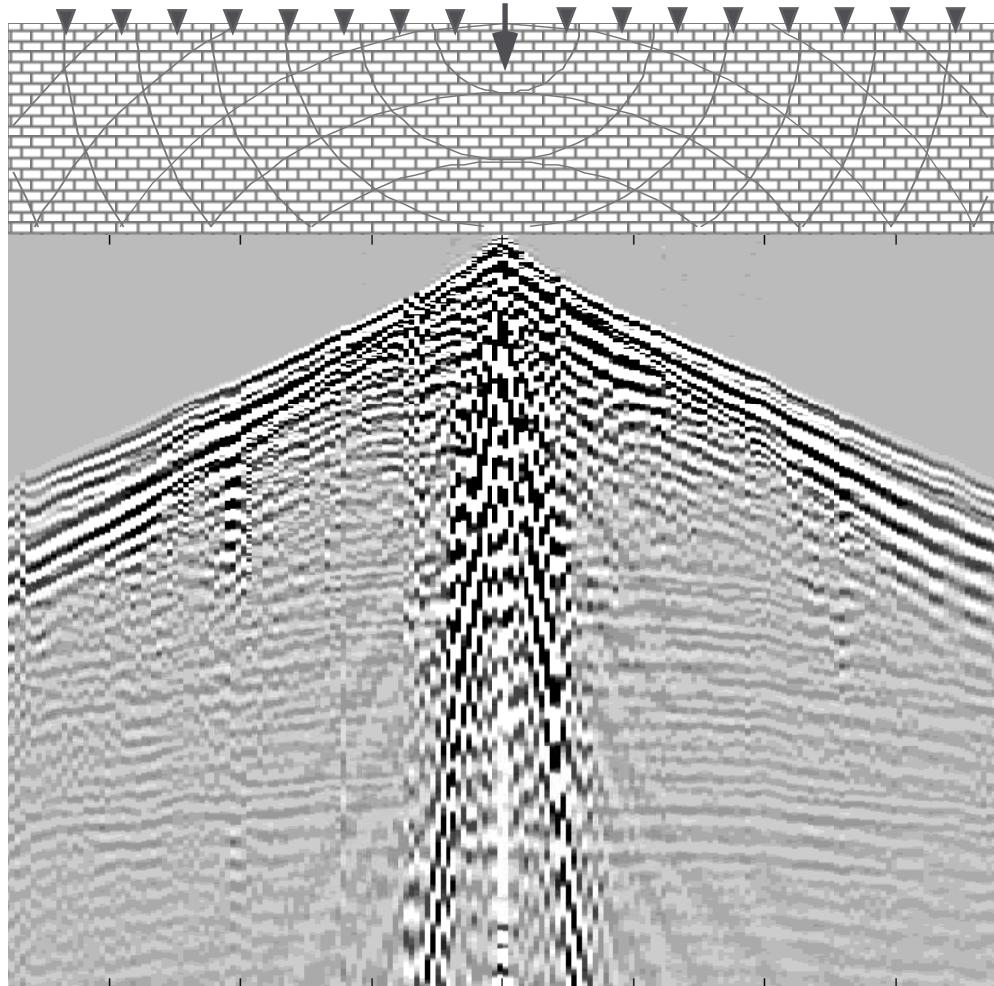


Methods of Seismic Data Processing

Gary F. Margrave

Geophysics 557/657
Course Lecture Notes, Winter 2006



The Department of Geology and Geophysics
The University of Calgary

**Methods of Seismic Data Processing
Geophysics 557/657
Course Lecture Notes
420 Pages
Winter 2005**

by
G.F. Margrave, Associate Professor, P.Geoph.
The CREWES Project
Department of Geology and Geophysics
The University of Calgary
Calgary, Alberta, T2N-1N4
403-220-4604
gary@geo.ucalgary.ca

Table of Contents

Section Title	Page Number
Chapter 1: Synthetic Seismograms	30 pages
The Big Picture	1-2
Elastic Waves	1-7
Well Logs	1-9
Gardner's Rule	1-11
The Wave Equation	1-14
Traveling Waveforms	1-17
Normal Incidence Reflection Coefficients	1-19
Synthetic Seismogram Algorithms	1-23
Synthetic Seismogram Examples	1-28
P-S Synthetic Seismogram Construction	1-30
Chapter 2: Signal Processing Concepts	76 pages
Convolution	2-2
Convolution by Replacement	2-5
Convolution as a Weighted Sum	2-6
Matrix Multiplication by Rows	2-7
Matrix Multiplication by Columns	2-8
Convolution as a Matrix Operation	2-9
Fourier Transforms and Convolution	2-13
Fourier Analysis and Synthesis	2-19
Fourier Analysis Example	2-21
Fourier Transform Pairs	2-23
The Dirac Delta Function	2-25
The Convolution Theorem	2-27
Sampling	2-29
The Discrete Fourier Transform	2-33
The Fast Fourier Transform	2-37
Filtering	2-38
The Z Transform	2-39
Crosscorrelation	2-44
Autocorrelations	2-46
Spectral Estimation	2-48
Wavelength Components	2-53
Apparent Velocity (or phase velocity)	2-56
The 2-D F-K Transform	2-58
F-K Transform Pairs	2-62
-p Transforms	2-63
Properties and uses of the -p Transform	2-68
Inverse -p Transforms	2-71
Least Squares -p and f-k Transforms	2-74
Chapter 3: Amplitude Effects	32 pages
Seismic Wave Attenuation	3-2
True Amplitude Processing	3-8
Automatic Gain Correction (AGC)	3-9
Trace Equalization (TE) or Trace Balancing	3-13
Constant Q Effects	3-14
Minimum Phase Intuitively	3-18
Minimum Phase and the Hilbert Transform	3-21

Minimum Phase and Velocity Dispersion	3-25
Array Theory	3-27
Chapter 4: The Convolutional Model and Deconvolution	61 pages
Bandlimited Reflectivity	4-2
The Convolutional Model	4-4
Frequency Domain Spiking Deconvolution	4-12
Finding a Wavelet's Inverse	4-20
Wiener Spiking Deconvolution	4-23
Prediction and Prediction Error Filters	4-28
Gapped Predictive Deconvolution	4-32
Burg (Maximum Entropy) Deconvolution	4-36
The Minimum Phase Equivalent Wavelet	4-39
Vibroseis Deconvolution	4-41
Deconvolution Pitfalls	4-47
Reflectivity Color	4-55
Q Example	4-58
Chapter 5: Surface Consistent Methods	29 pages
Seismic Line Coordinates	5-2
A Surface Consistent Convolutional Model	5-5
Surface Consistent Methods	5-9
Statics and Datums	5-12
Statics with Uphole Times	5-17
Surface Consistent Residual Statics	5-19
Refraction Statics	5-25
Chapter 6: Velocity Definitions and Simple Raytracing	26 pages
Velocity in Theory and Practice	6-2
Instantaneous Velocity	6-3
Vertical Traveltime	6-4
Vins as a Function of Vertical Traveltime	6-6
Average Velocity	6-8
Mean Velocity	6-10
RMS Velocity	6-11
Interval Velocity	6-13
Snell's Law	6-18
Raytracing in a $v(z)$ Medium	6-20
Measurement of the Ray Parameter	6-24
Raypaths when $v = v_0 + cz$	6-25
Chapter 7: Normal Moveout and Stack	38 pages
Normal Moveout	7-2
Stacking Velocity	7-5
Normal Moveout and Reflector Dip	7-6
NMO for a $V(z)$ Medium	7-10
Dix Equation Moveout	7-13
Normal Moveout Removal	7-15
Extension of NMO and Dip to $V(z)$	7-17
NMO for Multiple Reflections	7-22
CMP Stacking	7-27
Post Stack Considerations	7-30
ZOS: A Model for the CMP Stack	7-34
Fresnel Zones	7-36

Chapter 8: Migration Concepts	52 pages
Raytrace Migration of Normal Incidence Seismograms	8-2
Time and Depth Migrations, A First Look	8-5
Elementary Constant Velocity Migration	8-6
Huygen's Principle and Point Diffractors	8-9
The Exploding Reflector Model	8-14
F-K Migration, Geometric Approach	8-20
F-K Migration, Mathematics	8-25
F-K Wavefield Extrapolation	8-27
Recursive F-K Wavefield Extrapolation for $v = v(z)$	8-31
The Extrapolation Operator	8-33
Vertical Time-Depth Conversions	8-36
Time and Depth Migration in Depth	8-37
Kirchhoff Migration	8-40
Finite Difference Concepts	8-43
Finite Difference Migration	8-46
Chapter 9: The Third Dimension	32 pages
Impulse Responses	9-2
Wave Propagation	9-6
Fresnel Zones	9-7
Wavelength Components	9-10
Apparent Velocity (or phase velocity)	9-13
The F-K Transform	9-15
F-K Transform Pairs	9-19
F-L transform Computation	9-20
3-D Migration by Double 2-D	9-24
Exploitable Symmetries	9-27
Mapping Strategies	9-29
Time migration of traveltimes maps	9-31
Chapter 10: Seismic Resolution Limits	35 pages
Resolution Concepts	10-2
Linear $v(z)$ resolution theoru for zero offset seismic data	10-18
Chapter 11: Study Guide	9 pages
Geophysics 557 Final Exam Study Guide	11-2
Exam Sampler	11-7

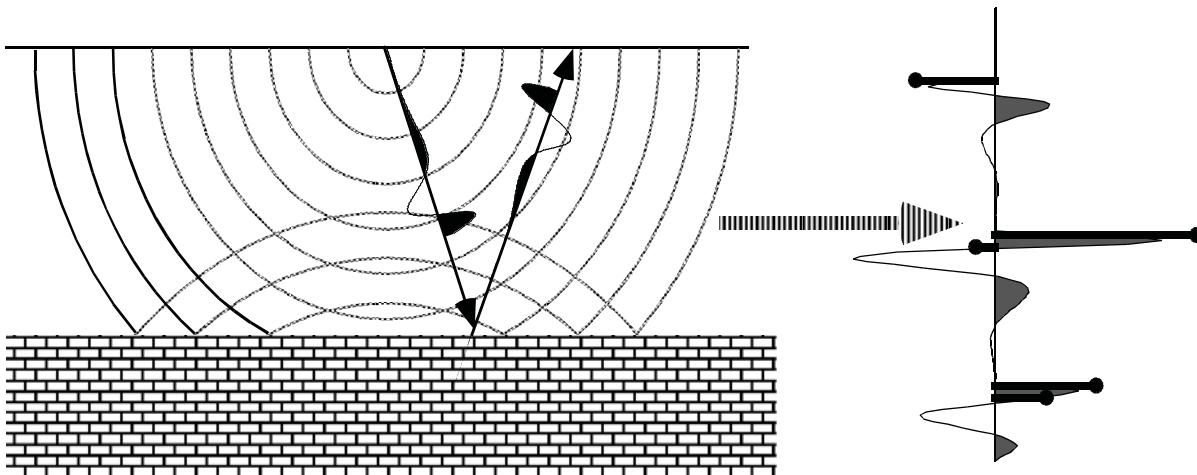
Methods of Seismic Data Processing

**Lecture Notes
Geophysics 557**

**Chapter 1
Synthetic Seismograms**

The Big Picture

The simplest model of seismic data is that of a wavelet convolved with reflectivity. The picture is simple and appealing. A compact pulse of sound is sent down into the earth and scaled copies of it are reflected from the major formation boundaries.



These echoes are recorded over the extent of the seismic experiment and analyzed. Since each echo is a scaled copy of the source waveform, simple comparison makes it is easy to deduce the relative strength of the different reflecting horizons. The estimated set of reflection coefficients is called the reflectivity function of the earth beneath the survey.

Its a nice concept but is it valid? How can it be defended from basic physical principles? What assumptions (there are always assumptions in physics) are required? When are they justified and when are they not?

The Big Picture

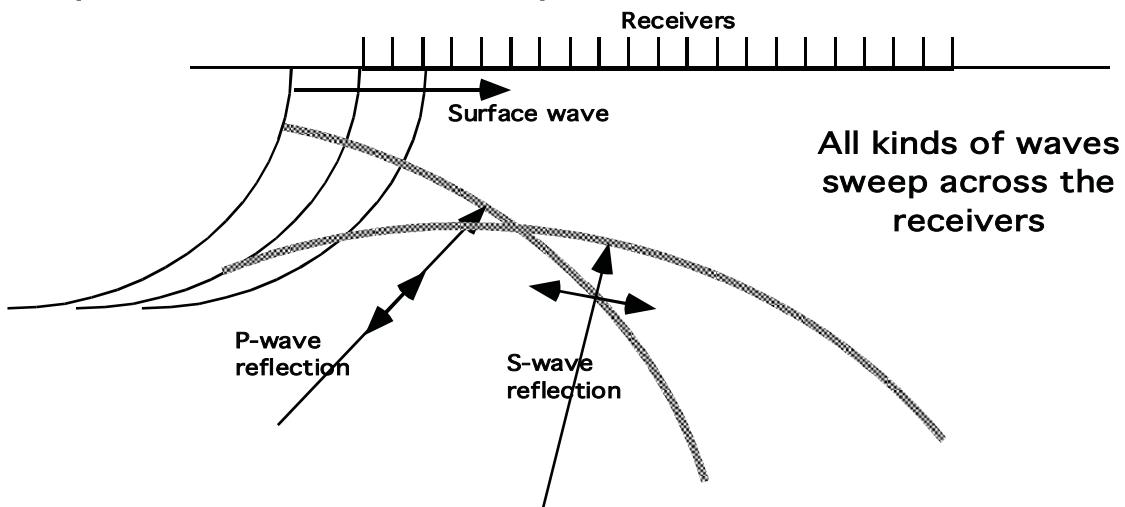
Once we start to think about the idea, we can immediately come up with a lot of questions such as:

- How can we proceed if we don't know the source waveform?
- What if several echos are very closely spaced?
- How can we tell where the echo came from?
- Isn't there attenuation of seismic energy and doesn't this change the source waveform?
- What is convolution anyway? (And why should I care?)
- What about multiple bounce echos? Don't they confuse things?
- If things are so simple, how come seismic processing is so complicated? Maybe those processors are just fooling us ...
- How can I decide how much source energy I need?
- What are the limits of the detail that can be resolved?
- What are the tradeoffs with Vibroseis and dynamite?
- What is reflectivity anyway? (And why should I care?)
- What's this band-limited stuff?
- Why can't I just trust the seismic processor to take care of these messy details?

I'm sure that you can think of more questions as well. All of these questions have their relevance and I hope to address many of them in this course. At the end, you should have a good understanding of the strengths and weaknesses of the convolutional model and this should help you form a healthy, sceptical view of final seismic images.

The Big Picture

Seismic data processing is typically divided into many steps though the reality is that the seismic reflection process does not cleanly separate into discrete packages. We have a source which sends out a complicated, largely unknown waveform which expands, attenuates, reflects, transmits, changes modes, and generally scatters about while a set of receivers placidly records whatever comes their way. And generally what hits the recorders is far more complicated than the simple direct echos that we want:

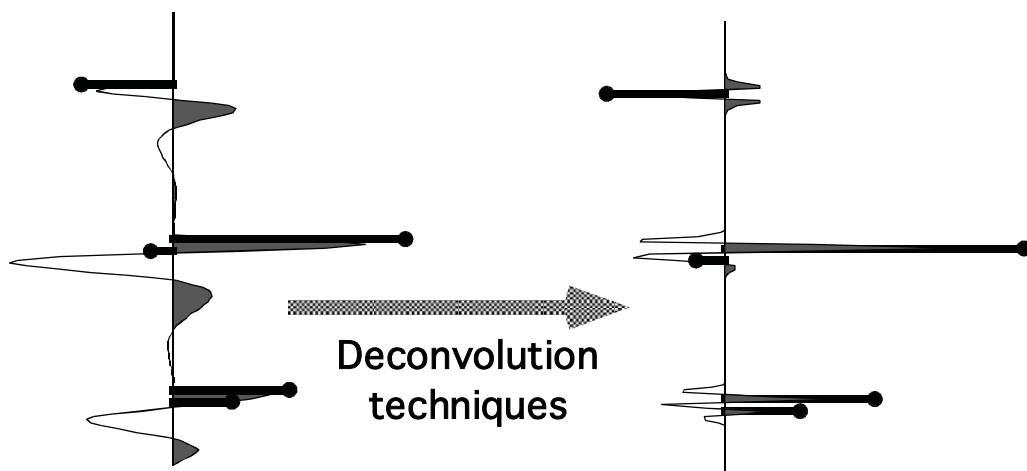


God would not process seismic data the way we do. (I've received a revelation on that point. ;-}) Instead, He would back the waves down into the earth undoing all physical effects at the point where they occurred. We are prevented from doing this largely because of ignorance of the subsurface structure. That is, in order to undo the physical effects of wave propagation, we require knowledge of the subsurface properties that control those effects. Unfortunately, those are the very properties which we hope to discover with the seismic experiment in the first place. Problems of this sort are common in geophysics and are called "inverse problems".

The Big Picture

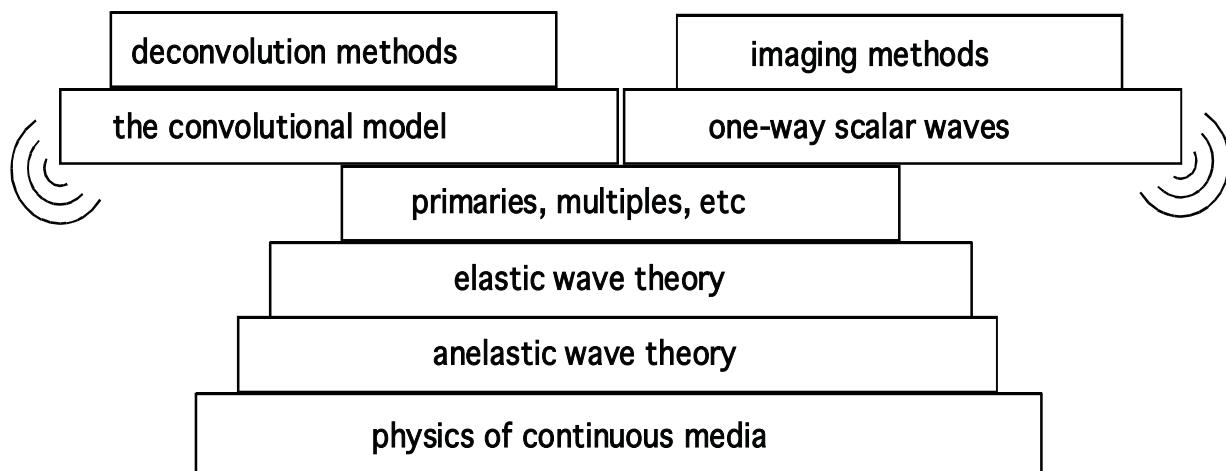
So, faced with the need to find a solution in spite of almost total ignorance, we subdivide, compartmentalize, assume, and approximate until we reach a restatement of the problem which is so vastly simplified that we can actually solve it. An example of such a tremendous oversimplification is the "convolutional model" of the seismic trace which is of central importance to deconvolution theory.

