\pi\int_{0}^{1/R}\left(\tilde{\sigma}_0(k_r) - \frac{1}{2}\tilde{D}(k_r)\tilde{G}(k_r)\right)k_r^{-1}\tilde{D}(k_r)\,dk_r$$

$$= 2\pi\int_{0}^{1/R}\frac{\tilde{\sigma}_0(k_r)\tilde{D}(k_r)}{k_r}dk_r - \pi\int_{0}^{1/R}\frac{\tilde{D}^2(k_r)\tilde{G}(k_r)}{k_r}dk_r \qquad 8.169$$

$$= 2\pi\int_{0}^{1/R}\frac{\tilde{\sigma}_0(k_r)\tilde{D}(k_r)}{k_r}dk_r - \pi\int_{0}^{1/R}\frac{\tilde{D}(k_r)\Delta\tilde{\sigma}(k_r)}{k_r}dk_r$$

At this point, I will assume that the prestress and the slip are both random variables with pdf's given by

$$\tilde{\sigma}_0(k_r) \approx C_{\sigma_0}k^{-\alpha} \qquad 8.170$$

$$\tilde{D}(k_r) \approx C_D k^{-\beta} \qquad 8.171$$

I can now approximate the change in potential energy 8.132 as

$$\Delta W = 2\pi \int_0^{1/R} E\left[\frac{\tilde{\sigma}_0(k_r)\tilde{D}(k_r)}{k_r}\right]dk_r - \pi \int_0^{1/R} E\left[\frac{\tilde{D}(k_r)\Delta\tilde{\sigma}(k_r)}{k_r}\right]dk_r$$

$$= 2\pi \int_0^{1/R} \left\{E\left[\tilde{\sigma}_0(k_r)\right]E\left[\tilde{D}(k_r)\right] + \mathrm{cov}\left[\tilde{\sigma}_0(k_r), \tilde{D}(k_r)\right]\right\} k_r^{-1}dk_r - \pi \int_0^{1/R} E\left[\tilde{D}(k_r)\Delta\tilde{\sigma}(k_r)\right]k_r^{-1}dk_r$$

8.172

$$\begin{aligned}\Delta W &= \iint_S \left(\sigma_0(x,y) - \frac{1}{2}\mu f(y)\frac{\partial}{\partial x}D(x)\right)D(x)f(y)\,dxdy \\ &= \iint_S \left(\sigma_0(x,y)f(y)D(x) - \frac{1}{4}\mu f^2(y)\frac{\partial}{\partial x}D^2(x)\right)dxdy\end{aligned} \qquad 8.173$$

For the moment, I will greatly simplify things by assuming that $f(y) \approx \Pi(y)$. I can now rewrite 8.173 as

$$\Delta W \approx W \int_0^L \left(\sigma_0(x)D(x) - \frac{1}{4}\mu\frac{\partial}{\partial x}D^2(x)\right)dx \qquad 8.174$$

The observation that the ratio of $\bar{D}/L \approx 10^{-5}$ can be produced by a random-phase slip that has an amplitude spectrum of $\tilde{D}(k) = k^{-\alpha}$, where $\alpha \approx 1.2$ (see Zheng-Liu and others). From 8.130,

$$\Delta W = \frac{E_R}{\eta} \qquad 8.175$$

So 8.175 is rewritten as

$$E_R \approx \eta W \int_0^L \left(\sigma_0(x)D(x) - \frac{1}{4}\mu\frac{\partial}{\partial x}D^2(x)\right)dx \qquad 8.176$$

Now radiated energy is estimated to be $E_R \approx \sigma_{eff} P$ (see 8.7), so 8.176 becomes

$$E_R \approx \sigma_{eff} P \approx \eta W \int_0^L \left(\sigma_0(x)D(x) - \frac{1}{4}\mu\frac{\partial}{\partial x}D^2(x)\right)dx \qquad 8.177$$

Or

$$\sigma_{eff} W \int_0^L D(x)dx \approx \eta W \int_0^L \left(\sigma_0(x)D(x) - \frac{1}{4}\mu\frac{\partial}{\partial x}D^2(x)\right)dx \qquad 8.178$$

$$\sigma_{eff}\int_0^L\left[D(x)+\frac{1}{4}\mu\frac{\partial}{\partial x}D^2(x)\right]dx\approx\eta\int_0^L\left(\sigma_0(x)D(x)\right)dx$$

$$\sigma_{eff}\int_0^L\left[D(x)+\frac{1}{2}\mu D(x)\frac{\partial}{\partial x}D(x)\right]dx\approx\eta\int_0^L\left(\sigma_0(x)D(x)\right)dx$$

$$\sigma_{eff}\int_0^L\left[1+\frac{1}{2}\mu\frac{\partial}{\partial x}D(x)\right]D(x)dx\approx\eta\int_0^L\left(\sigma_0(x)D(x)\right)dx \qquad 8.179$$

$$\Rightarrow\sigma_{eff}D(x)\left[1+\frac{1}{2}\mu\frac{\partial}{\partial x}D(x)\right]=\eta\sigma_0(x)D(x)$$

$$\sigma_{eff}\left[1+\frac{1}{2}\mu\frac{\partial}{\partial x}D(x)\right]=\eta\sigma_0(x)$$

$$E_D=\iint_S\sigma_F\left(x,y;D,\dot{D}\right)D(x,y)\,dxdy \qquad 8.180$$

IF we assume that the prestress and final stress are uniform in x and y, then 8.132 becomes

$$\Delta W_{uniform}=\frac{1}{2}(\sigma_0+\sigma_1)\bar{D}S \qquad 8.181$$

Of course the prestress and final stress can only be uniform for one unique distribution of slip which depends on the geometry of the rupture surface. If we make the very restrictive assumption that the friction is uniform and equal to the final stress, then

$$E_D=\sigma_1\bar{D}S+E_G \qquad 8.182$$

Where E_G is fracture energy. If we assume that the radiated energy is the change in potential energy minus the dissipation energy, then in the case of uniform stress and friction,

$$\begin{aligned}E_R&=\frac{1}{2}(\sigma_0+\sigma_1)\bar{D}S-\sigma_1\bar{D}S-E_G=\frac{1}{2}(\sigma_0-\sigma_1)\bar{D}S-E_G\\&=\frac{1}{2}(\Delta\sigma)\bar{D}S-E_G\\&\equiv\Delta W_0-E_G\end{aligned} \qquad 8.183$$

Where E_G is the fracture energy, which is meant to signify the transition from static to dynamic friction (sometimes called the breakdown energy). In the case of slip weakening friction, $E_G = \frac{1}{2}\left(\sigma_{yield} - \sigma_1\right) D_0 S$. Assuming slip weakening friction, the potential energy change available for seismic radiation is $\Delta W_0 \equiv \frac{1}{2}\Delta\sigma \bar{D} S - E_G$. Unfortunately, it is not possible for seismologists to determine absolute stress, we can only infer stress changes. In response, many seismologists have substituted ΔW_0 for ΔW in their definition of seismic efficiency. I will call this "crack seismic efficiency" (η_{crack}). In this case

$$\eta_{crack}^{observed} \equiv \frac{E_R^{observed}}{\Delta W_0} = \frac{2E_R^{observed}}{\Delta\sigma \bar{D} S} \tag{8.184}$$

We also know that

$$\begin{aligned} \eta_{crack} &= 1 - \frac{E_G}{\Delta W_0} \\ &= 1 - \frac{\left(\sigma_{yield} - \sigma_1\right) D_0}{\Delta\sigma \bar{D}} \end{aligned} \tag{8.185}$$

Or

$$D_0 = \frac{\left(1 - \eta_{crack}\right)\Delta\sigma \bar{D}}{\sigma_{yield} - \sigma_1} \tag{8.186}$$

In principle, seismologically observed quantities can be used to determine everything in 8.184. Then assuming that the Byerly friction is the yield stress, then the slip weakening distance can be determined, that is provided that σ_1 is known (which it is not). However, in this case, $E_G^{crack} = \left(1 - \eta_{crack}\right)\Delta\sigma \bar{D} S$.

Equations 8.181 through8.186 all assume that the frictional sliding stress is the final stress after all rupture has ceased. The resulting definition of the change in potential energy is **very, very** restrictive. It assumes constant friction and it thus **excludes slip pulses**. Furthermore, all of these definitions assume that the prestress and stress drop are spatially uniform. This is clearly not the case. In particular, the slip and the stress drop are both complex functions of space that

are correlated. Multiplying their average values is clearly inappropriate. The best description of the radiated energy is to subtract 8.180 from 8.132, or

$$E_R = \iint_S \left[\sigma_0(x,y) - \frac{1}{2}\Delta\sigma(x,y) - \sigma_F\left(D,\dot{D}\right) \right] D(x,y)\,dxdy$$

8.187

As you are aware, I believe that the prestress and stress drop is very heterogeneous. Unfortunately, I don't know any simple average relations between radiated energy, slip and stress drop.

With regards to the fracture energy, I like the definition that $E_G \equiv \left(E_f - D\sigma_{\min}\right)dS$. This is rather arbitrary. However, the key issue about dynamics is whether or not the energy available to radiate increases with additional rupture. That is, if $\frac{dE_R}{dt} = 0$, then the event has ended. This can be accomplished solely with the distribution of prestress, which then determines the distributions of friction and slip. One does not need large fracture energies to stop an event. If the friction is strongly rate weakening, it produces slip pulses and the relationship between slip and dynamic friction and stress drop becomes very complex (see Elbanna and Heaton).

One potential research direction is to use seismological evidence to determine $E_R^{observed}$, $D(x,y)$, and $\Delta\sigma(x,y)$. We could then assume a friction law together with the assumption that slip is a Joffe slip pulse to determine $E_F(x,y) = F_{Joffe}\left(D(x,y)\right)$. This is similar to the approach taken in Elbanna and Heaton. In this case we can anticipate that $E_F \approx CD^{\nu}$, where C and ν are constants obtained from numerical simulations of pulses. Then 8.187 could be written

$$\begin{aligned} E_R^{observed} &= \iint_S \sigma_0(x,y) D(x,y)\,dxdy - \iint_S \left\{ CD(x,y)^{\nu} + \frac{1}{2}\Delta\sigma(x,y) D(x,y) \right\} dxdy \\ &= \iint_S \left\{ \left[\sigma_0(x,y) - \frac{1}{2}\Delta\sigma(x,y) \right] D(x,y) - CD(x,y)^{\nu} \right\} dxdy \end{aligned}$$

8.188

Finally, we can constrain put a constraint on the prestress.

$$\iint_S \sigma_0(x,y)D(x,y)\,dxdy = E_R^{observed} + \iint_S \left\{ \frac{1}{2}\Delta\sigma(x,y)D(x,y) + CD(x,y)^{\nu} \right\} dxdy$$

8.189

The system could work consistently even with no friction or fracture energy as long as

$$E_R^{observed} = \iint_S \left\{ \left[\sigma_0(x,y) - \frac{1}{2}\Delta\sigma(x,y) \right] D(x,y) \right\} dxdy \quad 8.190$$

As an example we use the relationship between radiated energy and seismic moment of Kanamori. We recall that, on average,

$$E_R \approx \frac{M_0}{2\times 10^4} = \frac{\mu \bar{D} S}{2\times 10^4} = \sigma_e \bar{D} S = \sigma_e P \quad 8.191$$

Where effective energy $\sigma_e \approx 2MPa$. Substituting this into 8.188, we obtain

$$\sigma_e S\bar{D} \approx= \iint_S \left\{ \left[\sigma_0(x,y) - \frac{1}{2}\Delta\sigma(x,y) \right] D(x,y) - CD(x,y)^{\nu} \right\} dxdy$$

8.192

If we assume that $\frac{\partial}{\partial D}\left[\sigma_0(x,y) - \frac{1}{2}\Delta\sigma(x,y) \right] = 0$, or that the final slip at a point is independent from the difference between the initial stress and the stress drop at the point (implausible assumption), then 8.192 becomes,

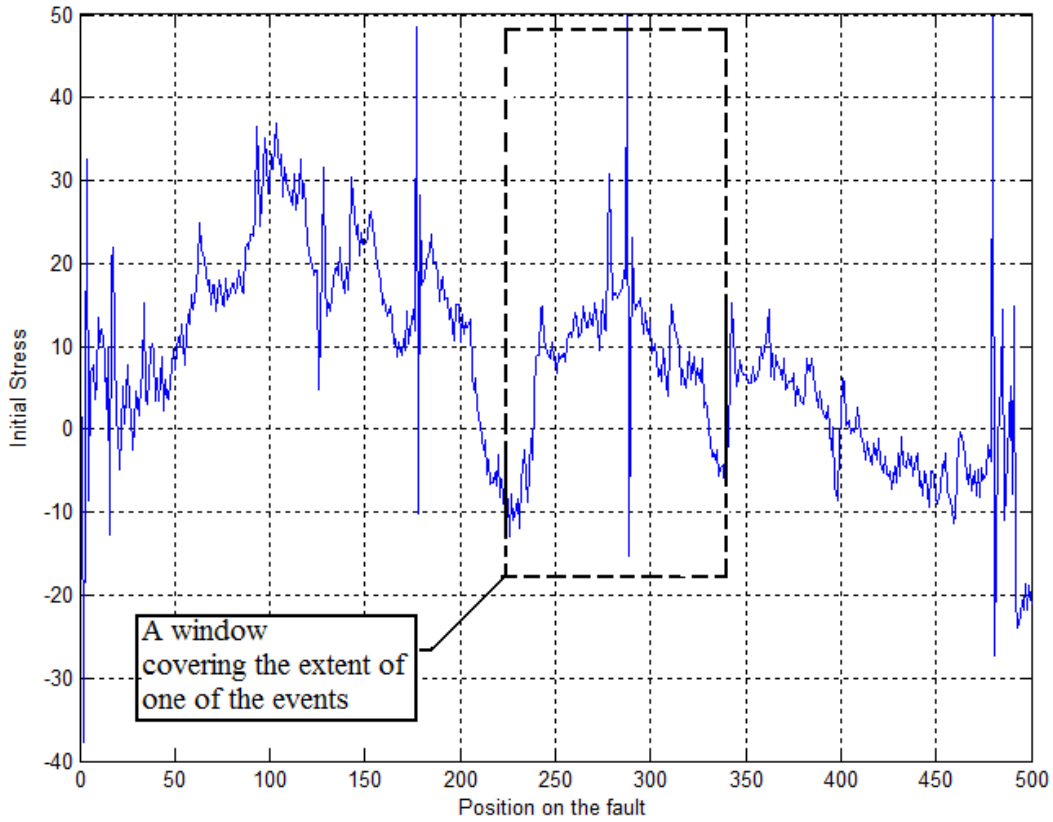
$$\sigma_e \approx \overline{\left[\sigma_0(x,y) - \frac{1}{2}\Delta\sigma(x,y) \right]} - \frac{E_F}{P} \quad 8.193$$