Continuing with sweeping generalities, we can group most physically based seismic processes into one of two groups: imaging processes and deconvolution processes. Imaging processes attempt to determine the correct spatial position of the echos and are typified by nmo removal, cmp stacking, and migration. Deconvolution processes attempt to remove the illuminating waveform and maximize the resolution of the seismic image. Examples are gain recovery, statistical deconvolution, inverse Q filtering, and wavelet processing.



The Big Picture

In order to understand the implications of our simplified theories, it is important to understand as much as possible about the more realistic physics that we are approximating. Therefore, in addition to studying mathematical simplifications such as the convolutional model, we will not hesitate to examine of the most important physical mechanisms involved in seismic wave propagation.



Elastic Waves

The simplest elastic material requires 2 fundamental constants to describe the relation between stress and strain known as Hooke's law:

$$\begin{aligned}\sigma_{ii} &= \lambda\Delta + 2\mu\varepsilon_{ii}, \quad i=x,y,z & \Delta &= \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} \\ \sigma_{ij} &= \mu\varepsilon_{ij}, \quad i=x,y,z, \quad i \neq j & \text{(Sherriff and Geldart,} \\ & & \text{Exploration Seismology, 1981)}\end{aligned}$$

Here σ_{ij} denotes the components of the stress tensor and ε_{ij} the components of the strain tensor. λ and μ are called the Lame constants and μ is also often known as the shear modulus. μ is zero for a fluid. Other constants are often also referenced such as Young's modulus, E , Poisson's ratio, σ , and the bulk modulus, k . These constants are all related in various ways and any two suffice to describe the elastic material.

$$E = \frac{\mu(3\lambda+2\mu)}{\lambda+\mu} \quad \sigma = \frac{\lambda}{2(\lambda+\mu)} \quad k = \frac{3\lambda+2\mu}{3}$$

The description of elastic wave in such a medium, requires the application of Newton's second law ($f=ma$). This leads to the incorporation of the density, ρ , as a necessary constant in the role of "mass" in Newton's second law. Thus, analysis of elastic waves in the most simple elastic solid (homogeneous and isotropic), requires three parameters: any two of: λ , μ , E , σ , and k , plus the density, ρ .

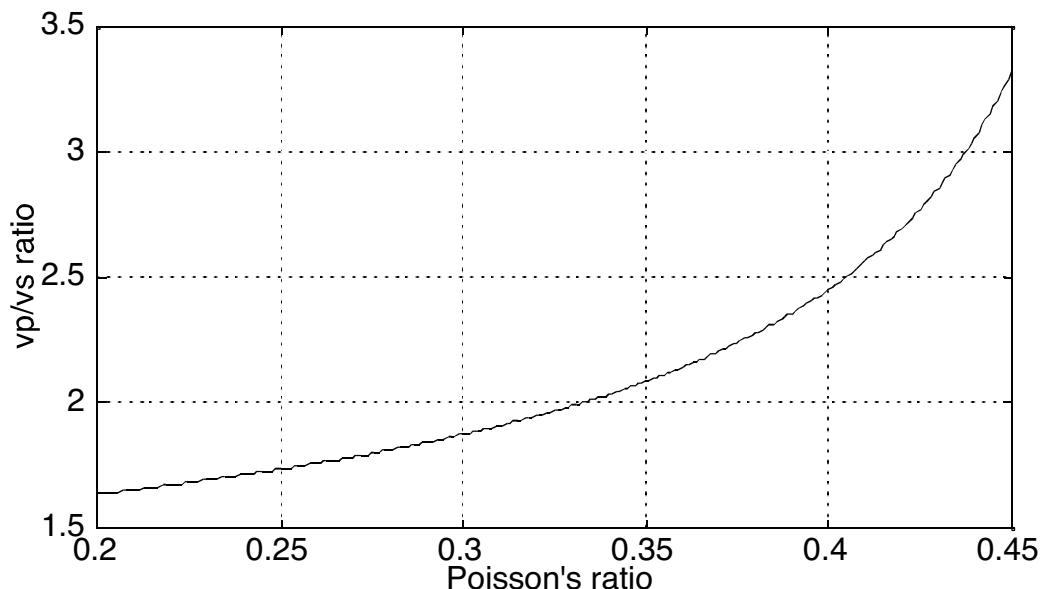
Elastic Waves

It is well established in theory^{1,2,3} that a homogeneous, isotropic elastic solid supports two distinct types of body waves: compressional and shear. Compressional or P waves are characterized by particle motion parallel to the direction of wave propagation. Shear or S waves have particle motion transverse to the direction of wave propagation. P and S waves have velocities of propagation given by:

$$\alpha = \sqrt{\frac{(\lambda+2\mu)}{\rho}} \quad \beta = \sqrt{\frac{\mu}{\rho}}$$

We may choose to regard α and β as fundamental constants (together with ρ). Some relationships are:

$$\lambda = \rho(\alpha^2 - 2\beta^2) \quad \mu = \rho\beta^2 \quad \sigma = \frac{\alpha^2 - 2\beta^2}{2(\alpha^2 - \beta^2)} \quad \frac{\alpha}{\beta} = \sqrt{\frac{2(1 - \sigma)}{1 - 2\sigma}}$$



1: Waters, Reflection Seismology, 1987

2: Sherriff and Geldart, Exploration Seismology, 1982

3: Aki and Richards, Quantitative Seismology Theory and Methods,
1980,

Well Logs

Well logging is a technology designed to make geophysical measurements in a bore hole. Well logs are the most common way to get information about the elastic parameters of rocks which are needed for making synthetic seismograms. Three very common logs, which are of interest to us, are

SON ... P-wave interval transit time

SSON ... S-wave interval transit time

RHOB ... density

The interval transit time logs are usually provided in units of microseconds/lu (lu= meters or feet). Thus, the P and S wave velocities are found as:

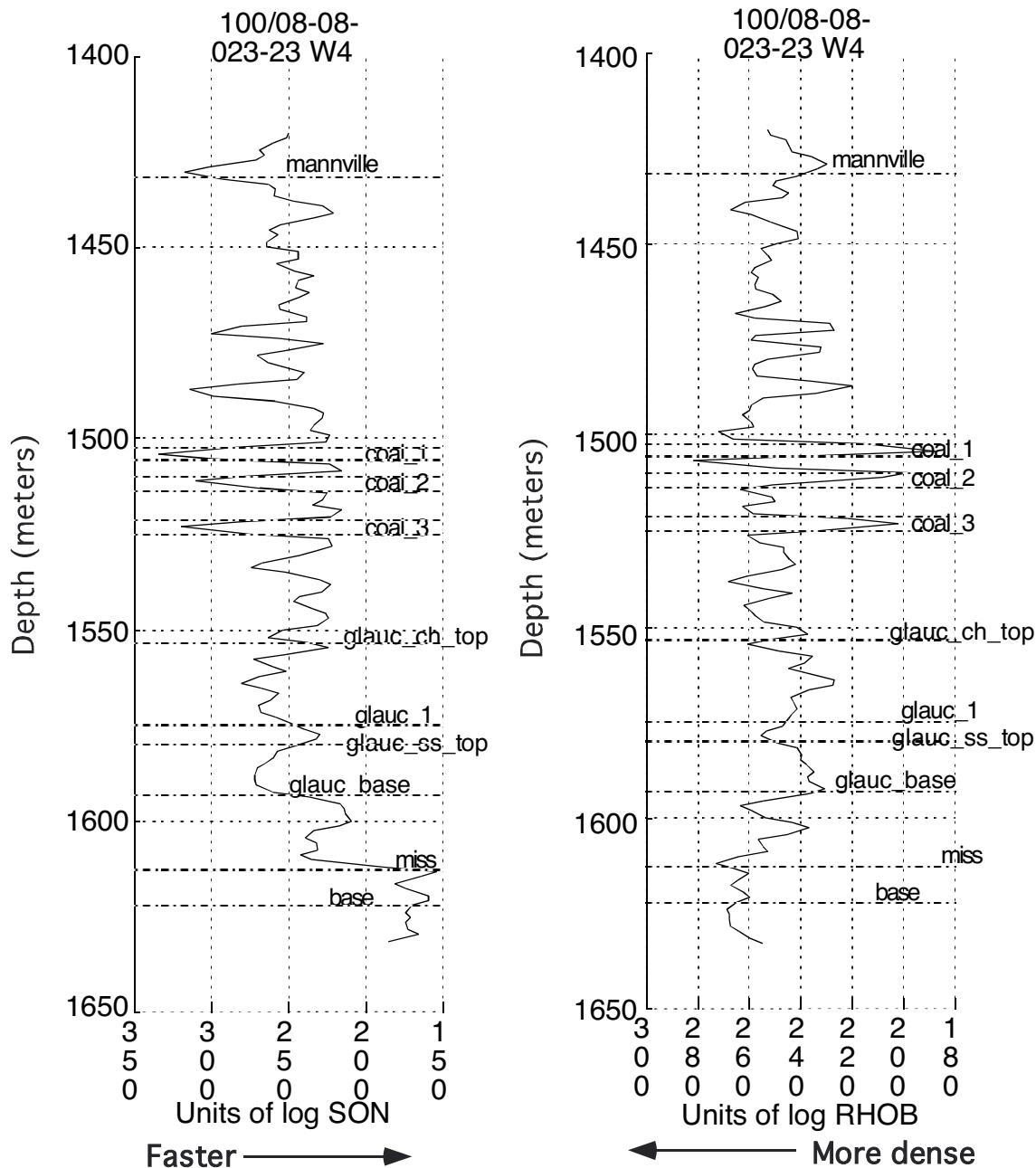
$$\alpha = \frac{10^6}{\text{son}} \quad \beta = \frac{10^6}{\text{sson}}$$

Units for density logs can vary. Be careful to work with consistent units.

Digital well logs are usually packaged in ascii flat files in either GMA or LAS format. The LAS format is more modern and flexible and is to be preferred.

Well Logs

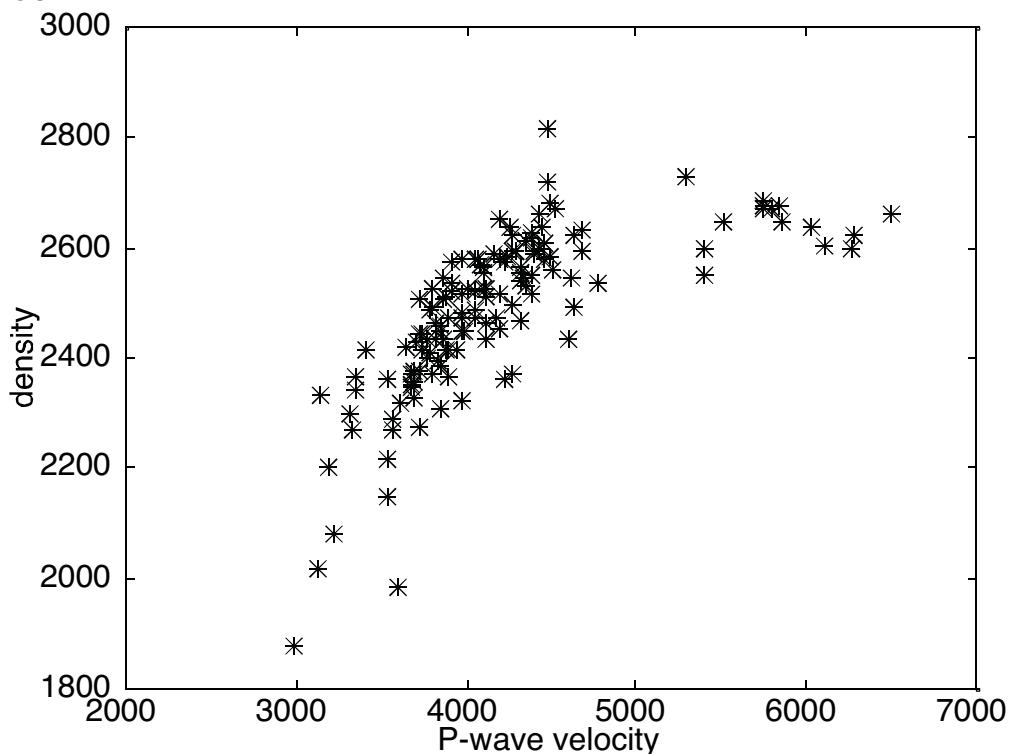
Here are some example logs from 8-8, an oil well in the Blackfoot field



Why do these logs appear to have a negative correlation?

Gardner's Rule

Well logs are often inadequate, incomplete, or missing. One common example of this comes from the fact that sonic logs (SON) are run much more frequently than density logs. Thus we are often faced with the need to create a seismogram without density information. Gardner et al. (1), followed the reasonable approach of seeking an empirical relationship between P-wave velocity and density. Below is a crossplot of ρ and v_p for Blackfoot 8-8 which indicates a reasonable correlation exists:



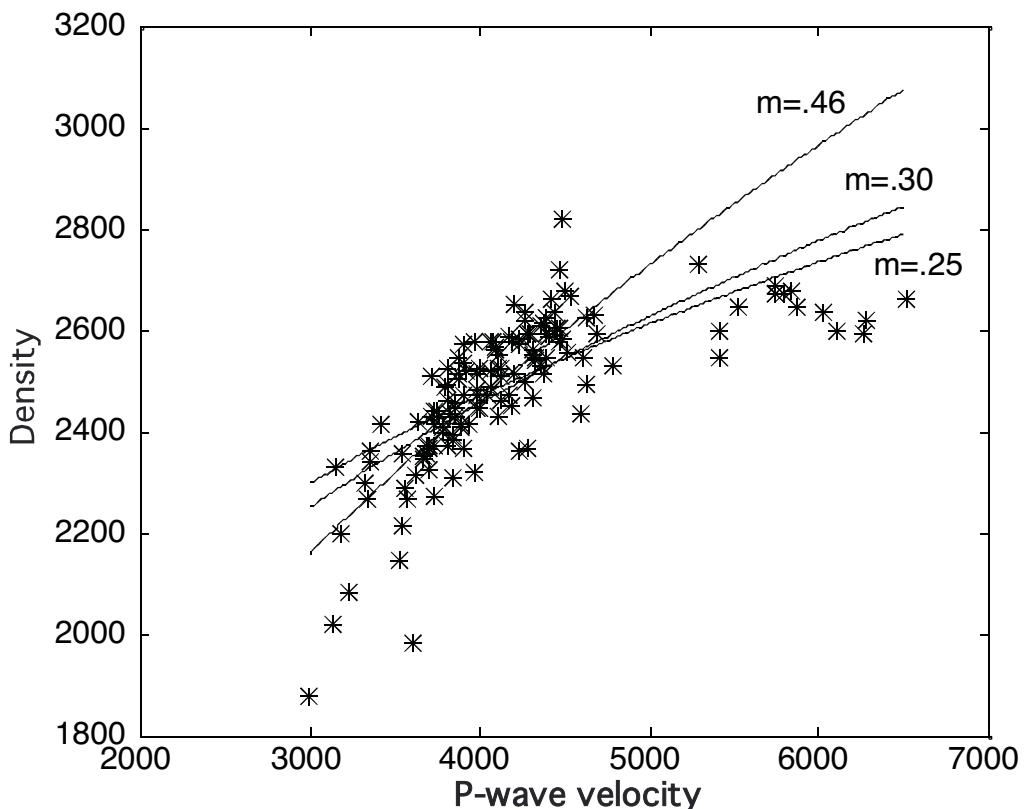
¹ Gardner, G.H.F., Gardner, L. W., and Gregory, A.R., 1974, Formation velocity and density - the diagnostic basis for stratigraphic traps, Geophysics, 39, 770-780

Gardner's Rule

Gardner et al. sought and found a relationship of the form:

$$\rho = a\alpha^m$$

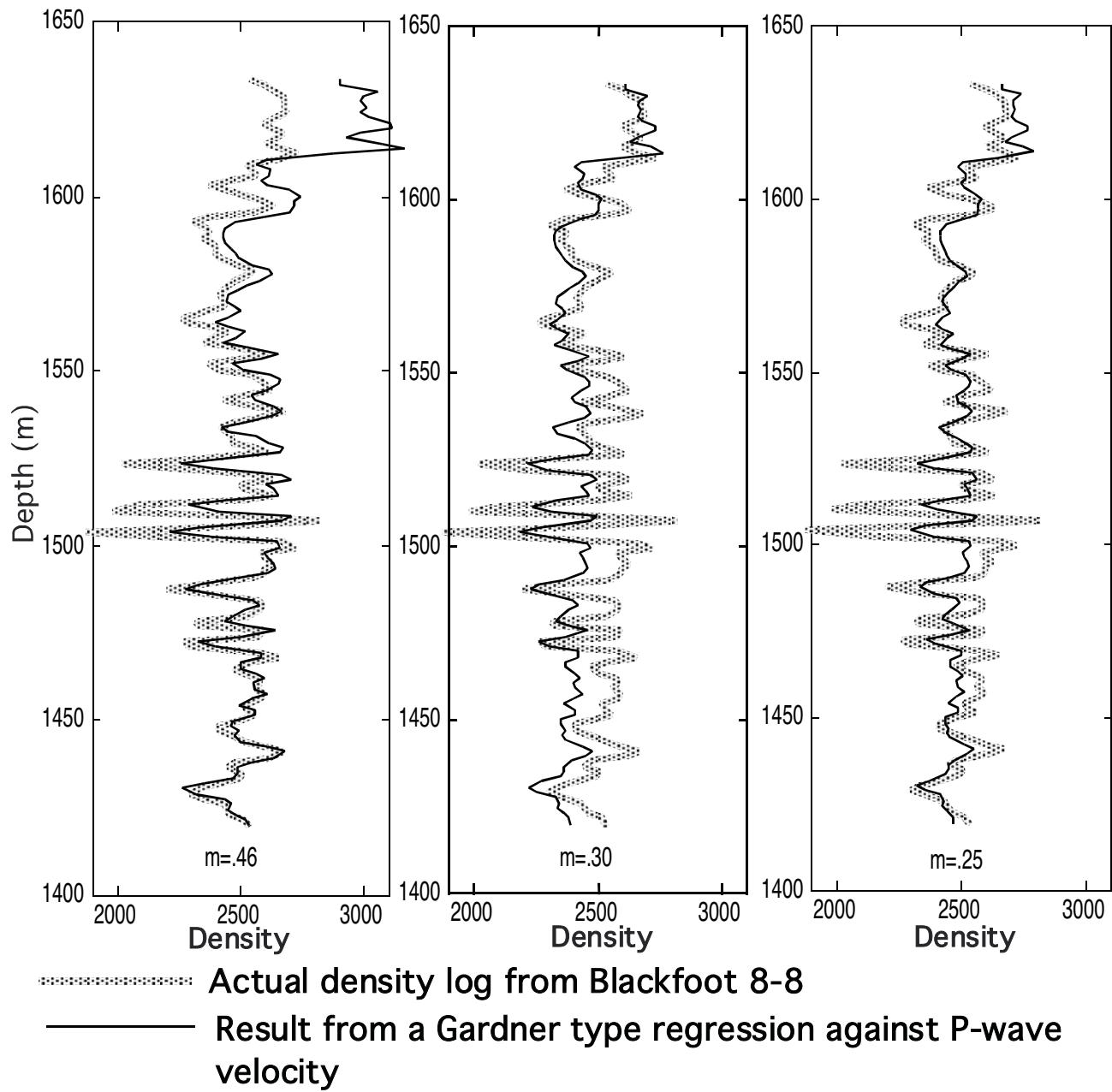
The constants a and m can be determined from fitting a straight line to a plot of $\log(\rho)$ versus $\log(\alpha)$. Below are the results of several such fits to Blackfoot 8-8.