If $\overline{\Delta\sigma} \approx 3MPa$, then $\overline{\sigma_0(x,y)} \approx \sigma_e + 1.5MPa + \frac{E_F}{P} \approx 3.5MPa + \frac{E_F}{P}$. The fact that there is minimal melting implies that $\frac{E_F}{P} < 2Mpa$. then $\overline{\sigma_0(x,y)} < 5.5MPa$. This is a restatement of the stress paradox. Stresses inferred from seismology are far smaller than stresses measured in the laboratory, and they are small compared to the stresses required to resist gravity in mountain ranges. I am convinced that the key to resolving these problems is to establish the detailed connection between prestress and stress drop. In dynamic models that produce slip pulses (strong rate weakening friction), the amplitude of the slip

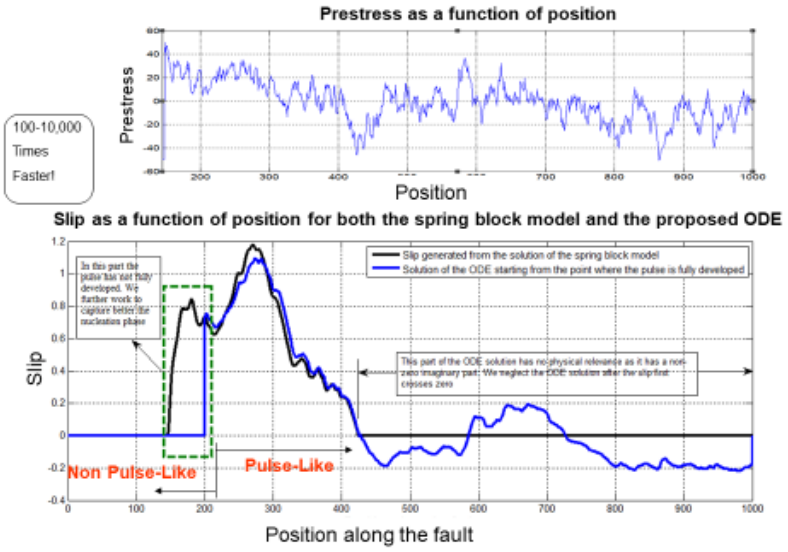
pulse varies in a complex way that is mostly determined by the prestress and the hypocentral location (see the pulse energy equation by Elbanna and Heaton).

Thank you to Hiroo Kanamori, Valere Lambert, and Victor Tsai for the comments.

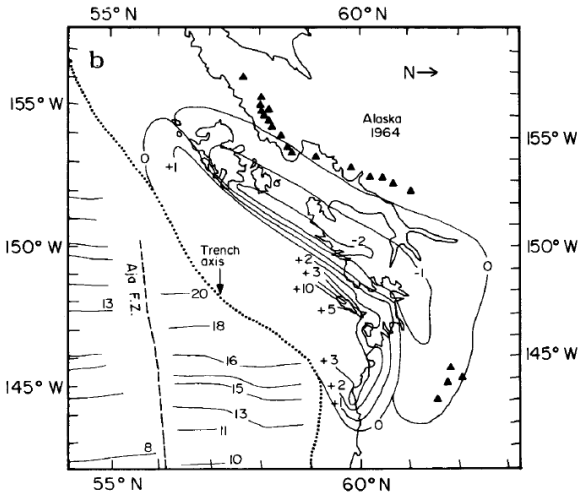
Strength-length scaling
b-value and prestress
b-value and brittleness
creeping fault
discrete vs continuum
far from equilibrium Reynolds number

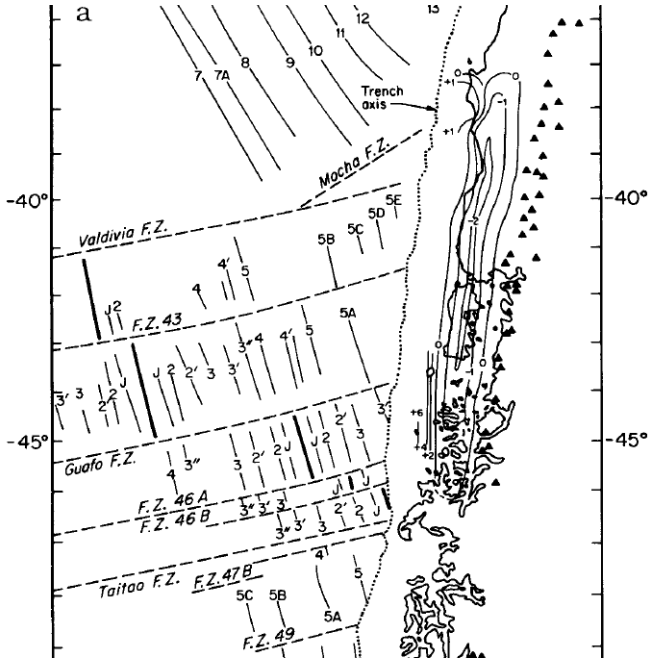


A Test Case



32





Derivation of Brune spectrum from the observation of near-source high-frequency magnitude saturation

We observe that very near-source accelerations (high-frequencies) appear to be incoherent noise with a peak acceleration that is independent of magnitude. This leads us to the following hypothesis.

Hypothesis: Radiated high-frequency energy $E_R^{\omega \gg \omega_c}$ scales with the rupture area S , independent of the average slip on that rupture surface, or

$$E_R^{\omega \gg \omega_c} \sim S \qquad 8.2$$

This hypothesis means that if we double the rupture area, then we double the radiated high-frequency energy.

What does this mean for the scaling of radiated seismic waves?

Let us **assume** that the **seismic wave is approximately a non-negative function of duration** T_c, whose **integrated area scales with seismic moment, and which has a power-law high-frequency spectral decay**. Then,

$$\tilde{U}_R \sim M_0 \text{sinc}^{\alpha}\left(\omega/\omega_c\right) \tag{8.3}$$

Where $\omega_c = 2\pi/T_c$. Equation **Error! Reference source not found.** has asymptotes

$$\tilde{U}_R \sim \begin{cases} M_0 & \omega \ll \omega_c \\ M_0\left(\omega/\omega_c\right)^{-\alpha} & \omega \gg \omega_c \end{cases} \tag{8.194}$$

Now let us **assume that the duration of the signal is proportional to the dimension of the fault**, or that $\omega_c \sim \sqrt{S}$. In addition, **we will assume that** $M_0 \sim S\bar{D} \sim S^{3/2}$. Therefore, $\omega_c \sim M_0^{-1/3}$, and 8.194 becomes

$$\tilde{U}_R \sim \begin{cases} M_0 & \omega \ll \omega_c \\ M_0^{1-\alpha/3}\omega^{-\alpha} & \omega \gg \omega_c \end{cases} \tag{8.195}$$

Now the radiated energy spectrum scales as the square of the Fourier amplitude spectrum, or

$$E_R^{\omega \gg \omega_c} \sim \left(\tilde{U}_R^{\omega \gg \omega_c}\right)^2 \sim M_0^{2\left(1-\alpha/3\right)}\omega^{-2\alpha} \tag{8.196}$$

Now our original hypothesis was that

$$E_R^{\omega \gg \omega_c} \sim S \sim M_0^{2/3} \tag{8.197}$$

Relations 8.196 and 8.197 can only be simultaneously true if $\alpha = 2$. Therefore our hypothesis that the high-frequency very near-source ground motion is incoherent noise of constant amplitude that is independent of the size of the slip implies that

$$\tilde{U}_R \sim M_0 \text{sinc}^2\left(\omega/\omega_c\right) \sim \begin{cases} M_0 & \omega \ll \omega_c \\ M_0^{1/3}\omega^{-2} & \omega \gg \omega_c \end{cases} \tag{8.198}$$

Relationship 8.198 is identical to the Brune spectrum (1970), but without the stress drop scaling. That is, the high-frequency radiation is independent of the stress drop. Therefore our hypothesis is identical to assuming Brune's spectral scaling for seismic moment (but not for stress drop).

We can also anticipate the following asymptotic behavior for any ground motion prediction equations. When the distance is large compared to the source dimension, and when the predominant periods of the ground motion are large

compared to the source duration, we expect the ground motion amplitudes to scale with the seismic moment, or

$$\log U_{far\&lowfreq} \sim \log M_0 \sim {}^{3}\!/_{2}\, M \tag{8.199}$$

For very near-source long-periods (e.g. displacement), we expect the peak amplitude to scale with the size of the slip on the nearby fault segment, or

$$\log U_{near\&lowfreq} \sim \log \bar{D} \sim \log M_0^{1/3} \sim {}^{1}\!/_{2}\, M \tag{8.200}$$

Gutenberg-Richter Frequency Magnitude Relation

$$\log\left(N > M\right) = a - bm \tag{8.201}$$

$$\frac{d\left(N > M\right)}{dM} = \frac{d}{dM}\left(10^{a-bM}\right) = -b\ln(10)10^{a-bM} \tag{8.202}$$

$$N'(M - \frac{\Delta M}{2} < M < M + \frac{\Delta M}{2}) = -b\ln(10)10^{a-bM}\,\Delta M \tag{8.203}$$

$$\log\left[N'(M - \frac{\Delta M}{2} < M < M + \frac{\Delta M}{2}) \right] = \log\left[\ln(10)^{-b}\,\Delta M \right] + a - bM \tag{8.204}$$

$$N'(M - \frac{\Delta M}{2} < M < M + \frac{\Delta M}{2}) = 10^{\log\left[\ln(10)^{-b}\Delta M\right] + a - bM} = C''10^{-bM} \tag{8.205}$$

$$M = C - \frac{2}{3}\log M_0 = C' + \log\left(L^2 D\right)^{-2/3} \tag{8.206}$$

If $L \sim D$, then

$$M = C'' + \log L^{-2} \tag{8.207}$$

$$N'(M - \frac{\Delta M}{2} < M < M + \frac{\Delta M}{2}) = C''10^{-b\left(C'' + \log L^2\right)} = \mathcal{C}L^{-2b} \tag{8.208}$$

Now the total rupture area $A_{total}(M,\Delta M)$ for all earthquakes with $M-\frac{\Delta M}{2}<M<M+\frac{\Delta M}{2}$ is

$$A_{total}(M,\Delta M)=N'L^2=\Upsilon L^{2(1-b)} \qquad 8.209$$

That is, if the b-value is 1, then the total rupture area is the same for integrated area of each magnitude. That is the sum of the rupture area of all 2's is the same as the rupture area of all 3's, is the same as all 4's, etc. What this means is that given a b-value of 1, and given that a point has just experienced slip, then it is equally likely that it could have come from any magnitude earthquake. Given that different magnitude earthquakes have different slips, any slip is as likely as any other. This is only true in a logarithmic sense. That is, a fault is equally likely to experience slip between D and C x D, regardless of the value of D and a constant, C (e.g. given, slips between 1 and 2 mm are just as common as slips between 4 and 8 m.

but with a mirror image fault
with plan view of displacement amplitude vs. distance from fault.
cross section of static motion of a dip-slip fault.
Static offsets in a half-space
near-fault particle motions
Aagaard's figures of different rupture velocities
Spectral representations
Stress drop
Steady-state ruptures do not radiate
Stress in cylindrical hole.
Fracture energy vs yield stress
Residual stress
Energy of a slip pulse.
Frictional energy
Chaotic ruptures
Self-organization
Random walks

Appendix A

A Generalized power-law spectral scaling law

The fact that very near-source accelerations (high-frequencies) appear to be incoherent noise with a peak acceleration that is independent of magnitude motivates the following hypothesis.

Assumption 1: Radiated high-frequency energy $E_R^{f \gg f_c}$ scales with the rupture area, $S = LW$, some unknown power of the stress drop, γ, and is independent of the average slip on that rupture surface, or

$$E_R^{f \gg f_c} \sim S \Delta\sigma^{\gamma} \qquad 8.210$$

This assumes that high-frequency motions are characterized by random phase, which means that if we sum two motions, then the energies sum linearly. This means that if we double the rupture area, then we double the radiated high-frequency energy. I have also included a yet unknown static stress drop dependence.

My hypothesis of near-source energies (equation 8.210) and Brune's assumptions are identical if $\gamma = 2$, and if $\sigma_{eff} = \Delta\sigma$. In essence, Brune (1970) also assumed that the near-source high-frequency radiated energy scales linearly with the rupture area (see **Error! Reference source not found.**).

Assumption 2: Assume that the far-field **seismic wave is approximately a non-negative function of duration T_c.** This is equivalent to saying that the motions are approximately the solutions for the far-field S-waves in an elastic whole space. A non-negative function of duration T_c has a flat amplitude spectrum between zero frequency (value is proportional to the integrated area of the function, which is proportional to the Potency) and the corner frequency, $f_c = 1/T_c$. If the S-wave is non negative then its spectrum at periods shorter than T_c is less than the amplitude spectrum for periods longer than T_c. Brune also explicitly made this assumption.

Assumption 3: For simplicity, **assume that the high-frequency spectrum has a power-law high-frequency spectral decay $f^{-\alpha}$**. That is, **assume** that

$$\tilde{U}_R \sim \frac{P}{1 + \left(\frac{f}{f_c} \right)^{\alpha}} \qquad 8.211$$

Where $\tilde{U}_R$ is the Fourier amplitude spectrum of the radiated (far-field terms) S-wave, potency $P = S\bar{D}$, and corner frequency, $f_c = 1/T_c$. Equation 8.211 has asymptotes

$$\tilde{U}_R \sim \begin{cases} P & f \ll f_c \\ P\left(f/f_c\right)^{-\alpha} & f \gg f_c \end{cases} \quad 8.212$$

Brune also used this assumption, but he assumed that $\alpha = 2$ (based on the analytic solution of an instantaneous shear on a half space).