Gardner et al. determined and recommended $m=.25$ as a reasonable value. However, as we can see, the data support quite a range of alternatives. (The value of α is largely dependent on the units used and is not quoted here.) Thus, the careful application of Gardner's rule requires a bit of analysis.

Gardner's Rule

Here are the three pseudo density logs from the three fits on the previous page.



The Wave Equation

The great success of physics in explaining our world and fueling the growth of technology is based fundamentally upon differential equations and more specifically partial differential equations. PDE's are the mathematical statement of the application of basic physical laws to complex systems. For example, a consideration of a constant density fluid leads to the 'scalar wave equation' which is central to most geophysical imaging algorithms. The SWE is a direct consequence of Newton's second law and Hooke's law as applied to the fluid.

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} - \frac{1}{v^2(x,y,z)} \frac{\partial^2 \Psi}{\partial t^2} = f(x,y,z,t)$$

Here Ψ is the pressure, v is the velocity of wave propagation, and $f(x,y,z,t)$ represents any possible sources.

Though it is hardly obvious, the solutions to this equation are traveling waves. A great deal of interesting physical effects can be studied with the SWE including:

- propagation of primaries and multiples
- reflection and transmission at interfaces
- head waves and surface waves
- ray theory, Snell's law
- characterization of sources
- arrays of sources and receivers

The Wave Equation

There is a powerful method of solution of PDE's that is of considerable relevance exploration seismology. This is the method of solution by Green's functions. We will not develop it here but simply state the important results. The essence of the theory is to develop a solution to the PDE of interest for a "point source" and then to show how the response to arbitrary source configurations can be constructed from the elementary solution. The SWE, when specialized for the Green's function problem looks like:

$$\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} + \frac{\partial^2 G}{\partial z^2} - \frac{1}{v^2(x,y,z)} \frac{\partial^2 G}{\partial t^2} = \delta(x-x_o, y-y_o, z-z_o, t-t_o)$$

The term on the right of the equal sign is a Dirac delta function and represents a mathematical impulse at a single point in space, (x_o, y_o, z_o) , and at an instant of time, t_o . The solution to the Green's function problem, $G(x, y, z, t)$, is known exactly for constant velocity and approximately for a number of more complicated situations. G contains all physical effects due to the impulsive source and is properly called an "impulse response".

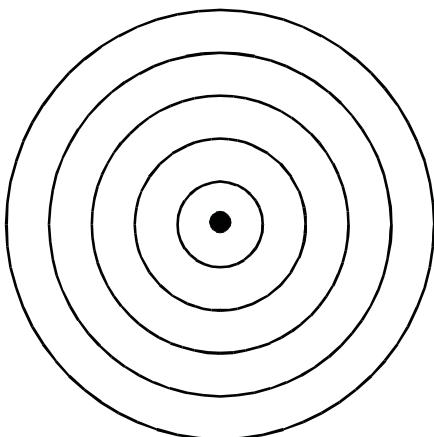
To obtain the response to general source configurations, we imagine the source to be composed of a set of scaled impulses. Then construct the Green's functions for all of these impulses and simply superimpose these Green's functions. This is an example of the mathematical process of "convolution". We will learn more about convolution later in this course. For now, it is enough to visualize it as a general superposition of scaled "impulse responses".

The Wave Equation

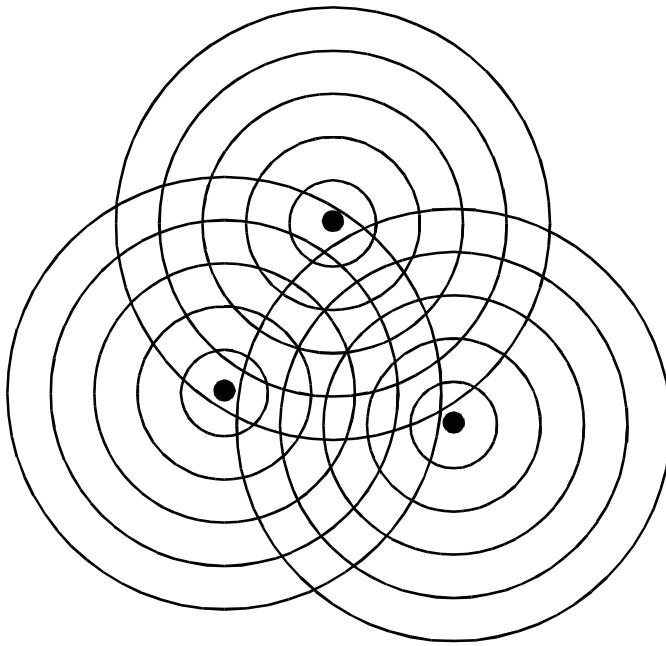
The result we have just obtained is so important that we restate it in different terms:

The wavefield due to a source having extended spatial and temporal form can be considered to be the convolution of the earth's impulse response with the extended source. This result holds for any linear wave equation and extends to elastic, anisotropic and attenuating media.

The two components of this result, the earth's impulse response, I_r , and the source waveform, w_s , are both abstract entities that are difficult to quantify. I_r is generally very complicated and contains all physical effects. w_s is a complete characterization of the source wavefield and can be considered as the specification of the wavefield at all points on a surface surrounding the source.



Impulse response



Response to 3 sources

Traveling Waveforms

The simplest mathematical wave equation is the scalar wave equation. In acoustic media or simple elastic media, compressional waves are governed by it. In 1-D, the scalar wave equation is:

$$\frac{\partial^2 \psi}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2} \quad (1)$$

Where ψ represents the propagating wave. We now show that

$$\psi = f(t \pm z/v) \quad (f \text{ is an arbitrary function})$$

is a solution to (1).

$$\frac{\partial f}{\partial z} = \pm \frac{1}{v} f' , \quad \frac{\partial^2 f}{\partial z^2} = \frac{1}{v^2} f''$$

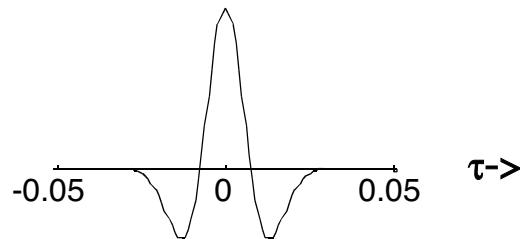
$$\frac{\partial f}{\partial t} = f' , \quad \frac{\partial^2 f}{\partial t^2} = f''$$

Substitution of the second partials of f into (1) results in an immediate identity. Thus f is a solution to (1) with the form of f being arbitrary except that it must be twice differentiable.

Traveling Waveforms

As an example of a waveform, consider the Ricker wavelet defined by:

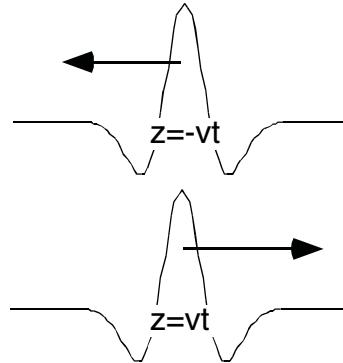
$$w(\tau) = \left(1 - 2(\pi f_{\text{dom}} \tau)^2\right) \exp\left(-(\pi f_{\text{dom}} \tau)^2\right)$$



Shown for $f_{\text{dom}}=30\text{Hz}$

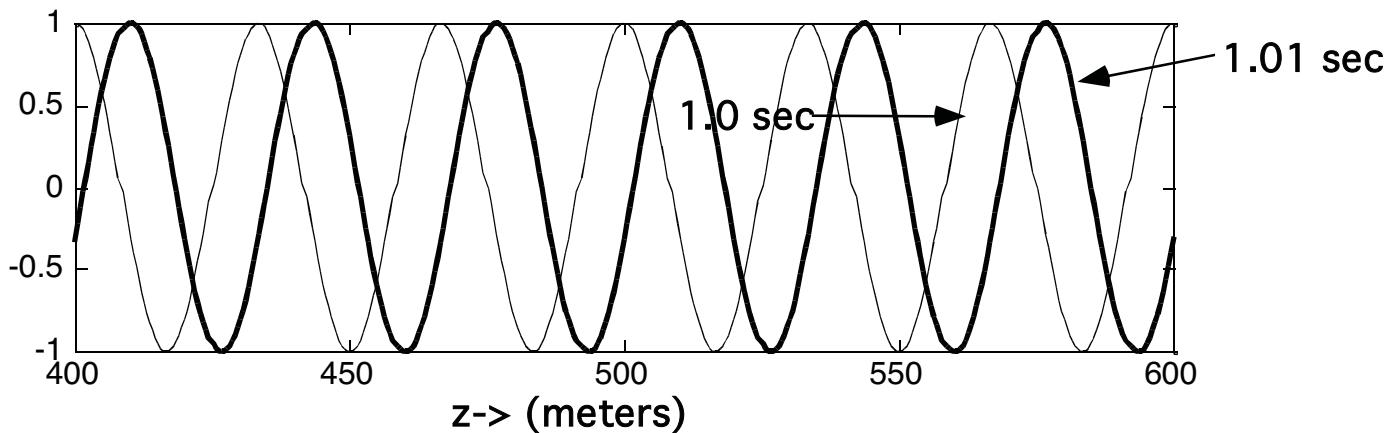
Note that the Ricker wavelet is centered where its argument equals zero. Thus $w(t+z/v)$ represents a wavelet centered at $t+z/v = 0$ or $z = -vt$. So we conclude:

$w(t+z/v)$ = Waveform traveling in the $-z$ direction



$w(t-z/v)$ = Waveform traveling in the $+z$ direction

Similarly, $\cos(\omega(t-z/v))$, $\cos(k(z - vt))$, and $\cos(\omega t - kz)$ all represent cosine waves traveling in the $+z$ direction.

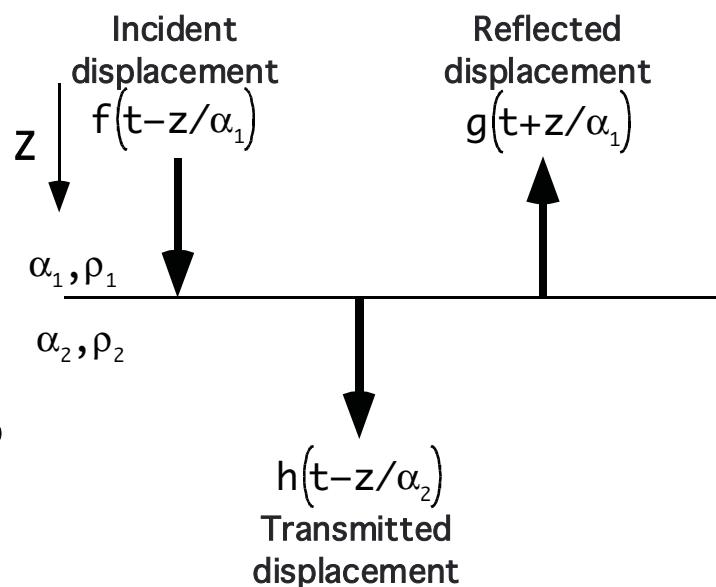


$\cos\left(2\pi 30(t-z/1000)\right)$ Plotted versus z for $t=1.0$ and 1.01 (sec)

Normal Incidence Reflection Coefficients

(Adapted from E.S. Krebes, Course Notes in Theoretical Seismology)

Consider a vertically traveling compressional wave incident on a horizontal interface. In order to describe the reflection and transmission that occur, it can be shown that two conditions must be satisfied:



continuity of displacement: $f + g = h$ (1)

continuity of normal pressure: ??? (2)

To develop a form for the second equation, we use Hooke's law which says stress is proportional to strain.

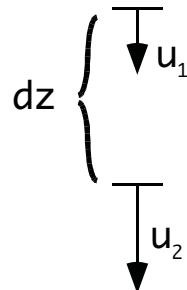
stress = (applied force)/area

strain = (change in length)/length

Normal Incidence Reflection Coefficients

Consider an infinitesimal elastic element whose ends undergo displacement u_1 and u_2 :

$$\text{Strain} = \frac{\Delta l}{l} = \frac{u_2 - u_1}{dz} \approx \frac{\partial u}{\partial z}$$



Now, invoking Hooke's law:

$$\text{stress} = \text{pressure} = \frac{\text{Force}}{\text{area}} = k \frac{\partial u}{\partial z}$$

Where k is a constant formed from the material constants. To determine k , we can use dimensional analysis:

$$\text{pressure} = \frac{\text{force units}}{(\text{length units})^2} = \frac{\text{mass } l/\text{sec}^2}{l^2} = \frac{\text{mass}}{l^3} \left(\frac{l}{\text{sec}} \right)^2$$

So, k looks like: $k = \rho \alpha^2$ Thus the pressure continuity equation is:

$$\rho_1 \alpha_1^2 \frac{\partial f}{\partial z} + \rho_1 \alpha_1^2 \frac{\partial g}{\partial z} = \rho_2 \alpha_2^2 \frac{\partial h}{\partial z} \quad (\text{evaluated at the interface})$$

$$\text{But since } \frac{\partial f}{\partial z} = \frac{-1}{\alpha_1} f', \quad \frac{\partial g}{\partial z} = \frac{1}{\alpha_1} g', \quad \frac{\partial h}{\partial z} = \frac{-1}{\alpha_2} h'$$

$$\rho_1 \alpha_1 f' - \rho_1 \alpha_1 g' = \rho_2 \alpha_2 h'$$

Which can be immediately integrated to give:

$$\rho_1 \alpha_1 f - \rho_1 \alpha_1 g = \rho_2 \alpha_2 h \quad (2)$$

Normal Incidence Reflection Coefficients

Assume that an interface occurs at $z=0$, then if the boundary conditions are applied there, the two equations determining normal incidence reflection and transmission are:

$$f + g = h \quad (1)$$

$$I_1 f - I_1 g = I_2 h \quad (2)$$

Where impedance = $I_k = \rho_k \alpha_k$, $k=\{1,2\}$

and where f, g , and h are understood to be evaluated at $z=0$.

Multiplying (1) by I_2 , and subtracting it from (2) leads to:

$$g = \frac{I_1 - I_2}{I_1 + I_2} f = -Rf$$

Similarly, we can obtain:

$$h = \frac{2I_1}{I_1 + I_2} f = Tf$$

The quantities R and T are known as the normal incidence reflection and transmission coefficients:

$$R = \frac{I_2 - I_1}{I_1 + I_2}, \quad T = \frac{2I_1}{I_1 + I_2}$$

Note that: $R+T = \frac{I_2 - I_1 + 2I_1}{I_1 + I_2} = 1$

Normal Incidence Reflection Coefficients

$$R = \frac{I_2 - I_1}{I_1 + I_2}, \quad T = \frac{2I_1}{I_1 + I_2}$$

R and T are often written in terms of the contrast and average of impedance across the layer:

$$I = \frac{1}{2}(I_1 + I_2), \quad \Delta I = I_2 - I_1$$
$$I_1 = I - .5\Delta I, \quad I_2 = I + .5\Delta I$$

Straight forward algebra then gives:

$$R = \frac{\Delta I}{2I} \approx \frac{1}{2} \frac{d(\ln(I))}{dz} \Delta z$$
$$T = 1 - R = \frac{I - .5\Delta I}{I}$$

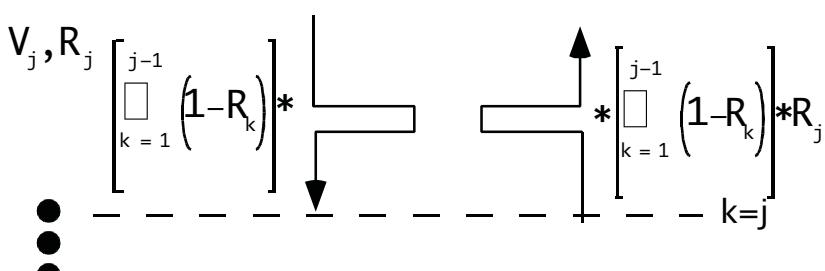
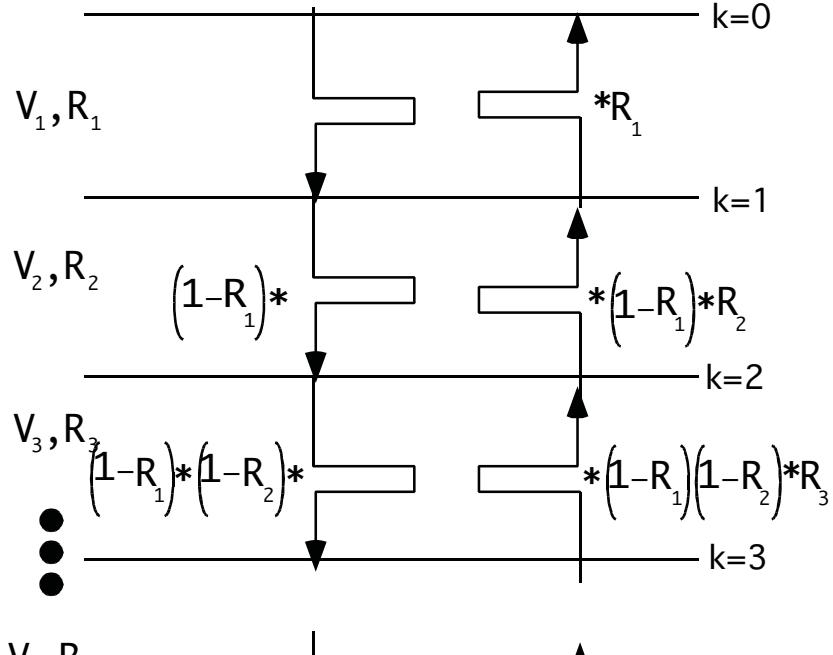
R can be written in terms of ρ and α as:

$$R = \frac{\Delta(\rho\alpha)}{2\rho\alpha} = \frac{\rho\Delta\alpha + \alpha\Delta\rho}{2\rho\alpha} = \frac{1}{2} \left(\frac{\Delta\alpha}{\alpha} + \frac{\Delta\rho}{\rho} \right)$$