Assumption 4: Now **assume that the duration of the signal is proportional to the dimension of the fault** (that is, assume constant rupture velocity), or that $f_c \sim \frac{1}{\sqrt{S}}$. Again, Brune used this assumption.

Assuption 5: Assume size similarity of the form $P \sim S\bar{D} \sim S^{3/2}\Delta\sigma$**, or alternatively** $LW \sim P^{2/3}\Delta\sigma^{-2/3}$. Brune also made this assumption. Therefore, $f_c \sim P^{-1/3}\Delta\sigma^{1/3}$, and 8.211 becomes

$$\tilde{U}_R(f) \sim \frac{P}{1+\frac{f^{\alpha}}{P^{-\alpha/3}\Delta\sigma^{\alpha/3}}} \quad 8.213$$

Which has asymptotes

$$\tilde{U}_R \sim \begin{cases} P & f \ll f_c \\ P^{1-\alpha/3}\Delta\sigma^{\alpha/3} f^{-\alpha} & f \gg f_c \end{cases} \quad 8.214$$

Now the radiated energy spectrum scales as the square of the Fourier amplitude spectrum, or at high frequencies

$$E_R^{f \gg f_c} \sim \left(\tilde{U}_R^{f \gg f_c}\right)^2 \sim P^{2\left(1-\alpha/3\right)}\Delta\sigma^{2\alpha/3} \quad 8.215$$

Now my original hypothesis 8.210 was that

$$E_R^{f \gg f_c} \sim S\Delta\sigma^{\gamma} \sim P^{2/3}\Delta\sigma^{\gamma-2/3} \quad 8.216$$

The $\Delta\sigma^{-2/3}$ in 8.216 comes from the hypothesis that the radiated high frequencies scale with the rupture area. This hypothesis implies that for the same P, a higher static stress drop has a smaller rupture area and therefore a smaller radiated high-frequency energy.

The scaling of high-frequency radiated energies with P and $\Delta\sigma$ given in the two independent relations 8.215 and 8.216 can be simultaneously true only if $\alpha = 2$ and also if $\gamma = 2$. Assuming $\gamma = 2$ is the same as saying that radiation of high frequency energy per unit of rupture area of scales with the square of the static stress drop. Since energy scales as the square of the wave amplitude, $\beta = 2$ is the same as saying near-source peak acceleration scales linearly with stress drop (this is Brune's assumption). This assumption is essentially the Brune spectral model (with $\Delta\sigma = \sigma_{eff}$). Notice that in this case

$$\tilde{U}_R \sim \frac{P}{1+\dfrac{f^2}{P^{-2/3}\Delta\sigma^{2/3}}} \sim \begin{cases} P & f \ll f_c \\ P^{1/3}\Delta\sigma^{2/3} f^{-2} & f \gg f_c \end{cases} \quad 8.217$$

This relation could have been derived from either one of two sets of assumptions. Assumption set 1: the spectrum is an f^{-2} power law with a maximum amplitude proportional to P and a spectral corner that is inversely proportional to the rupture dimension. Alternatively, assumption set 2: the spectrum is an unknown power law, the high-frequency near-source motions scale linearly with stress drop, and for equal stress drop, the high-frequency radiated energy scales with rupture area (random phase). These two sets of assumptions are equivalent and either set independently leads to the 2nd form of Brune's spectral model.

$$\tilde{U}_R \sim r^3\Delta\sigma \frac{1}{1+\dfrac{f^2}{P^{-2/3}\Delta\sigma^{2/3}}} \sim \frac{P}{1+\dfrac{f^2}{P^{-2/3}\Delta\sigma^{2/3}}} \quad 8.218$$

Which is identical to my earlier statement of Brune's spectral law.

Inconveniently, near-source high-frequencies (pga) seem to be independent of both slip amplitude (see Figure 8.14) and stress drop $\Delta\sigma$ (see Figure 8.11). That is, Figure 8.11 shows that near source pga's are approximately independent of P (for M>6) and $\Delta\sigma$. That is, $\gamma \approx 0$, or rewriting 8.216,

$$E_R^{f \gg f_c} \sim S \sim P^{2/3} \Delta\sigma^{-2/3} \quad 8.219$$

Which means that the high-frequency spectral amplitude scales as

$$\tilde{U}^{f \gg f_c} \sim \sqrt{S} \sim P^{1/3} \Delta\sigma^{-1/3} \quad 8.220$$

The lack of correlation between near-source pga and stress drop seen if Fig. 8.11 seems to suggest that the spectral asymptotes are given by

$$\tilde{U}_R \sim \begin{cases} P & f \ll f_c \\ P^{1/3} f^{-2} \Delta\sigma^{-1/3} & f \gg f_c \end{cases} \quad 8.221$$

Unlike the Brune relationship, we cannot assume a single power law with a single corner to fit the spectrum.

Scale Dependence of the Strength of the Earth's Crust

Thomas H Heaton

Introduction

What is it that earth scientists mean when they speak of the strength of the Earth's crust? Although the concept of strength is well defined for materials that yield uniformly with spatially homogeneous stress, there is little doubt that deformations and stresses are highly heterogeneous in the Earth's crust. I discuss more general definitions of strength that are applicable for materials that have spatially heterogeneous stress when they yield. These new definitions of strength are based on spatial averages of stress, and while they are compatible with a simple intuitive understanding of strength, they lead to the conclusion that materials that are deforming at multiple length scales have strengths that depend on the length scale of the observation. In particular, the strength decreases with increasing length scale.

Strength of a Material

It is natural to think of the strength of a material as the amplitude of stress at which a material begins to yield[2]. In practice, it is common to measure the strength of a material using some test apparatus that applies an increasing external

[2] Since stress is a tensor quantity, it is usually necessary to define strength as an imaginary surface in 6-dimensional stress space that separates elastic stress states from yielding stress states. If the material is isotropic, then this yield surface can be defined in the 3-dimensional space of the amplitudes of the principal stresses.

load to a sample until it begins to yield, F_Y^{EXT}. Figure 1 shows a schematic of a hypothetical test apparatus to measure the strength of a cubic sample of dimension L_0. The strength is determined by calculating a yield stress from the yield force and the appropriate cross sectional area L_0^2. For reasons that will become apparent shortly, I will consider the strength to depend on the scale of the sample and I will define it as

$$\Sigma(L_0) \equiv \left| F_Y^{EXT}(L_0) \Big/ L_0^2 \right|. \tag{8.222}$$

While this is the traditional way that strength is measured for materials, it is not really possible to make this measurement in the Earth's crust; there is no test apparatus large enough to cause the yielding of a sample that is ten's of kilometers in dimension. Of course, one can measure the yield stress of small samples of the crust, but we should be cautious since there may be mechanisms whose deformation physics depend on the scale of the material. Later I will discuss two other definitions of strength, one which is based on root-mean-square statistics, and the other is based on inelastic work.

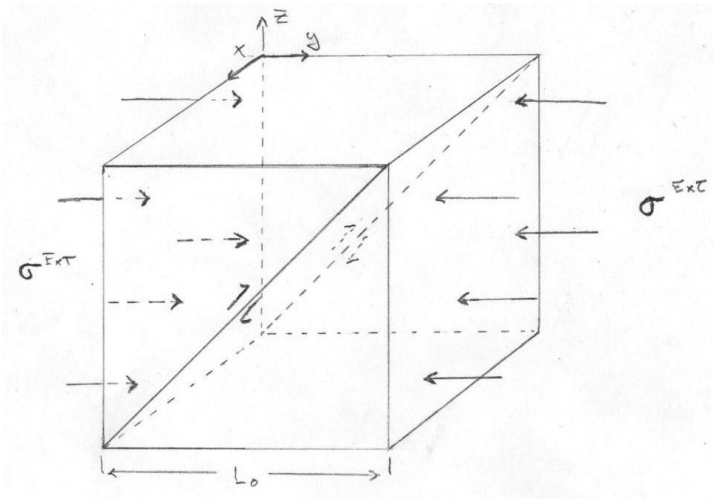


Figure 1. Hypothetical cubic test sample of dimension L_0 that is subjected to a uniform stress. The "strength" of this sample can be viewed as the size of the externally applied stress σ_Y^{Ext} that causes the sample to yield.

Since it is not possible to test the Earth's crust in an apparatus, it is common for earth scientists to estimate the strength of the crust by obtaining estimates of stress amplitudes for a section of the crust that is experiencing inelastic yielding (e.g., earthquakes; see Kanamori and Heaton, 2000). However it is important to recognize that this is a different measurement from the one that is made using a test apparatus in the laboratory. That is, if we return to our conceptual test

apparatus of Figure 1, instead of measuring the external yield force F_Y^{EXT}, we would directly measure stress σ_Y inside of our yielding sample. To keep the discussion relatively simple, let's choose to look at only one component of the shear stress in our medium. That is, I could alternatively define the strength of the sample to be the amplitude of internally measured stress within the yielding sample. Although the discussion up to this point may appear to be trivial, the fundamental issue is how to measure this internal yield stress. That is, stress is always spatially variable inside of any solid material. For example, any poly-crystalline material has enormous stresses at grain boundaries; these stresses are on the order of GPa's. Thermal stresses and dislocations are other sources of spatial variations in stress. Therefore, a more complete description of the stress in our medium would be

$$\sigma(x,y,z) = \sigma^{Ext}(L_0) + \sigma^{Int}(x,y,z), \qquad x,y,z \in [0,L_0) \tag{8.223}$$

where $\sigma^{Ext}(L_0)$ is the spatially uniform stress from the externally applied force and $\sigma^{Int}(x,y,z)$ is the spatial distribution of internal shear stress.[3] By definition, the spatial average of $\sigma^{Int}(\mathbf{x})$ over the entire sample is zero. However, the condition is stronger than that, the average internal stress over any cross section intersecting the x-axis must also be zero, or

$$\overline{\overline{\sigma^{Int}(\mathbf{x})}}^{L_0} \equiv \frac{1}{L_0^2}\int_0^{L_0}\int_0^{L_0} \sigma^{Int}(x,y,z)\,dydz = 0, \tag{8.224}$$

where I use the double bar notation to signify the average over a 2-dimensional surface that is perpendicular to the x-axis. Therefore, by using this decomposition, I can alternatively define the strength as

$$\Sigma(L_0) \equiv \left|\overline{\overline{\sigma(\mathbf{x})}}^{L_0}\right| \equiv \left|\frac{1}{L_0^2}\int_0^{L_0}\int_0^{L_0} \sigma(x,y,z)\,dydz\right| = \left|\sigma_Y^{Ext}(L_0)\right|. \tag{8.225}$$

Now suppose that at a given external load, then the sample could yield at either the scale length of the sample, or it could also yield only in some smaller section of our sample with the dimension, L_1. Therefore, we would conclude that the strength of the smaller cube is

[3] Internal stresses are sometimes called "residual stress", although we will use a more general description of these stresses than is usually considered.

$$\Sigma(L_1) \equiv \left| \overline{\overline{\sigma(\mathbf{x})}}^{L_1} \right| \equiv \left| \frac{1}{L_1^2} \int_0^{L_1} \int_0^{L_1} \sigma(x,y,z)\, dydz \right| = \left| \Sigma(L_0) + \frac{1}{L_1^2} \int_0^{L_1} \int_0^{L_1} \sigma^{Int}(x,y,z)\, dydz \right| \tag{8.226}$$

Therefore, $\Sigma(L_1) \neq \Sigma(L_0)$ unless $\overline{\overline{\sigma^{Int}(x,y,z)}}^{L_1} = 0$. Importantly, earthquakes seem to happen at multiple length scales $L_i, i = 1...n$, which implies that the stress in the crust is such that the crust is at its yield strength at multiple length scales. That is, the strength of the crust is independent of the length scale if, and only if, $\overline{\overline{\sigma^{Int}(x,y,z)}}^{L_i} = 0$ for all length scales L_i. If $i \to \infty$, then for the strength to be independent of scale length, then $\sigma^{Int}(\mathbf{x}) = 0$, which would imply homogeneous stress in the crust, which is inconsistent with the occurrence of earthquakes at multiple length scales. Therefore, if the crust is failing at all length scales, then the strength of the crust, as defined by (8.225), must depend on the scale length.

More generally, since the definitions of strength given by (8.225) and (8.222) are equivalent, this implies that if our laboratory apparatus recreates the same physics as earthquake failures in the crust, then we would also measure changes in the strength of samples as a function of the size of the sample. As we will see, the key to a deeper understanding of this problem is to characterize the internal stress distribution in the crust.

A Statistical Description of Stress in the Crust

In the previous section I defined the strength of a yielding volume as the amplitude of the average stress on a plane that cuts through the volume, that is, equation(8.225). I am particularly interested in how strength Σ varies as a function of the length scale L of the yielding volume. In the case of the Earth's crust, I **make the following key assumption. The crust has evolved into such a state of stress that it can fail at any length scale.** That is, I assume that earthquakes of any size are possible within our hypothetical crust (Bak and others, 1987). In this case the distribution of stress $\sigma(x,y,z)$ tells us the length scale dependence of the strength in the crust. All we need to do is to estimate the amplitude of the spatial average of $\sigma(x,y,z)$ as a function of the length scale.