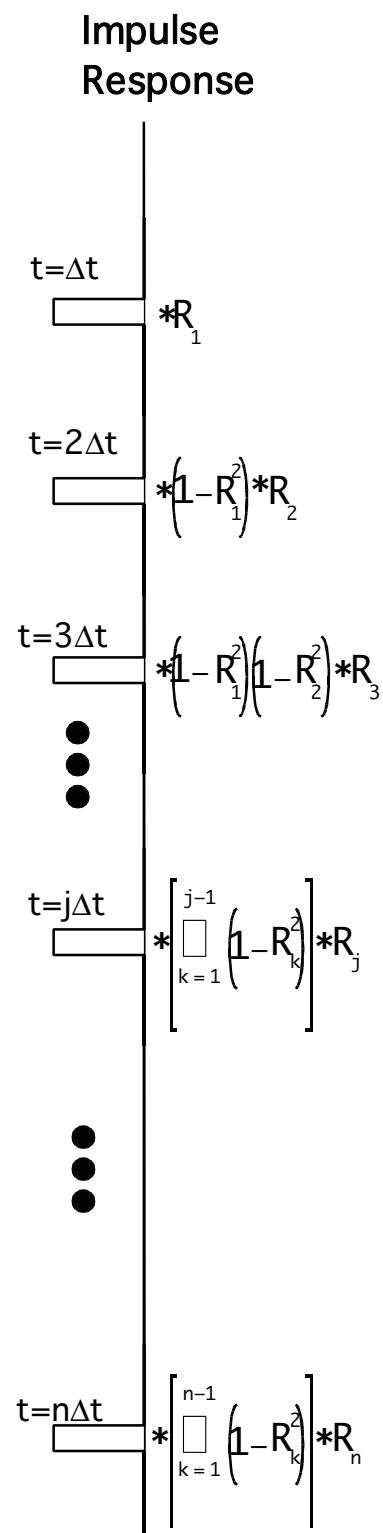
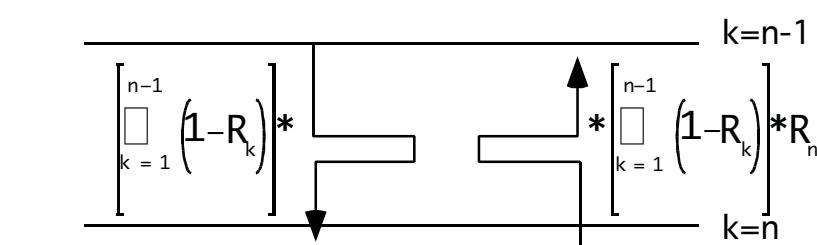
Note that the definition of R is such that an impedance increase gives a positive RC but that the reflected pulse is flipped in polarity.

Simple "Primaries Only" Impulse Response.

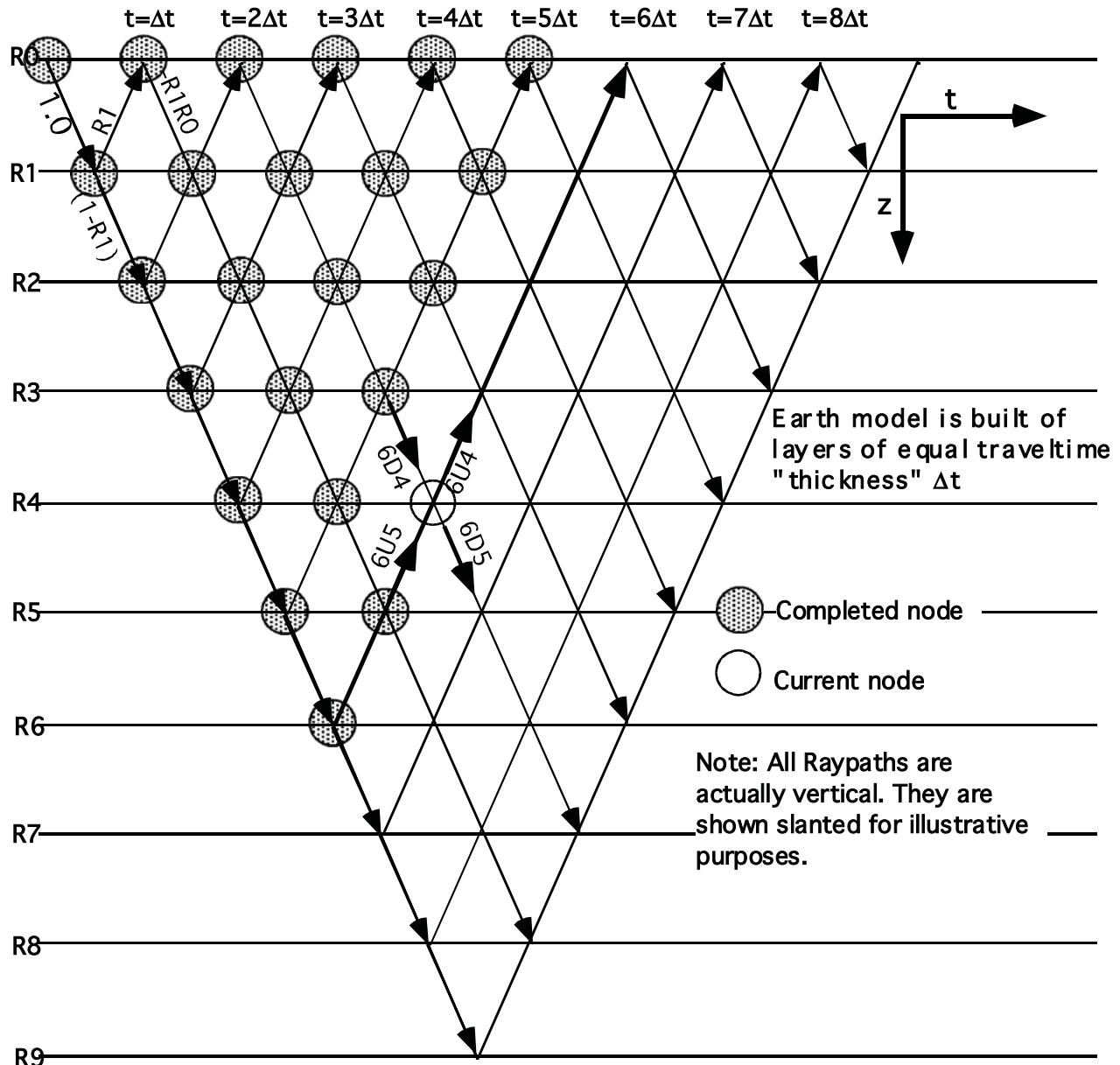
Layered Earth, Normal Incidence, Acoustic Model



Model layers have a constant traveltime $\Delta t = 2 \frac{\Delta Z_j}{V_j}$
"thickness":



Computation of a 1-D Synthetic Seismic Impulse Response (Including All Multiples)



At the designated point, $6D_4$ and $6U_5$ are known and we wish to compute $6U_4$ and $6D_5$:

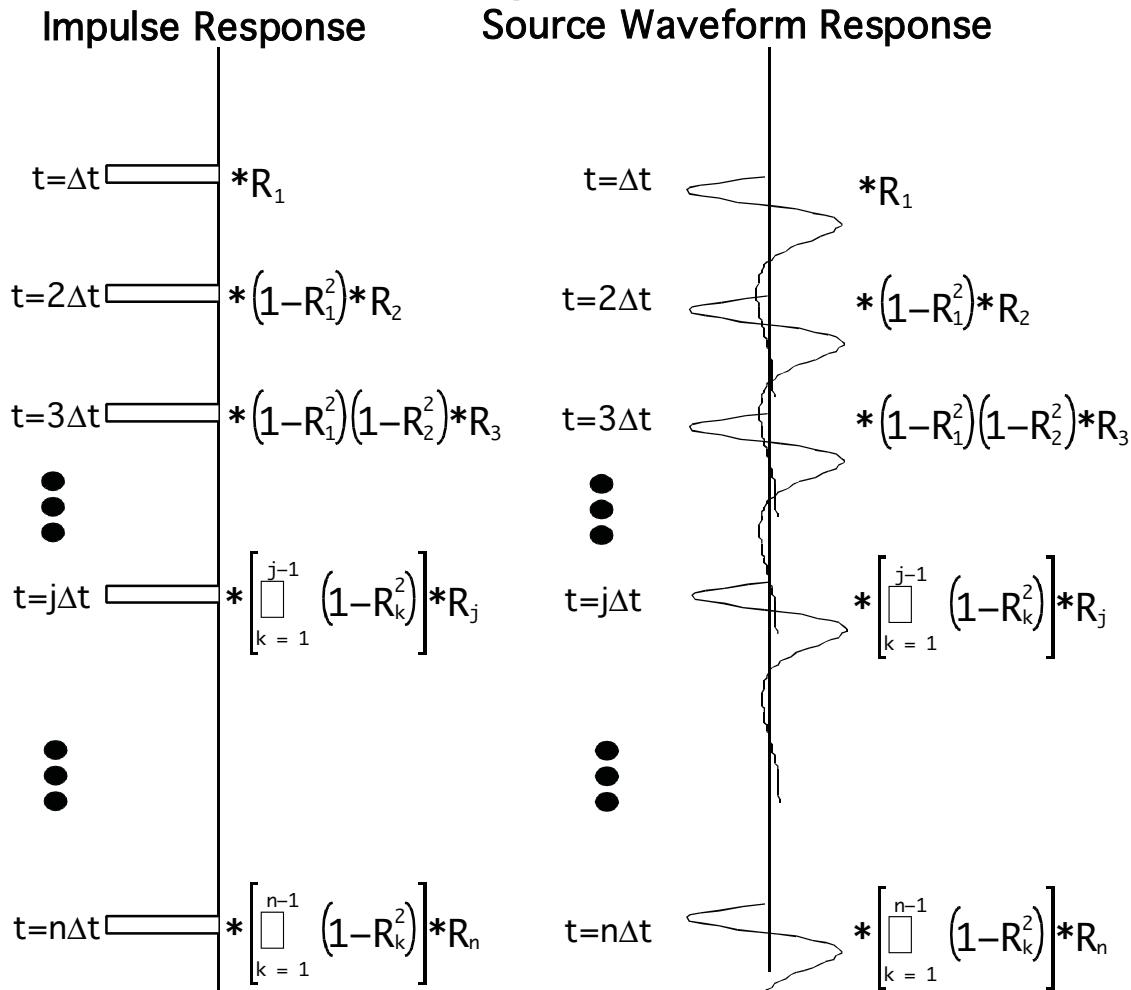
$$6U_4 = R_4 * 6D_4 + (1+R_4) * 6U_5$$

$$6D_5 = (1-R_4) * 6D_4 - R_4 * 6U_5$$

The complete seismogram is obtained by recursive calculation beginning in the upper left. All nodes on any upward traveling ray are completely calculated before proceeding to the next depth.

Adapted from: Reflection Seismology, K.H. Waters, 1981
J.H. Wiley

From Impulse Response to Source Waveform Response



The "primaries only" impulse response consists of a time series of scaled and delayed impulses

To obtain the source waveform response from the impulse response, simply replace each spike of the impulse response by the product of the spike and source waveform. This is the mathematical process of convolution

Impulse Responses and Seismograms

For a linear earth, it can be shown that if we are given the waveform signature of a non-impulsive source and the impulse response of an earth model, then:

$$s(t) = I_r(t) \bullet w_s(t)$$

where:

$I_r(t)$ is the earth impulse response

$w_s(t)$ is the source waveform

$s(t)$ is the earth response to the source waveform

The general proof of this result comes from "Green's function analysis" and is true for any linear wave equation (elastic, scalar, etc). Generally I_r contains all physical effects the theory is capable of producing, and usually that is more than we want.

The most common use of 1-D seismograms is in the interpretation of processed seismic sections. In this case most of the physical effects (multiples, transmission losses, attenuation) have been removed in the processing. Therefore, common practice replaces $I_r(t)$ with $r(t)$ where:

$r(t)$ = normal incidence reflection coefficients positioned in 2-way vertical traveltime

Thus:

$$s(t) = r(t) \bullet w_s(t)$$

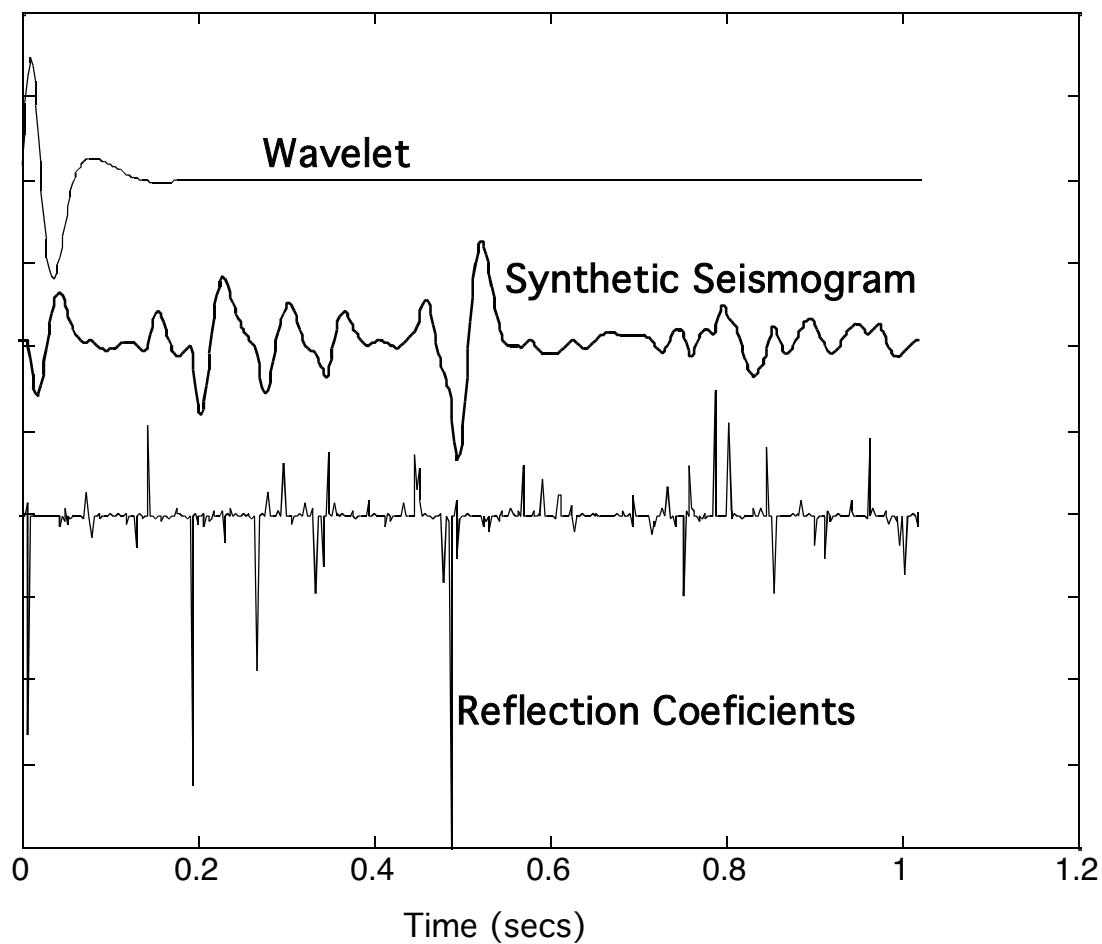
$s(t)$ given by this result is the most common 1-D seismogram computed in exploration geophysics.

1-D Synthetic Seismogram Summary

- A complete solution, generating all multiples and transmission effects, can be constructed. Some methods also include attenuation.
- Assumptions: ray theory, 1-D, normal incidence
- Geophysical well logs, providing P-wave velocities and densities, are used. They are usually resampled to a variable depth layering with equal Dt steps.
- Method is inherently algorithmic. No analytic closed form solution available.
- In practice, multiples and transmission losses are not usually included. Reflection coefficients in time are simply convolved with a source response.

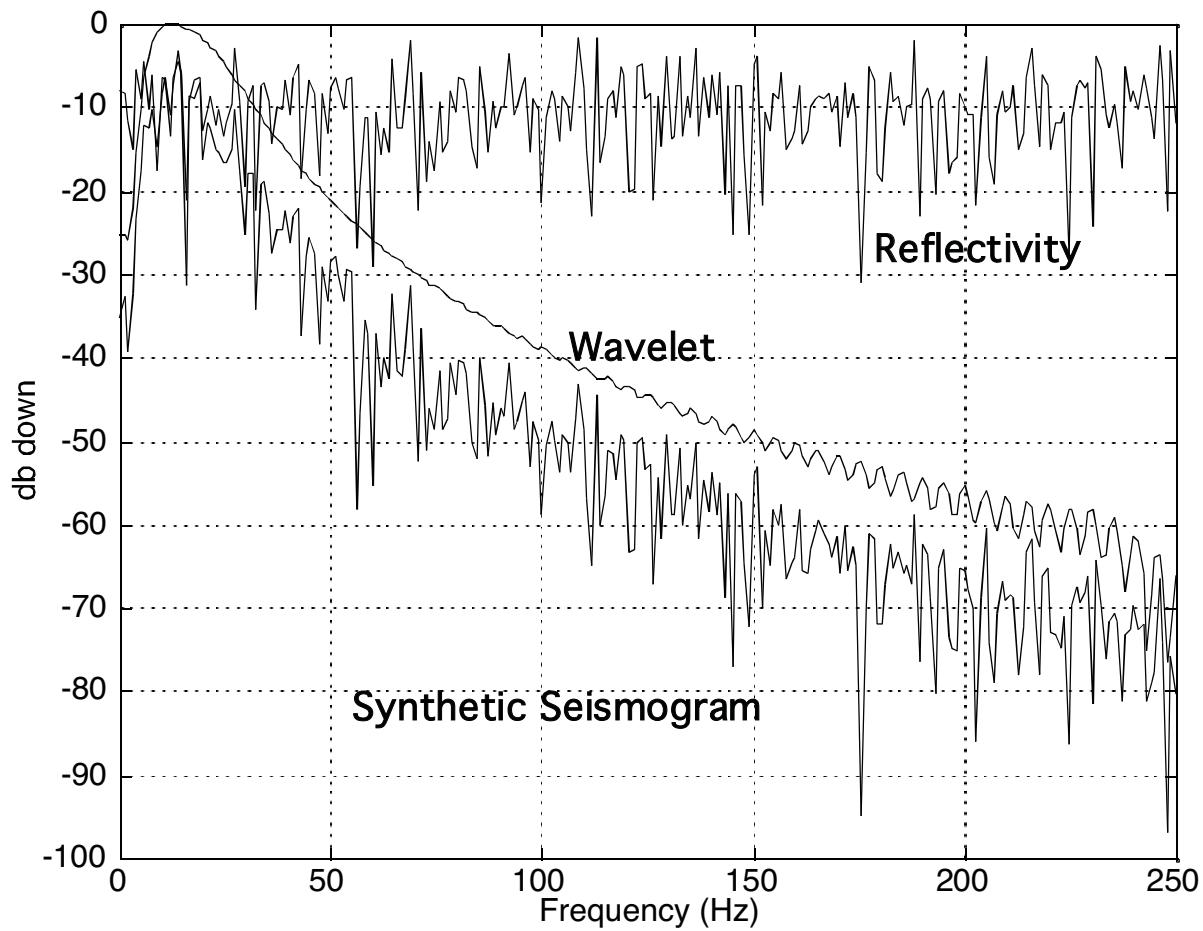
Example of Synthetic Seismogram Creation by Convolution of Reflectivity and Wavelet.