At this point, it is convenient to assume that $\sigma(x,y,z)$ is a random stationary function of *y* and *z*. The assumption of stationarity means the statistical properties of the stress are invariant with respect to position within the medium. That is, the joint statistical distribution of $\sigma^{Int}(x_1,y_1,z_1),...,\sigma^{Int}(x_n,y_n,z_n)$ is the

same as that of $\sigma^{Int}(x_1+x, y_1+y, z_1+z), ..., \sigma^{Int}(x_n+x, y_n+y, z_n+z)$. Assuming stationarity means that the process σ^{Int} is entirely determined by its covariance function

$$\Gamma(x, y, z, x', y', z') \equiv \mathrm{cov}\left[\sigma^{Int}(x, y, z), \sigma^{Int}(x', y', z')\right]$$
$$\equiv E\left\{\left[\sigma^{Int}(x, y, z) - E\left(\sigma^{Int}(x, y, z)\right)\right]\left[\sigma^{Int}(x', y', z') - E\left(\sigma^{Int}(x', y', z')\right)\right]\right\} \quad (8.227)$$

which is also known under the names of the autocorrelation function or the power spectrum, and where E is taken to mean *expected value*. Note that the stationarity assumption implies that the covariance between two points depends only on the distance between those points. Therefore, we can define $\gamma(\xi - \xi') \equiv \Gamma(\xi, \xi')$

I recognize that stress in the Earth's crust is clearly not stationary with respect to position; for example, there are systematic variations in crustal thickness, material properties, total deformation, and deformation rate. Nevertheless, this assumption is necessary for later analysis in this paper and it may be approximately valid within regions of the crust. Furthermore, if I choose an x-axis that is perpendicular to a fault plane, then I am essentially assuming that the statistical properties of stress on the fault plane are independent of the position on the fault plane.

I can now restate my definition of strength in (8.225) as

$$\Sigma(L) \equiv E\left(\left|\overline{\overline{\sigma(x, y, z)}}^{L}\right|\right) \quad (8.228)$$

Where E is the expected value of the amplitude of the stress averaged over the volume of length scale L. As an alternative, it can make the mathematics simpler if I introduce an alternative definition of strength that I will call the root-mean-square strength $\Sigma_{rms}(L)$ and which I define as

$$\Sigma_{rms}^2(L) \equiv E\left[\left(\overline{\overline{\sigma(x, y, z)}}^{L}\right)^2\right] \quad (8.229)$$

Using rms in the definition ensures that strength is always a positive amplitude, but it has the advantage that it allows me to relate strength to the statistical variance of stress. In particular, I show in Appendix I (see page 316 of Dwass) that

$$\begin{aligned}\Sigma_{rms}^2(L) &\equiv \left[E\left(\overline{\overline{\sigma(x,y,z)}}^{L}\right)\right]^2 + Var\left(\overline{\overline{\sigma(x,y,z)}}^{L}\right) \\ &= \left(\sigma_Y^{Ext}(L_0)\right)^2 + Var\left(\overline{\overline{\sigma^{Int}(x,y,z)}}^{L}\right)\end{aligned} \tag{8.230}$$

This means that I can find the length scale dependence of *rms* strength by finding the dependence of the variance of the spatially averaged stress on the length scale of the spatial averaging.

I can use Parseval's theorem to calculate the variance of a stationary function that is described by its covariance. That is, the integrated energy in the space domain, which is the variance times the length of the function, is identically equal to the integral of the power spectrum at all frequencies. That is, if $\overline{\overline{\tilde{\sigma}^{Int}}}\left(k_x,k_y,k_z\right)$ is the 3-dimensional Fourier transform of the spatially averaged internal stress, then it has a variance given by

$$\begin{aligned}Var\left(\overline{\overline{\sigma^{Int}(x,y,z)}}^{L}\right) &= \frac{1}{L_0^2}\int_0^{L_0}\int_0^{L_0}\left|\overline{\overline{\sigma(x,y,z)^{Int}}}^{L}\right|^2 dydz \\ &= \frac{1}{L_0^2}\int_{L_0^{-1}}^{\infty}\int_{L_0^{-1}}^{\infty}\left|\overline{\overline{\tilde{\sigma}(k_x,k_y,k_z)^{Int}}}^{L}\right|^2 dk_y dk_z\end{aligned} \tag{8.231}$$

If I can characterize how the integral in (8.231) depends on scale length *L*, then I will have solved our problem. At this point, I assume that $\sigma^{Int}(x,y,z)$ is a mean-zero (i.e., $E\left[\sigma^{Int}(x,y,z)=0\right]$) Gaussian stationary process. That is, I assume that the internal stress can be approximated by a spatially filtered version of Gaussian white noise. I am unaware of any physical basis for this assumption, and I use it because it makes the mathematics simpler.

I am particularly interested in the obtaining the statistical properties of $\overline{\overline{\sigma^{Int}(x,y,z)}}^{L}$, which is the spatial average of the stress over a 2-dimensional surfaces of dimension L. Obtaining the spatial average of the stress is equivalent to 2-dimensional convolution of the stress with some 2-dimensional function $g\left(y/L, z/L\right)L^{-2}$, where *g* has circular symmetry with respect to *y* and *z*, and *g* has unit characteristic length and integrated area (e.g., a Gaussian, a box function, etc.). That is, $g\left(y/L, z/L\right)L^{-2}$ has a characteristic length of L and an integrated area of unity. Therefore

$$\begin{aligned}\overline{\overline{\sigma^{Int}(x,y,z)}}^{L} &= L^{-2}\sigma^{Int}(x,y,z) * g\left(y/L, z/L\right) \\ &= L^{-2}\sigma^{Int}(x,y,z) * \left[g'\left(y/L\right)\delta(z)\right] * \left[g'\left(z/L\right)\delta(y)\right]\end{aligned} \tag{8.232}$$

Where g' is the 1-dimensional version of *g*. Therefore I can conclude that

$$\overline{\overline{\tilde{\sigma}^{Int}\left(k_x,k_y,k_z\right)}}^{L} = L^{-2}\tilde{\sigma}^{Int}\left(k_x,k_y,k_z\right)\tilde{g}'\left(Lk_y\right)\tilde{g}'\left(Lk_z\right) \tag{8.233}$$

I can substitute (8.233) into (8.231) to conclude that

$$Var\left(\overline{\overline{\sigma^{Int}(x,y,z)}}^{L}\right) = \frac{1}{L_0^2}\int_{L_0^{-1}}^{\infty}\int_{L_0^{-1}}^{\infty}\left|\tilde{\sigma}^{Int}\left(k_x,k_y,k_z\right)\tilde{g}'\left(Lk_y\right)\tilde{g}'\left(Lk_z\right)\right|^2 dk_y dk_z \tag{8.234}$$

At this point, it is helpful to be more specific about g'. If I were to assume g' to be a box rectangle function in the space domain (corresponding to an un-weighted running mean), then it's transform would be a sinc function in the wavenumber domain. If, on the other hand, I assume that g' is a sinc function in the space domain, then its transform is a rectangle function in the wavenumber domain and the solution becomes particularly simple. That is if

$$g'(y) \equiv \frac{\sin(y)}{y} \tag{8.235}$$

Then

$$\tilde{g}'\left(k_y\right) = \Pi\left(k_y\right) \equiv \begin{cases}1, & -\frac{1}{2} < k_y < \frac{1}{2} \\ 0, & \text{otherwise}\end{cases} \tag{8.236}$$

Substituting (8.236) into (8.234), I obtain

$$\begin{aligned}Var\left(\overline{\overline{\sigma^{Int}(x,y,z)}}^{L}\right) &= \frac{1}{L_0^2}\int_{L_0^{-1}}^{\infty}\int_{L_0^{-1}}^{\infty}\left|\tilde{\sigma}^{Int}\left(k_x,k_y,k_z\right)\Pi\left(Lk_y\right)\Pi\left(Lk_z\right)\right|^2 dk_y dk_z \\ &= \frac{1}{L_0^2}\int_{L_0^{-1}}^{L^{-1}}\int_{L_0^{-1}}^{L^{-1}}\left|\tilde{\sigma}^{Int}\left(k_x,k_y,k_z\right)\right|^2 dk_y dk_z\end{aligned} \tag{8.237}$$