Time Domain View



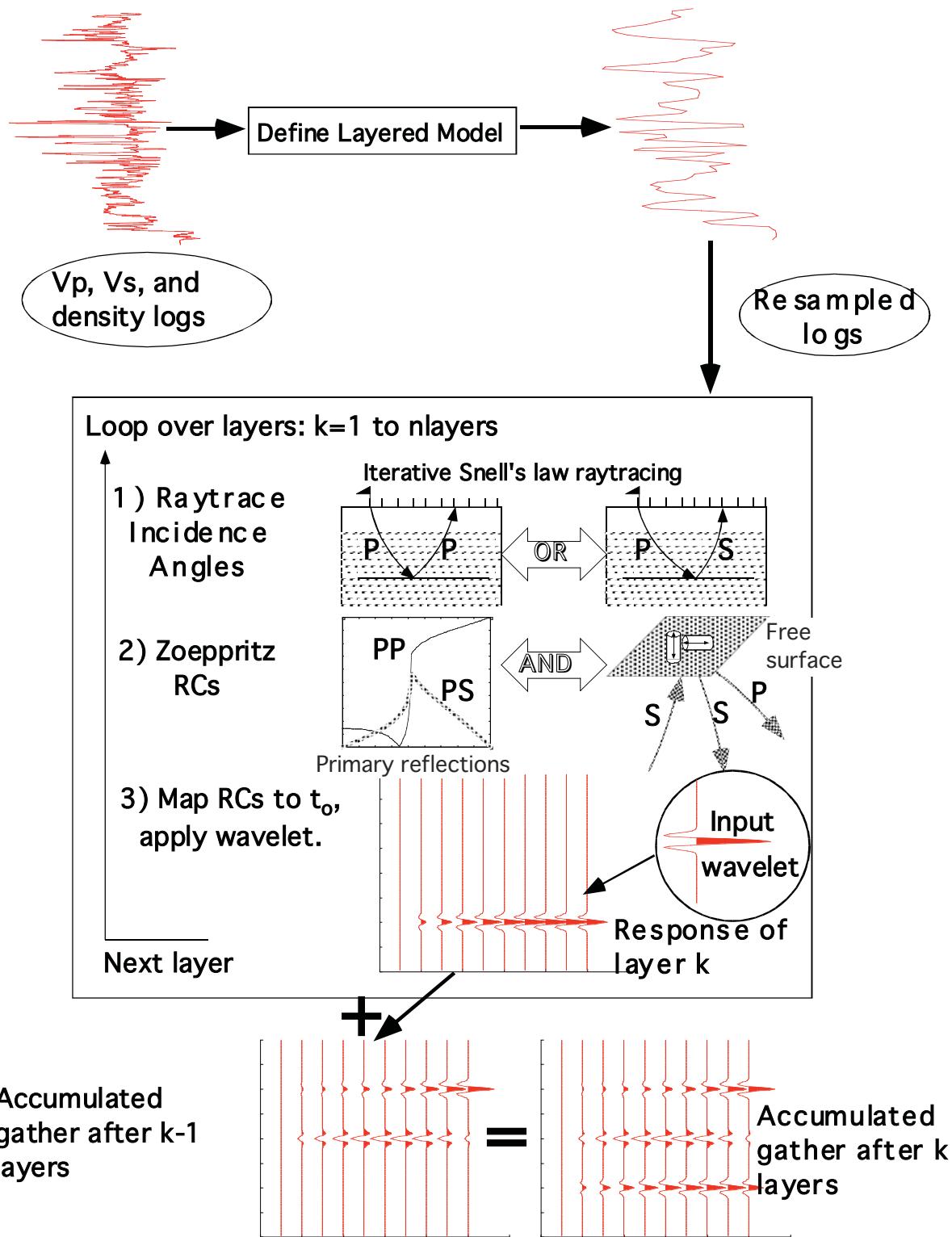
Example of Synthetic Seismogram Creation by Convolution of Reflectivity and Wavelet.

Frequency Domain View



P-S Synthetic Seismogram Construction

The SYNTH Algorithm



Methods of Seismic Data Processing

**Lecture Notes
Geophysics 557**

**Chapter 2
Signal Processing**

Convolution

Convolution is the mathematical process of "shifting, scaling, and summing" a waveform to produce an output by superposition. Generally, two input signals are required, say r and w , with w being the waveform and r a series of scaling coefficients. For example, let $r = [1 \ 0 \ 0 \ -.5 \ .5 \ 0 \ -1]$ and let $w = [-.5 \ 1 \ -.5]$, then the convolution of r and w is:

		Output sample number																			
		j →	0	1	2	3	4	5	6	7	8										
Input sample number k ↓	0	$r_0w_0 \ r_0w_1 \ r_0w_2 = r_0 * w$																			
	1	-.5	1	-.5	0	0	0	0	0	0	0										
	2	0	0	0	0	0	0	0	0	0	0										
	3	0	0	0	.25	-.5	.25	0	0	0	0										
	4	0	0	0	0	-.25	.5	-.25	0	0	0										
	5	0	0	0	0	0	0	0	0	0	0										
	6	0	0	0	0	0	0	.5	-1	.5											
		<hr/>																			
$s = r \cdot w$																					
<table border="1" style="width: 100%;"><tr><td>-.5</td><td>1</td><td>-.5</td><td>.25</td><td>-.75</td><td>.75</td><td>.25</td><td>-1</td><td>.5</td><td></td><td></td></tr></table>											-.5	1	-.5	.25	-.75	.75	.25	-1	.5		
-.5	1	-.5	.25	-.75	.75	.25	-1	.5													

Convolution

In the previous slide, we described a tabular method for computing the convolution of r and w to yield s . This can be written mathematically as follows:

$$S = r \bullet w$$
$$S_j = \sum_k r_k w_{j-k}$$

To see that this summation expression is equivalent to the tabular method, consider the example of $j=4$:

$$S_4 = r_0 w_{4-0} + r_1 w_{4-1} + r_2 w_{4-2} + r_3 w_{4-3} + r_4 w_{4-4} + r_5 w_{4-5} + r_6 w_{4-6}$$
$$S_4 = r_0 w_4 + r_1 w_3 + r_2 w_2 + r_3 w_1 + r_4 w_0 + r_5 w_{-1} + r_6 w_{-2}$$

Note that the length of s is the combined lengths of r and w less 1:

$$\text{length}(s) = \text{length}(r) + \text{length}(w) - 1$$

Thus, mathematically, everytime a convolution is performed the result increases in length. This creates a bit of a header (bookkeeping) problem in seismic data processing and is not usually allowed. That is, if a seismic trace is convolved with a filter operator, the result is truncated at the same length as the seismic trace.

Convolution

We have seen that the convolution of discretely sampled vectors is written:

$$S_j = \sum_k r_k w_{j-k}$$

The analogous result for continuous functions is:

$$s(t) = \int_{-\infty}^{\infty} r(\tau)w(t-\tau)d\tau$$

We now show that the order of convolution is immaterial.
Let:

$$\tau' = t - \tau, \quad d\tau' = -d\tau, \quad \tau = t - \tau'$$

Then: $s(t) = - \int_{\infty}^{-\infty} r(t-\tau')w(\tau')d\tau'$

And: $s(t) = \int_{-\infty}^{\infty} r(t-\tau')w(\tau')d\tau'$

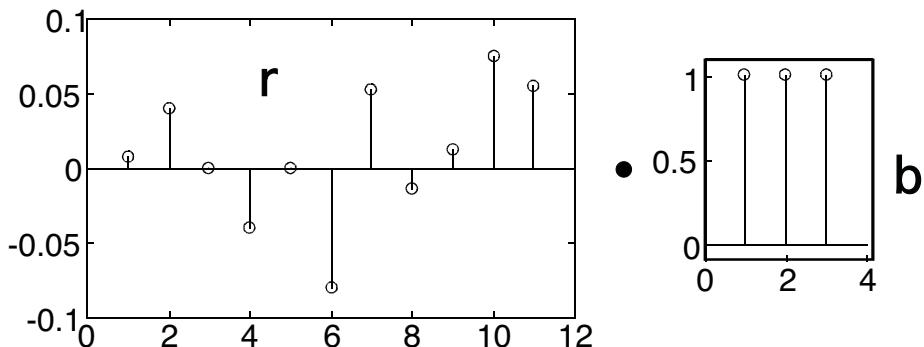
So: $S = r \bullet w = w \bullet r$

We also note that convolution is linear in the sense that:

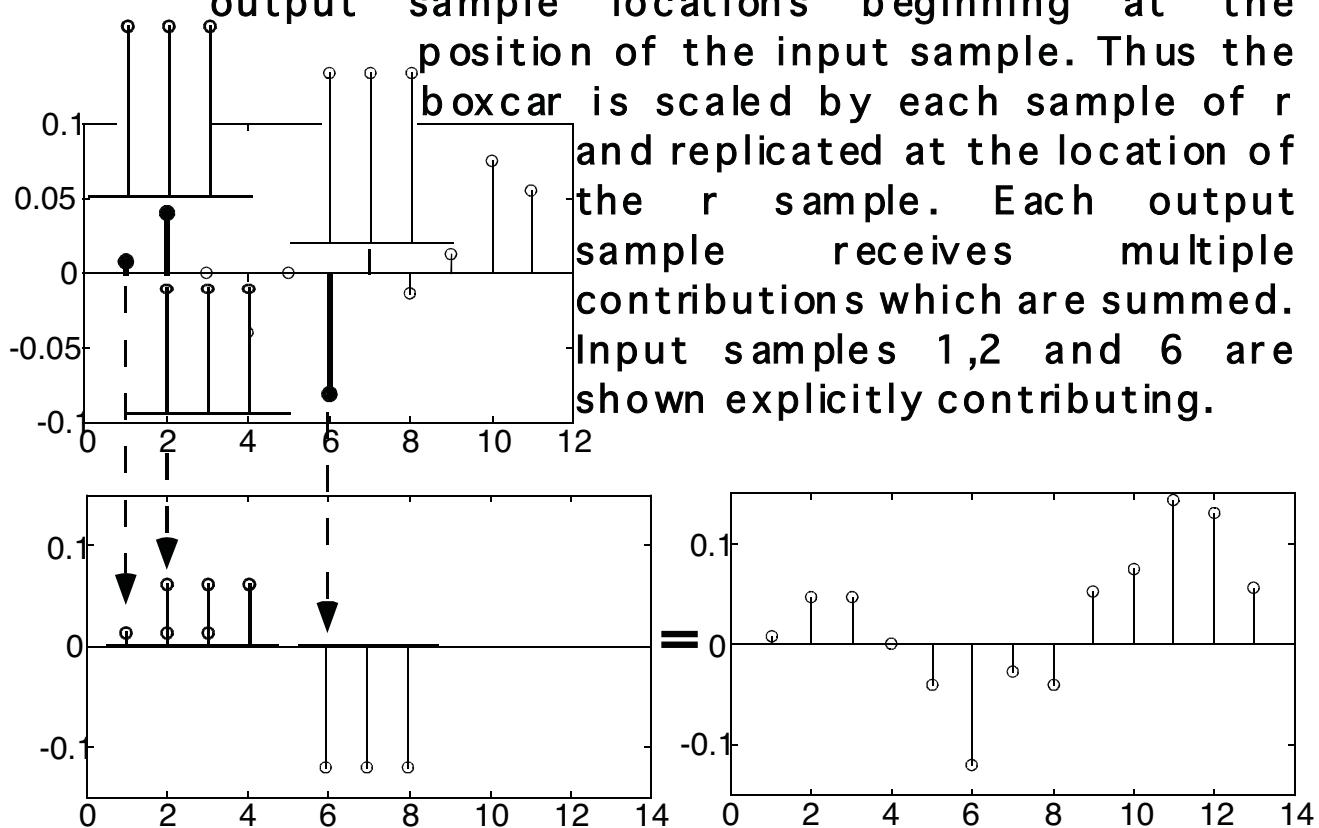
$$(a+b) \bullet c = a \bullet c + b \bullet c$$

Convolution by Replacement (emphasis on input samples)

Consider the discrete convolution of a three point boxcar, b, with an eleven point time series, r.

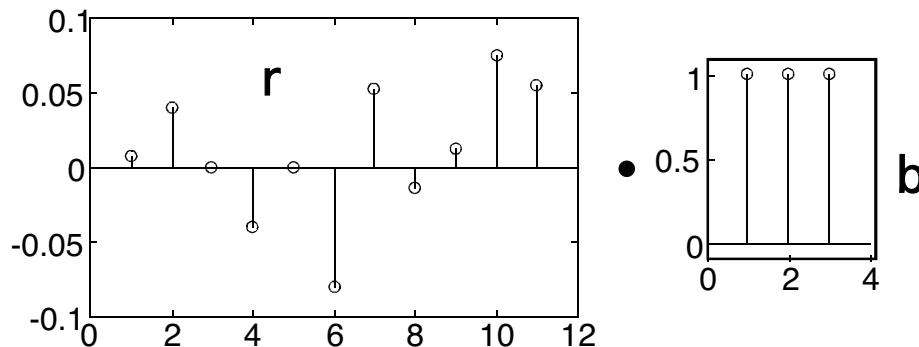


Each input sample is considered separately. The boxcar is multiplied by the input sample resulting in a scaled boxcar. The scaled boxcar contributes to output sample locations beginning at the position of the input sample. Thus the boxcar is scaled by each sample of r and replicated at the location of the r sample. Each output sample receives multiple contributions which are summed. Input samples 1,2 and 6 are shown explicitly contributing.

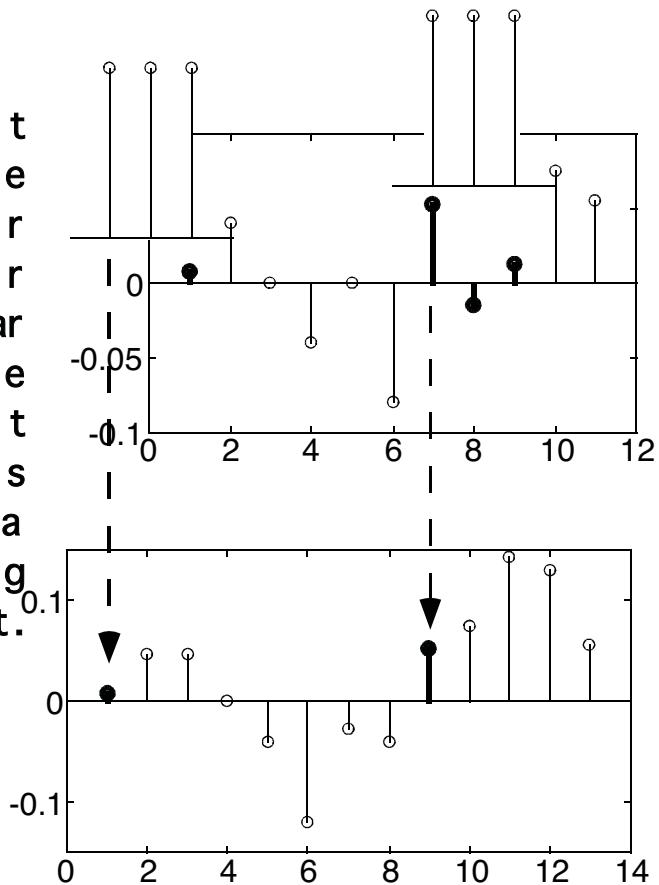


Convolution as a Weighted Sum (emphasis on output samples)

Consider the discrete convolution of a three point boxcar, b, with an eleven point time series, r.



To compute an output sample, position the boxcar over some r samples, multiply the r samples by the boxcar weights, and sum. The computation of output samples 1 and 7 is illustrated. This is a process of smoothing or averaging the input.



Matrix Multiplication by Rows

Consider the a 4x4 matrix equation such as:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} \quad \text{eqn 1}$$

This is equivalent to the following system of equations:

$$c_1 = a_{11}b_1 + a_{12}b_2 + a_{13}b_3 + a_{14}b_4$$

$$c_2 = a_{21}b_1 + a_{22}b_2 + a_{23}b_3 + a_{24}b_4 \quad \text{eqns 2a-2d}$$

$$c_3 = a_{31}b_1 + a_{32}b_2 + a_{33}b_3 + a_{34}b_4$$

$$c_4 = a_{41}b_1 + a_{42}b_2 + a_{43}b_3 + a_{44}b_4$$

Thus the elements of the vector C are computed by taking each row of A, multiplying it by the vector B, and summing the results. This process is familiar to most students of linear algebra as "matrix multiplication by rows". It can be written symbolically as two nested computation loops:

```
c=zeros(1,4);
for irow=1:4
    for jcol=1:4
        c(irow)=c(irow) + a(irow,jcol)*b(jcol);
    end
end
```

Matrix Multiplication by Columns

Matrix multiplication "by columns" is less well known than the corresponding process "by rows" but it provides a useful intuitive insight to convolution. Examination of equations 2a-2d shows that the columns of A have been multiplied by a single corresponding element of B. Thus we can express the matrix multiplication as a sum of column vectors, each one being a scaled version of a column of A.