Substituting (8.237) into (8.230) gives the functional dependence of *rms* strength on length scale.

$$\Sigma_{rms}^2(L) = \left(\sigma_Y^{Ext}(L_0)\right)^2 + \frac{1}{L_0^2}\int_{L_0^{-1}}^{L^{-1}}\int_{L_0^{-1}}^{L^{-1}} \left|\tilde{\sigma}^{Int}\left(k_x, k_y, k_z\right)\right|^2 dk_y dk_z \tag{8.238}$$

This relationship tells us how to determine the strength of the sample at smaller scales L than the scale L_0 of the entire sample. Conversely, if I already knew the rms strength of the material at a smaller scale L, then I could determine its value at the larger scale L_0 simply by rearranging (8.238) as

$$\left(\sigma_Y^{Ext}(L_0)\right)^2 = \Sigma_{rms}^2(L) - \frac{1}{L_0^2}\int_{L_0^{-1}}^{L^{-1}}\int_{L_0^{-1}}^{L^{-1}} \left|\tilde{\sigma}^{Int}\left(k_x, k_y, k_z\right)\right|^2 dk_y dk_z \tag{8.239}$$

That is, if I could make in situ measurements of stress in the crust at the scale length of meters, then I could infer the *rms* strength at the length scale of kilometers by using (8.239), provided of course that I can characterize $\left|\tilde{\sigma}^{Int}\left(k_x, k_y, k_z\right)\right|$.

The nature of stress heterogeneity in the crust

.As mentioned in the introduction, we can view the question of strength of the crust as being equivalent to characterizing $\left|\tilde{\sigma}\left(k_x, k_y, k_z\right)\right|$. That is, through millennia of deformation, the crust has evolved into a state of stress whose statistical characteristics are determined by its spatial power spectrum. In a very fundamental way, knowing this power spectrum tells us the strength of the crust. However, a power spectral definition of strength is definitely not as simple as the more traditional notion of measuring a force necessary to yield a sample (i.e., equation (8.222)). We saw that these two notions of strength can be connected through equation (8.238).

What is the nature of $\left|\tilde{\sigma}\left(k_x, k_y, k_z\right)\right|$? To keep things as simple as possible, we will assume that the power spectrum is isotropic in space. Although we have explicitly acknowledged the 3-dimensionality of the crust, we have adopted a definition of strength that is based on the failure of 2-dimensional surfaces. In some ways it is attractive to drop one of the dimensions from this discussion. However, we will later discuss seismicity data that is inherently 3-dimensional. If the stress power spectrum is isotropic in three dimensions, then we can consider that we can write it as $\left|\tilde{\sigma}\left(k_r\right)\right|$, where $k_r = \sqrt{k_x^2 + k_y^2 + k_z^2}$.

If the stress in the crust was a simply a constant (which it's clearly not), then the strength is also a constant, independent of the scale. In this case, $\left|\tilde{\sigma}^{Int}\left(k_r\right)\right| = 0$

and the power spectrum of stress is simply a 3-dimensional impulse function at the origin, or $\left|\tilde{\sigma}\left(k_r\right)\right| = \Sigma\left(L_0\right)\delta\left(k_r\right)$.

A more interesting case is to investigate the possibility that the power spectrum can be approximated as a power law, or assume that $\left|\tilde{\sigma}\left(k_r\right)\right| = Ck_r^{-\beta}$, where β is positive and where the outer scale of the sample is infinite. In this case, the strength approaches zero as the scale becomes infinite and (8.238) can be written as

$$\Sigma_{rms}^2\left(L\right) = \frac{C}{L_0^2}\int_0^{L^{-1}}\int_0^{L^{-1}}\left(k_y^2 + k_z^2\right)^{-\beta} dk_y dk_z \tag{8.240}$$

This double integral over a rectangular box in Cartesian coordinates can be transformed to a double integral over a circular area without any loss of generality. That is, we can rewrite (8.240) as

$$\begin{aligned}\Sigma_{rms}^2\left(L\right) &= C'\int_0^{L^{-1}}\int_0^{2\pi} k_r^{-2\beta}\left(k_r d\theta\right)dk_r \\ &= 2\pi C'\int_0^{L^{-1}} k_r^{1-2\beta} dk_r \\ &= \frac{2\pi C'}{\left(1-2\beta\right)}\frac{1}{L^{2(1-\beta)}}\end{aligned} \tag{8.241}$$

Or

$$\Sigma_{rms}\left(L\right) = \frac{C''}{\sqrt{1-2\beta}}\frac{1}{L^{1-\beta}},\ 0 \le \beta < 1 \tag{8.242}$$

The remarkably simple result of (8.242) only applies if the material is in a critical state at all length scales. Or, in other words, the material could experience an event of any size. The spectral decay parameter β must be smaller than 1, since if it is larger than 1, it would lead to the conclusion that long wavelength variations in stress are large compared to small ones. Such a stress state is incompatible with the concept that failure can occur at any length scale.

In many ways, either uniform stress, or random stress with a power law spectral content, are the stress states requiring the fewest parameters to describe them. Both of these idealized stress states have no inherent length scale. Of course, the stress in the Earth is determined by failure processes that introduce scale lengths that result in power spectra that are not easily parameterized. However, we will discuss evidence that suggests that earthquake failure processes may be consistent with power-law spectral behavior over a relatively broad range of scale lengths.

Numerical simulations using fractal tensors

Up to this point we have discussed stress as if it were a scalar quantity, which is clearly not the case. Smith and Heaton (2006) demonstrate a procedure to generate spatially varying stochastic stress tensors that 1) are isotropic (no preferred orientations), 2) has scalar invariants that are power law for any section through the material, and 3) has orientations of principal coordinate frames that vary stochastically in space with correlation at all length scales (i.e., the orientations are fractal). The procedure consists of parameterizing the stress at each point with three scalar invariants and with the orientation of the principal coordinate frame (another 3 degrees of freedom) for each point. For the purpose of this calculation, we are interested in spatial averages of shear stress, so we can assume that the scalar invariant corresponding to pressure is spatially constant (that is, it drops out of the problem).

We start with a cubic grid of discrete points. We then assign zero-mean Gaussian random numbers corresponding to two principal stresses. The third principal stress is then chosen so that the sum of the principal stresses is zero. These stress invariants are then spatially filtered in three dimensions using a wavenumber filter given by $\tilde{f}_{ijk} = n^{-3}\left(i^2 + j^2 + k^2\right)^{-\frac{1}{2}\beta}$, $i, j, k = 1...n$, where n is the number of points along a linear axis.

Kanamori, H., and T. Heaton, 2000, Microscopic and macroscopic physics of earthquakes, contained in Geocomplexity and the Physics of Earthquakes, Editors J. Rundle, D. Turcotte, and W. Klein, Geophysical Monograph 20, Published by the American Geophysical Union, D.C., 127-141.

BIBLIOGRAPHY

Doe, John B. *Conceptual Planning: A Guide to a Better Planet*, 3d ed. Reading, MA: SmithJones, 1996.

. City: University Press, 1998.

INDEX