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} b_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} b_2 + \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix} b_3 + \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix} b_4 = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$$

Written as computation loops, this amounts to reversing the order of the loops in the multiplications "by rows"

```
c=zeros(1,4);
for jcol=1:4
    for irow=1:4
        c(irow)=c(irow) + a(irow,jcol)*b(jcol);
    end
end
```

Convolution as a Matrix Operation

Consider the convolution of a reflectivity sequence, r , with a wavelet, w , to yield a seismic trace, s . This is usually written as the convolution integral:

$$s(t) = \int_{-\infty}^{\infty} w(t - \tau)r(\tau)d\tau$$

When we have discrete, finite length approximations to these quantities, the convolution is usually written as a summation. If r_j is the reflectivity series with $j=0,1,\dots,n$, and w_k is the possibly non-causal wavelet with $k=-m\dots 0\dots m$, then:

$$s_k = \Delta t \sum_{j=k+m}^{k-m} w_{k-j} r_j$$

Usually, in these expressions, the Δt term is dropped or set to unity. It is useful to write out a few terms of this summation:

$$s_0 = \dots + w_0 r_0 + w_{-1} r_1 + w_{-2} r_2 + \dots$$

$$s_1 = \dots + w_1 r_0 + w_0 r_1 + w_{-1} r_2 + \dots$$

The same operation can be achieved by matrix multiplication where the wavelet, w , is loaded into a special matrix called a Toeplitz or convolution matrix.

Convolution as a Matrix Operation

It is a simple exercise of matrix multiplication by rows to check that the following matrix equation computes the convolution of w with r

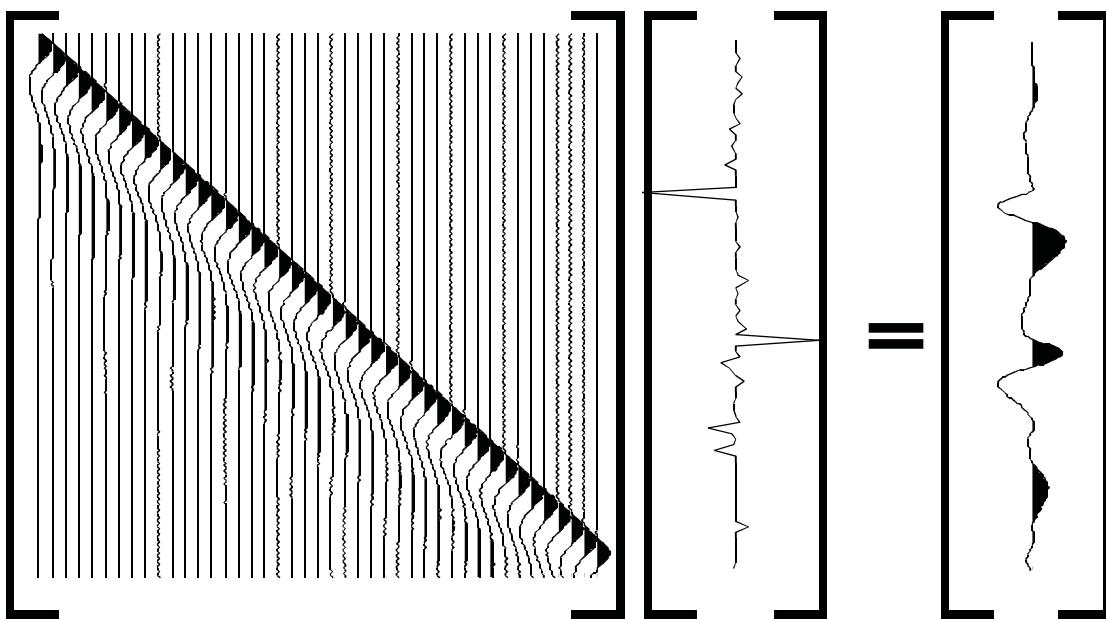
$$\begin{bmatrix} \dots & \dots & \dots & \dots & \vdots \\ w_0 & w_{-1} & w_{-2} & w_{-3} & \vdots \\ w_1 & w_0 & w_{-1} & w_{-2} & \vdots \\ w_2 & w_1 & w_0 & w_{-1} & \vdots \\ w_3 & w_2 & w_1 & \ddots & \vdots \\ \dots & \dots & \dots & \dots & w_0 \end{bmatrix} \begin{bmatrix} \vdots \\ r_0 \\ r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} = \begin{bmatrix} \vdots \\ s_0 \\ s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

Note the symmetry of the W matrix which has the wavelet samples constant along the diagonals. Another way to view W is that each column contains the wavelet with the zero time sample aligned on the main diagonal. Now, imagine doing the matrix multiplication by columns instead of rows and we get the most intuitive view of convolution "by replacement".

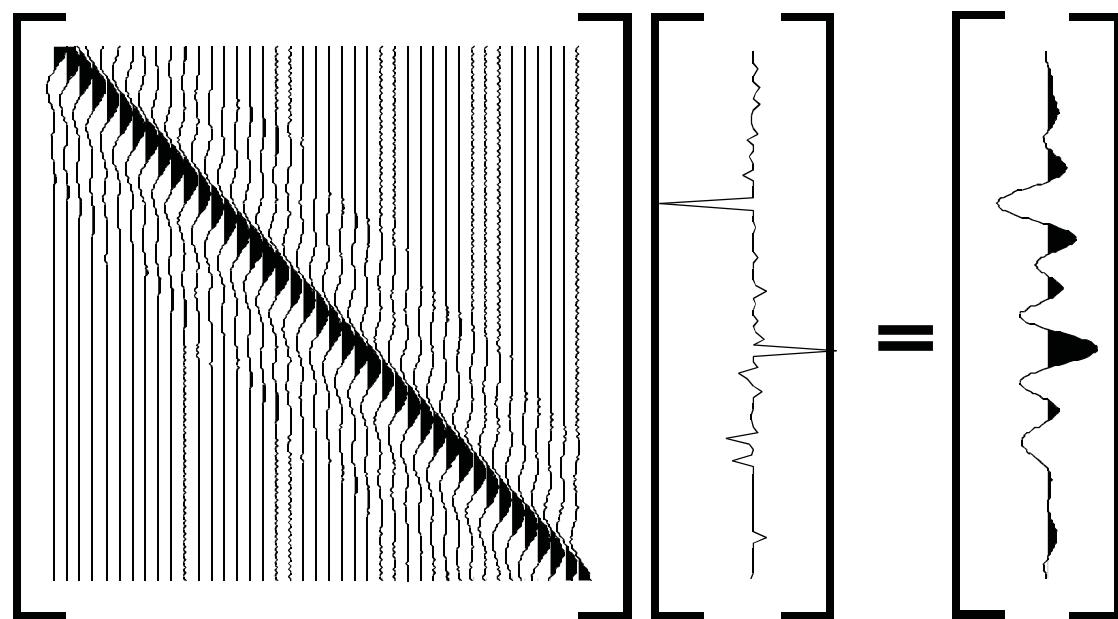
$$\begin{bmatrix} \vdots \\ w_0 \\ w_1 \\ w_2 \\ w_3 \\ \vdots \end{bmatrix} r_0 + \begin{bmatrix} \vdots \\ w_{-1} \\ w_0 \\ w_1 \\ w_2 \\ \vdots \end{bmatrix} r_1 + \dots = \begin{bmatrix} \vdots \\ s_0 \\ s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

Convolution as a Matrix Operation

As an example of convolution by matrix multiplication, here is an illustration of the convolution of a reflectivity series and a minimum phase wavlet to yield a 1-D seismogram.



As a second example, here is the convolution of a reflectivity series and a zero phase wavlet to yield a zero phase seismogram.



Convolution as a Matrix Operation

These examples of convolution by matrix multiplication show explicitly what is meant when we say that convolution is a stationary process. Intuitively, this phrase means that the operation does not change with time in some sense. Precisely, it means that the waveforms in the columns of the convolution matrix are all identical. That is, the wavelet which is scaled and used to replace each reflectivity spike does not change with time. As we shall see, many physical processes violate this assumption and it is quite possible to generalize the convolution operation to model nonstationary processes.

When the assumption of stationarity is made in the context of statistical deconvolution theory, it means precisely the same thing. We assume that the time series we measured (the seismic trace) is related to that which we want (the reflectivity) by a stationary convolution operation. Given that, we expect that a stationary inverse operator will suffice to recover the reflectivity.

Fourier Transforms and Convolution

Consider the convolution integral for continuous functions:

$$h(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau = f(t) \bullet g(t)$$

Now, let g be a complex sinusoidal function: $g(u) = e^{i\omega u}$

Then:
$$h(t) = \int_{-\infty}^{\infty} f(\tau)e^{i\omega(t-\tau)}d\tau = e^{i\omega t}F(\omega) \quad (1)$$

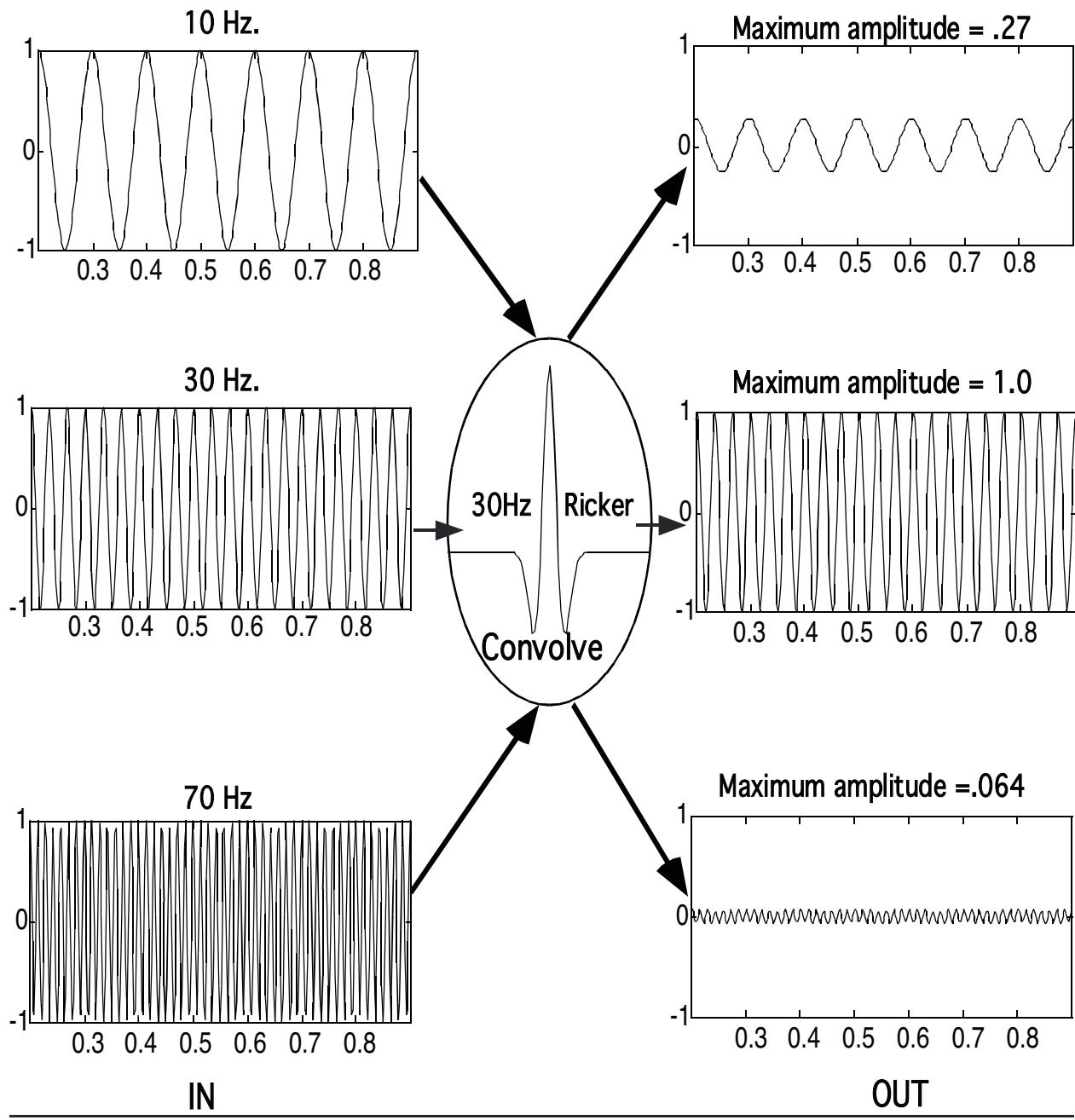
where
$$F(\omega) = \int_{-\infty}^{\infty} f(\tau)e^{-i\omega\tau}d\tau \quad (2)$$

This remarkable result shows that if we convolve ANY function, f , with a complex sinusoid, the result is the same complex sinusoid multiplied by a complex coefficient. This complex coefficient, $F(\omega)$, is computed from $f(t)$ and is known as the Fourier Transform of $f(t)$.

Those who have studied mathematical physics will recognize that this means that the complex sinusoids are eigenfunctions of the convolution operator and the Fourier Transform provides the eigenvalues.

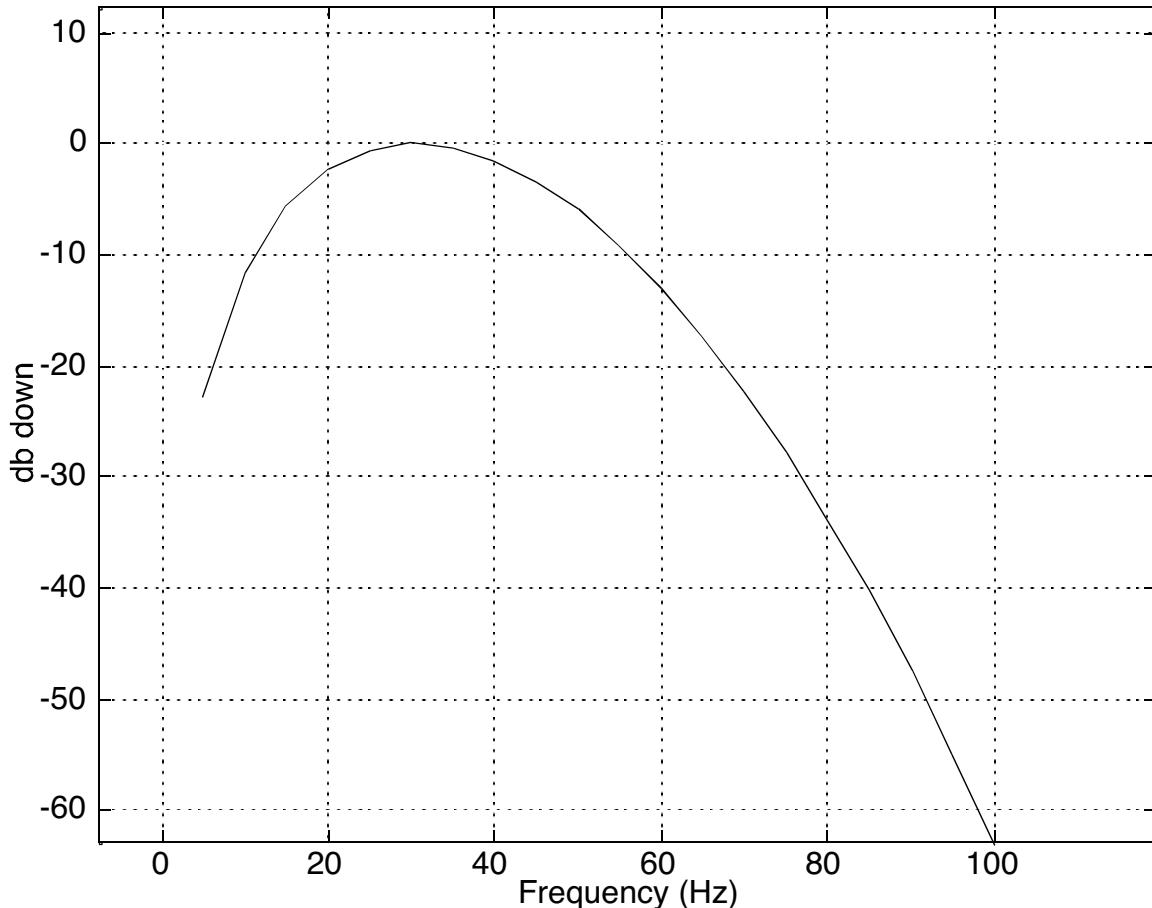
Fourier Transforms and Convolution

Here we see the result of convolving 10, 30, and 70 Hz complex sinusoids with a 30Hz Ricker wavelet. In each case, only the real parts of the complex sinusoids are plotted. We see that the 10 Hz sinusoid is diminished by 73%, the 70 Hz by 93%, and the 30Hz is unattenuated. (The distortions in the sinusoids are artifacts of the display not the convolution algorithm.)



Fourier Transforms and Convolution

Here is the actual Fourier amplitude spectrum of the 30Hz Ricker wavelet.



Since "decibels down" are computed by

$$\text{dbdown} = 20 \cdot \log_{10}(F(\omega)/F_{\max})$$

we can use the results from the previous figure to compute:

$$\text{dbdown}(10\text{Hz}) = 20 \cdot \log_{10}(0.27) = -11.4 \text{ decibels}$$

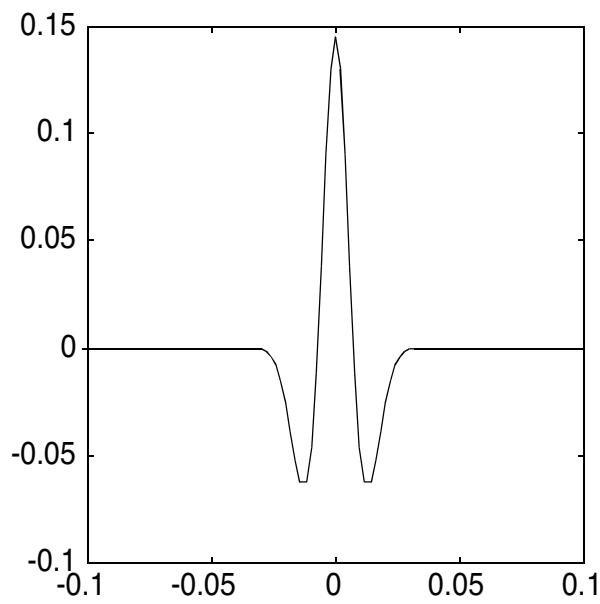
$$\text{dbdown}(30\text{Hz}) = 20 \cdot \log_{10}(1.0) = 0 \text{ decibels}$$

$$\text{dbdown}(70\text{Hz}) = 20 \cdot \log_{10}(0.064) = -23.9 \text{ decibels}$$

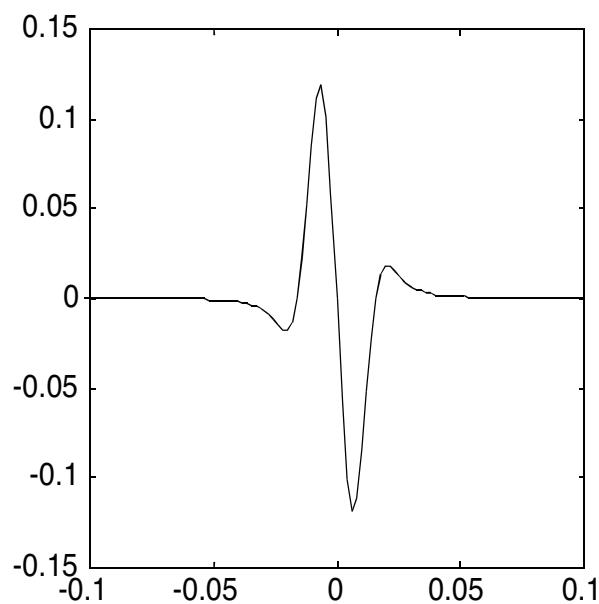
So $F(w)$, the Fourier Transform of a function $f(t)$, is a quick way of computing the relative attenuation of different sinusoids when they are convolved with $f(t)$.

Fourier Transforms and Convolution

A convolution can affect not only the amplitude of a sinusoid but its phase as well. The Ricker wavelet is known as a zero phase function which means that it does not have a phase effect. Let us repeat the analysis but this time with a function which has a known phase effect. For this purpose, we consider a Ricker wavlet with a 90° phase shift.



30 Hz. Ricker zero phase

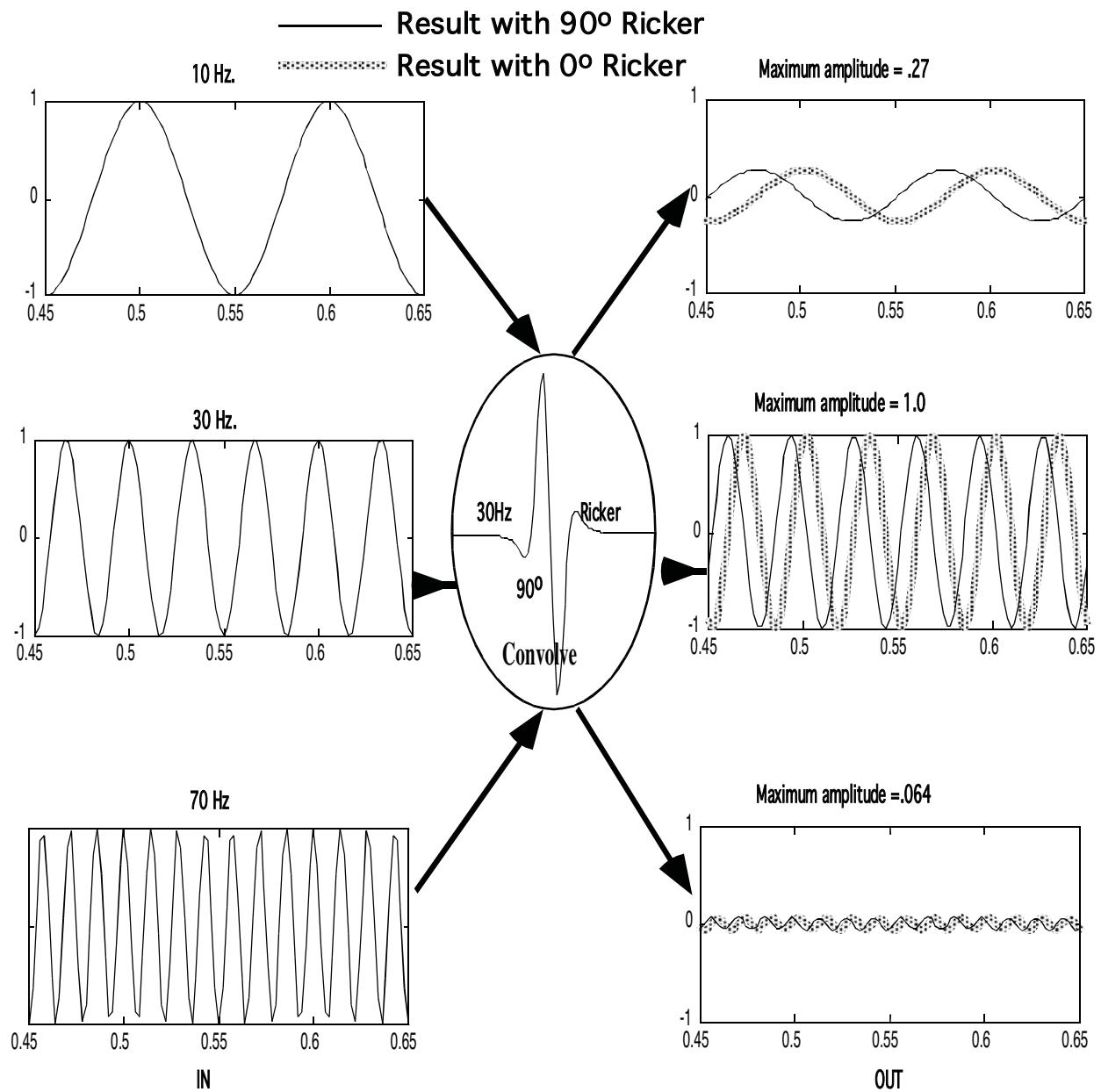


30 Hz Ricker 90° phase

Note that zero phase waveforms are always symmetric while 90° phase results in an antisymmetric waveform. We might expect the 90° Ricker to have the same effect on the amplitude of sinusoids but some additional effect as well. To see, we repeat the analysis of passing complex sinusoids through it.

Fourier Transforms and Convolution

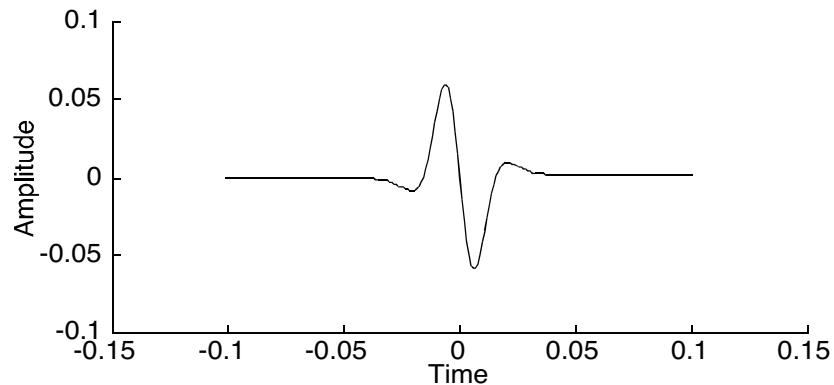
Here we repeat the result of convolving 10, 30, and 70 Hz complex sinusoids with a 30Hz Ricker wavelet but this time the Ricker has 90° phase. The amplitude attenuation of the sinusoids is the same as before but now there is an additional 90° phase lag. (When comparing this figure with -2- of this series, note that there has been an x-axis scale change on all plots.)



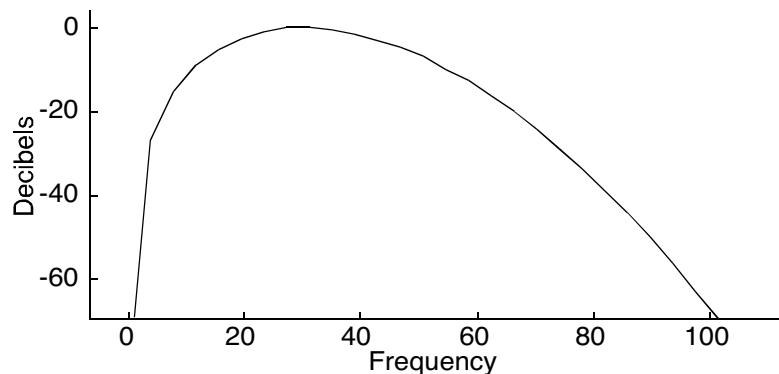
Fourier Transforms and Convolution

Here is a complete description of the 90° , 30 Hz. Ricker in the time domain and amplitude and phase spectrum in the Fourier domain. We have seen that the Fourier domain provides a convenient description of the effect of convolving the wavelet with complex sinusoids.

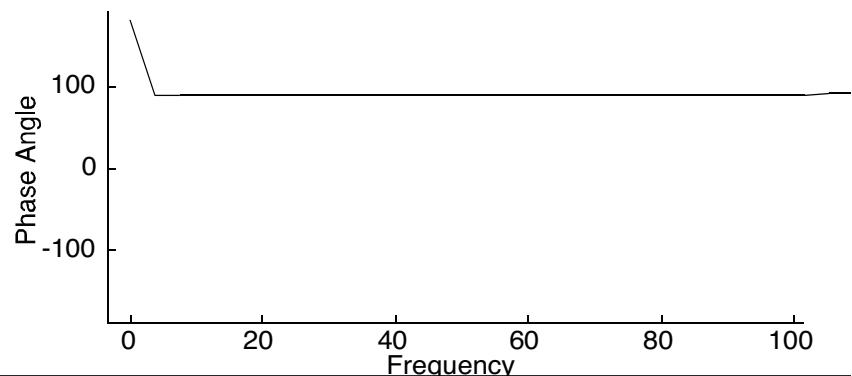
Time Domain



Fourier Domain
Amplitude
Spectrum



Fourier Domain
Phase Spectrum



Fourier Analysis and Synthesis

The great utility of the Fourier transform comes from its ability to decompose any function into a set of complex sinusoids. In the continuous case, the frequencies of the sinusoids range from $-\infty$ to ∞ and have amplitudes and phases which are computed from the forward Fourier transform:

$$H(\omega) = \int_{-\infty}^{\infty} h(t) e^{-i\omega t} dt$$

This equation computes the complex coefficients, $H(\omega)$, of the complex sinusoids which, when summed (integrated), will yield $h(t)$. Usually $H(\omega)$ is decomposed into two separate real functions:

amplitude spectrum: $A(\omega) = |H(\omega)| = \sqrt{\operatorname{Re}(H(\omega))^2 + \operatorname{Im}(H(\omega))^2}$

phase spectrum: $\phi(\omega) = \tan^{-1} \left(\frac{\operatorname{Im}(H(\omega))}{\operatorname{Re}(H(\omega))} \right)$

The inverse Fourier transform expresses the construction of $h(t)$ as a superposition of complex sinusoids:

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{i\omega t} d\omega$$

If we wish to use cyclical frequency, f , instead of angular frequency, ω , ($\omega = 2\pi f$) the Fourier transform pair is:

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i ft} dt$$

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{2\pi i ft} df$$

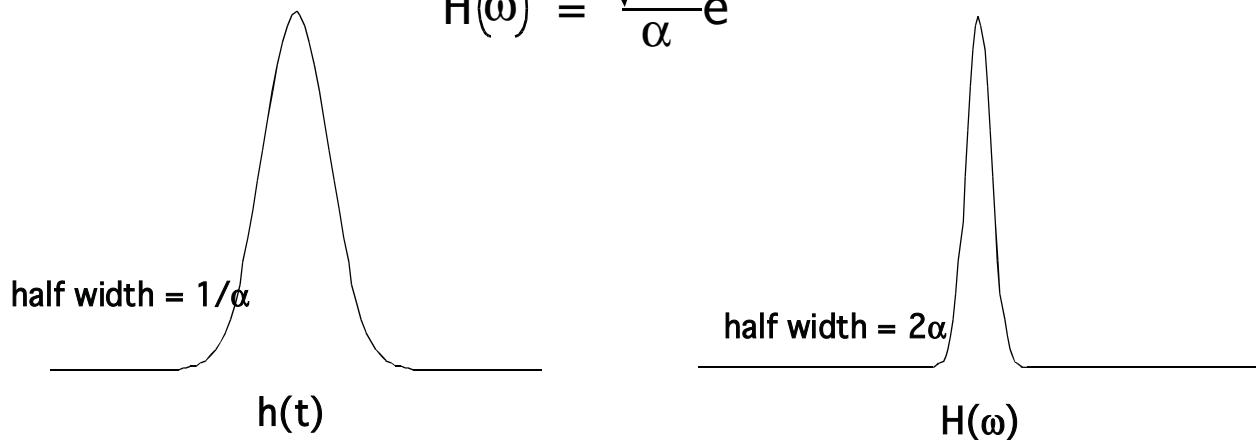
Fourier Analysis and Synthesis

As an example consider the Gaussian function:

$$h(t) = e^{-\alpha^2 t^2}$$

Using standard techniques of integral calculus, the Fourier transform of the Gaussian can be shown to be:

$$H(\omega) = \frac{\sqrt{\pi}}{\alpha} e^{-\omega^2/4\alpha}$$



Note that the half widths, as represented by their $1/e$ points are inversely proportional. In fact:

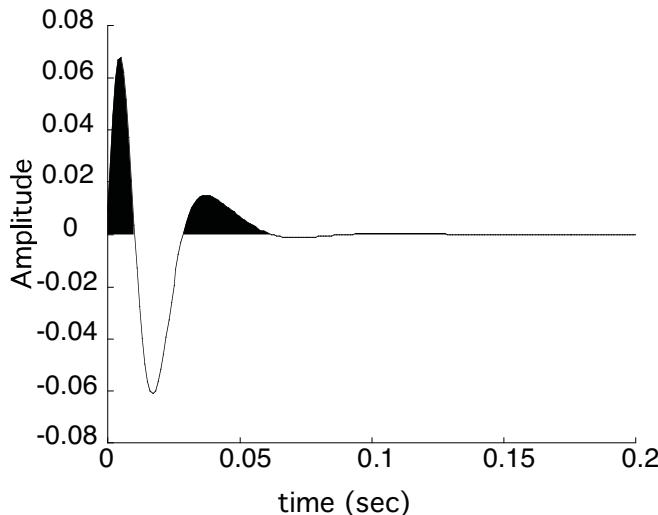
$$\Delta t \Delta \omega = (\alpha^{-1})(2\alpha) = 2$$

This is an example of a general property which says that the "width" of a time domain function is inversely proportional to its width in frequency. It can be shown, given a suitable measure of width, that:

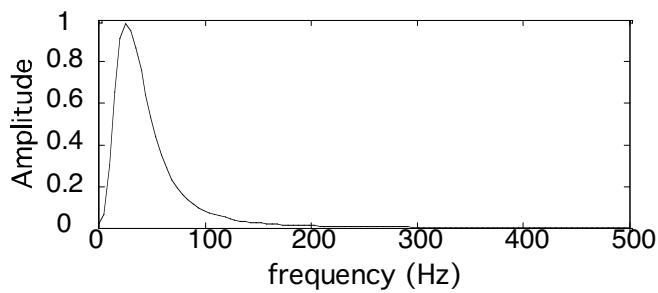
$$(\text{width in time})(\text{width in frequency}) \geq \text{a constant}$$

Bracewell (1978, The Fourier Transform and its Applications) shows the constant to be $1/2$ and that the equality holds for the Gaussian.

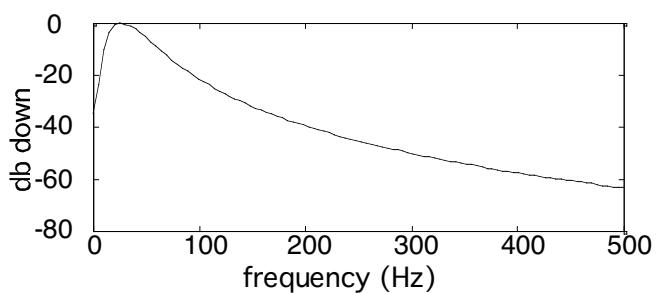
Fourier Analysis Example



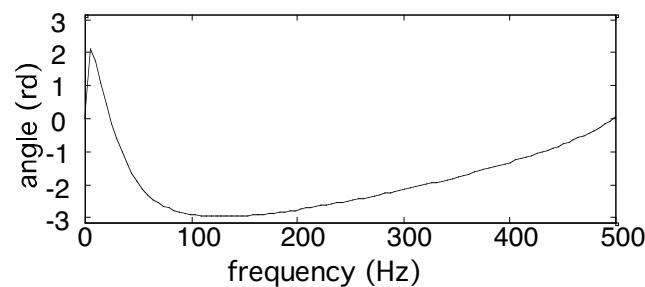
Here is a minimum phase wavelet constructed with a .001 sec sample rate and a 30 Hz dominant frequency.



This is the "amplitude spectrum of the wavelet displayed with a linear vertical scale. Note that the frequency axis stops at 500 Hz which is $1/(2 \cdot 0.001 \text{ sec})$.



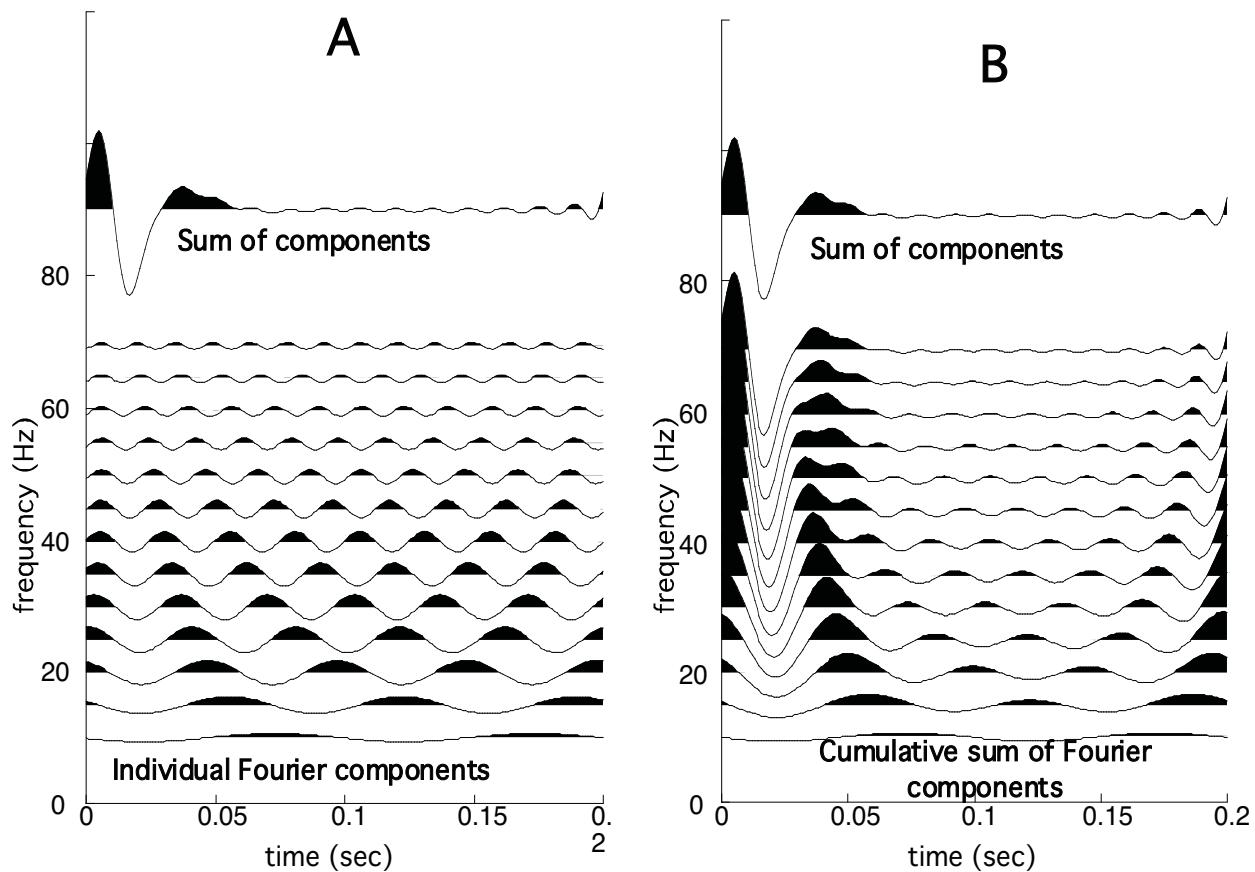
Here the amplitude spectrum is displayed with a decibel vertical scale:
$$\text{db} = 20 \cdot \log_{10}(A(f)/A_{\max})$$



This is the phase spectrum. Note that the vertical scale is in radians.

At this point, Fourier analysis may look like an exercise in graph making; however, its utility will become clear on the next page.

Fourier Analysis Example



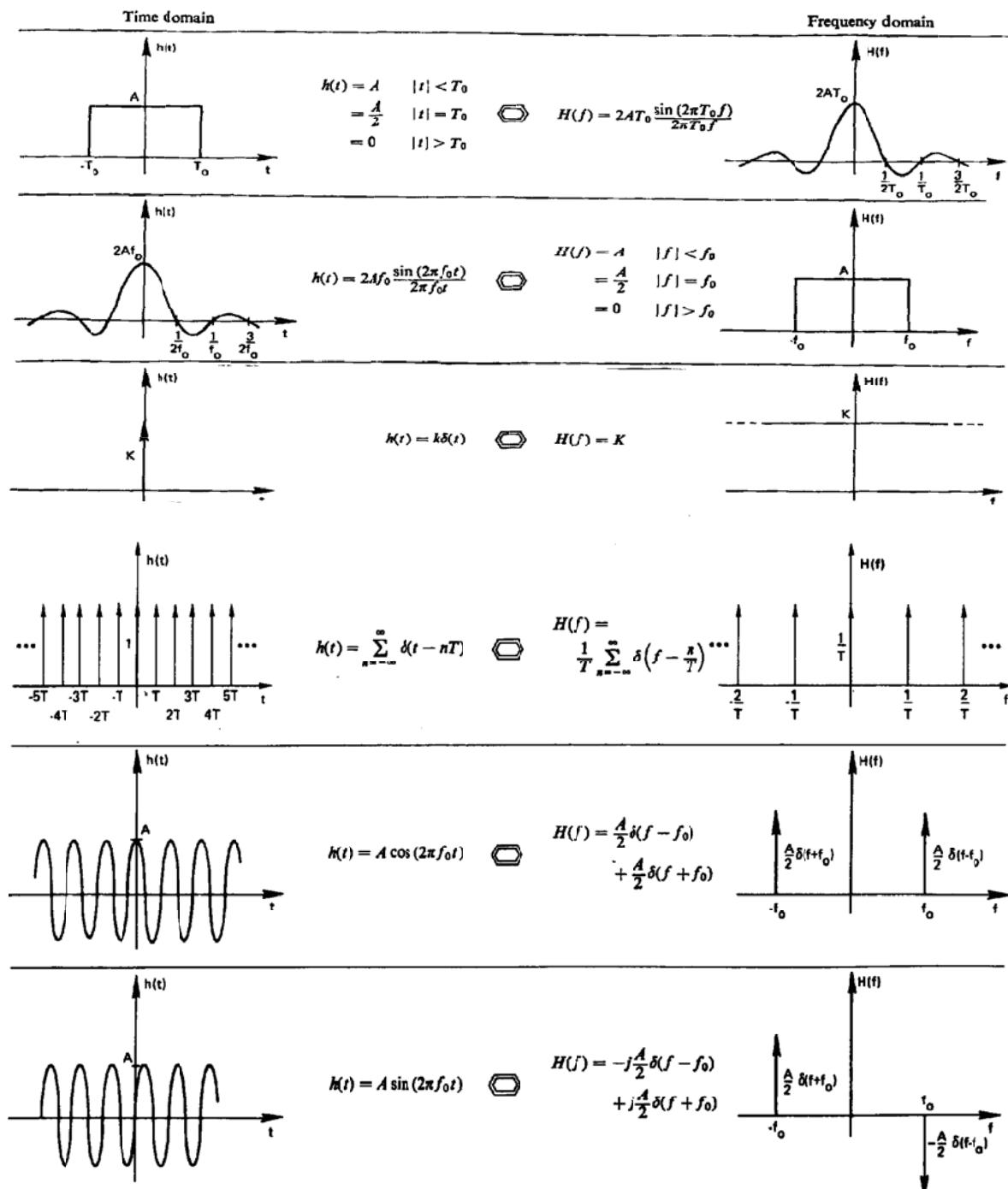
Here we see two equivalent ways of viewing the Fourier transform information on the previous page. In A, the individual Fourier components are shown from 10 to 70 Hz, properly scaled for their amplitude and phase. The sum of all 13 components yields the wavelet at the top which is quite similar to the true wavelet shown on the previous page. Adding in the remaining frequency components (0->10 Hz and 70->500 Hz) will reconstruct the wavelet exactly. The figure on the right contains the same information except that each trace is the sum of the frequency components between its frequency and 10 Hz. This gives a good illustration of how the wavelet takes form as its spectrum is summed.

Fourier Transform Pairs

The table below is reproduced from:

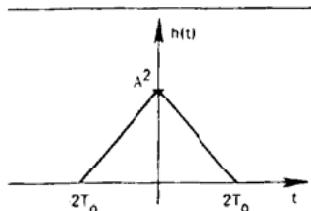
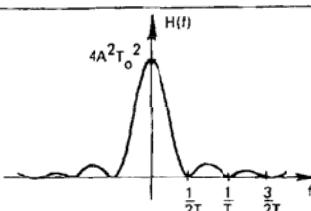
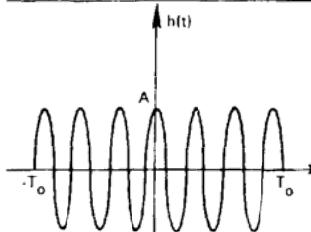
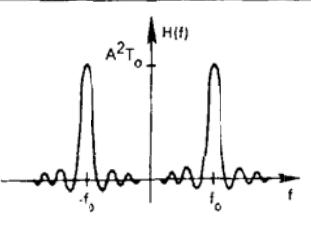
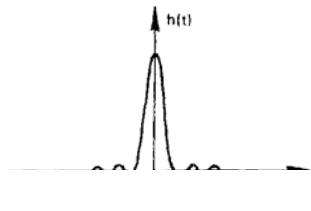
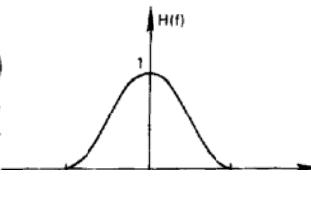
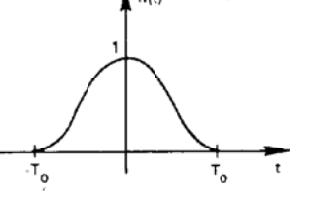
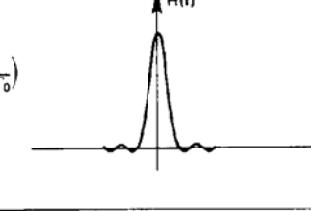
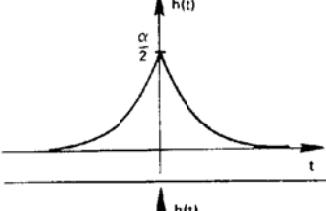
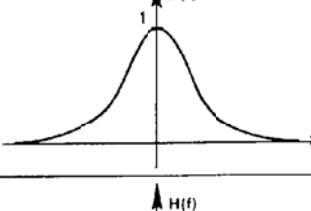
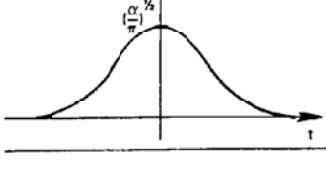
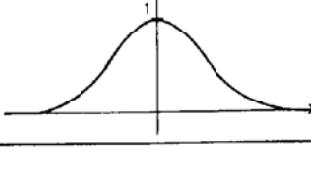
Brigham, E.O., 1974, The Fast Fourier Transform, Prentice Hall

Note: It is a remarkable fact that no signal can have finite length (i.e. compact support) in both the time and frequency domains.



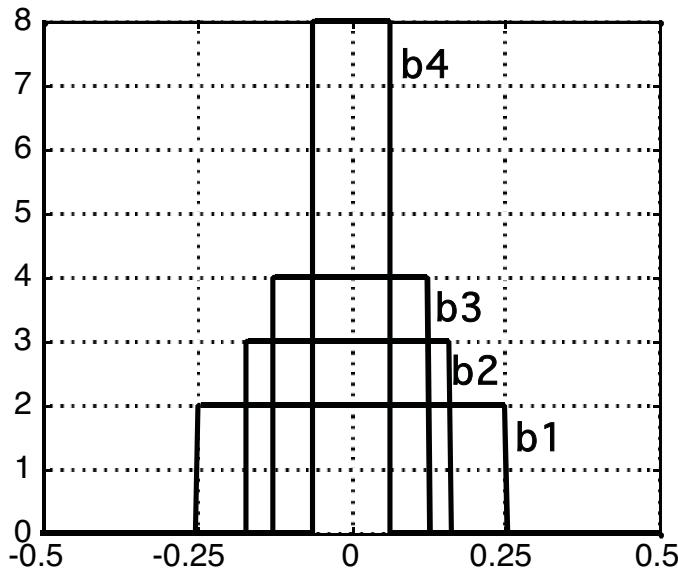
Fourier Transform Pairs

The table below is reproduced from:
 Brigham, E.O., 1974, The Fast Fourier Transform, Prentice Hall

Time domain		Frequency domain
	$h(t) = \begin{cases} -\frac{A^2}{2T_0}t + A^2 & t < 2T_0 \\ 0 & t > 2T_0 \end{cases}$	 $H(f) = A^2 \frac{\sin^2(2\pi T_0 f)}{(\pi f)^2}$
	$h(t) = \begin{cases} A \cos(2\pi f_0 t) & t < T_0 \\ 0 & t > T_0 \end{cases}$	 $H(f) = A^2 T_0 [Q(f + f_0) + Q(f - f_0)]$ $Q(f) = \frac{\sin(2\pi T_0 f)}{2\pi T_0 f}$
	$h(t) = \frac{1}{2} q(t) + \frac{1}{4} q(t + \frac{1}{2f_c}) + \frac{1}{4} q(t - \frac{1}{2f_c})$ $q(t) = \frac{\sin(2\pi f_c t)}{2\pi f_c t}$	 $H(f) = \frac{1}{2} + \frac{1}{2} \cos\left(\frac{\pi f}{f_c}\right)$ $= 0 \quad f > f_c$
	$h(t) = \begin{cases} \frac{1}{2} + \frac{1}{2} \cos\left(\frac{\pi t}{T_0}\right) & t \leq T_0 \\ 0 & t > T_0 \end{cases}$	 $H(f) = \frac{1}{2} Q(f) + \frac{1}{4} [Q\left(f + \frac{1}{2T_0}\right) + Q\left(f - \frac{1}{2T_0}\right)]$ $Q(f) = \frac{\sin(2\pi T_0 f)}{\pi f}$
	$h(t) = \frac{1}{2} \alpha \exp(-\alpha t)$	 $H(f) = \frac{\alpha^2}{\alpha^2 + 4\pi^2 f^2}$
	$h(t) = \left(\frac{\alpha}{\pi}\right)^{1/2} \exp(-\alpha t^2)$	 $H(f) = \exp\left(-\frac{\pi^2 f^2}{\alpha}\right)$

The Dirac Delta Function

The Dirac delta function was invented by P.A.M. Dirac to handle problems in the development of quantum mechanics. Since then, its unique ability to represent a "unit spike" in the continuous function domain. It can be defined as the limiting form of a sharply peaked function whose maximum proceeds to infinity as its width shrinks to zero in such a way that its area remains unity.



A series of boxcars with unit area converges in the limit to the delta function:

$$b_{\infty} = \delta(t)$$

It can be thought of as:

$$\delta(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases}$$

The most important property of the delta function is its behavior under integration. If $f(t)$ is any function, then:

$$\int_a^b f(t)\delta(t-t_0)dt = \begin{cases} f(t_0), & \text{if } a < t_0 < b \\ 0, & \text{otherwise} \end{cases}$$

This is known as the sifting property of the delta function.

The Dirac Delta Function

Consider the Fourier transform of the delta function:

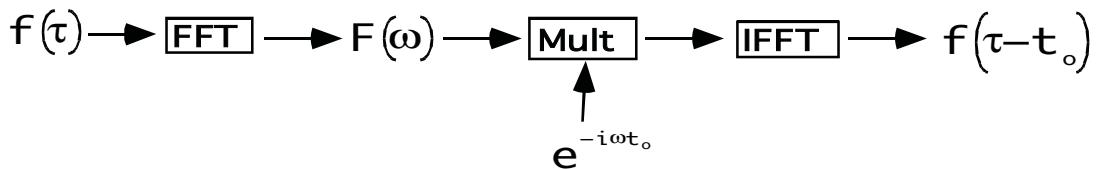
$$\int_{-\infty}^{\infty} \delta(t-t_0) e^{-i\omega t} dt = e^{-i\omega t_0}$$

Thus it has a constant, unit amplitude spectrum (also known as a "white" spectrum) and linear phase.

Consider the action of the delta function under convolution:

$$\int_{-\infty}^{\infty} \delta(t-t_0) f(\tau-t) dt = f(\tau-t_0)$$

Thus the delta function shifts $f(t)$ to place its origin at the location where the argument of the delta function vanishes. This is called a "static shift" in seismic data processing. Since convolution can be done in the Fourier domain by multiplication of transforms, we can conclude that a static shift can be done by:



That is, a static shift is equivalent to a linear phase shift. Finally, if we inverse Fourier transform the equation at the top of the page, we end up with a definition of the delta function in terms of its Fourier components:

$$\delta(\tau - t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(\tau-t_0)} d\omega = \int_{-\infty}^{\infty} e^{2\pi i f(\tau-t_0)} df$$

Thus the delta function has unit amplitude spectrum and a phase spectrum that is linear in frequency and with slope proportional to the time shift.

The Convolution Theorem

Consider the continuous convolution of f and g :

$$h(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \quad (1)$$

We can represent f and g in terms of their spectra as:

$$f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{i\omega\tau}d\omega \text{ and } g(t-\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega)e^{i\omega(t-\tau)}d\omega$$

Substituting these into (1):

$$h(t) = \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{i\omega\tau}d\omega \right] \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega)e^{i\omega(t-\tau)}d\omega \right] d\tau$$

Interchanging the order of integration

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)G(\omega) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\omega-\omega')\tau} d\tau \right] e^{i\omega t} d\omega d\omega'$$

The term in [] is the Dirac delta function.

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)G(\omega)\delta(\omega-\omega')e^{i\omega t} d\omega d\omega'$$

The delta function collapses one of the frequency integrals

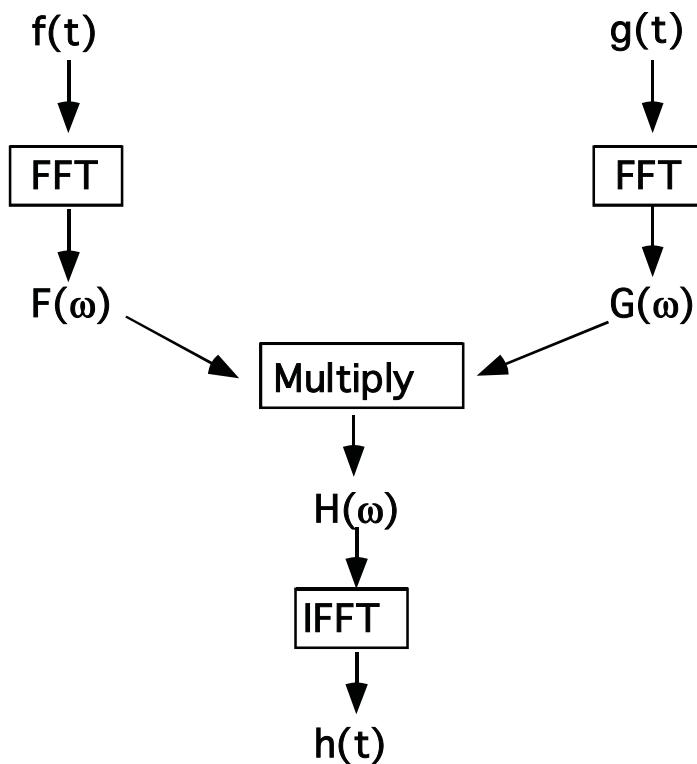
$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)G(\omega)e^{i\omega t} d\omega$$

Here we have $h(t)$ represented as the inverse Fourier transform of "something". By inference, that something must be the Fourier transform of h . Thus:

$$H(\omega) = F(\omega)G(\omega)$$

The Convolution Theorem

The result we have just derived is one of the most fundamental and important in all of signal processing. It tells us that we can convolve two signals by multiplying their spectra and inverse Fourier transforming the result. The reason that this is important is that there is an extremely fast algorithm for performing the digital Fourier transform called the fast Fourier transform (FFT). Using the FFT a convolution can be done by:



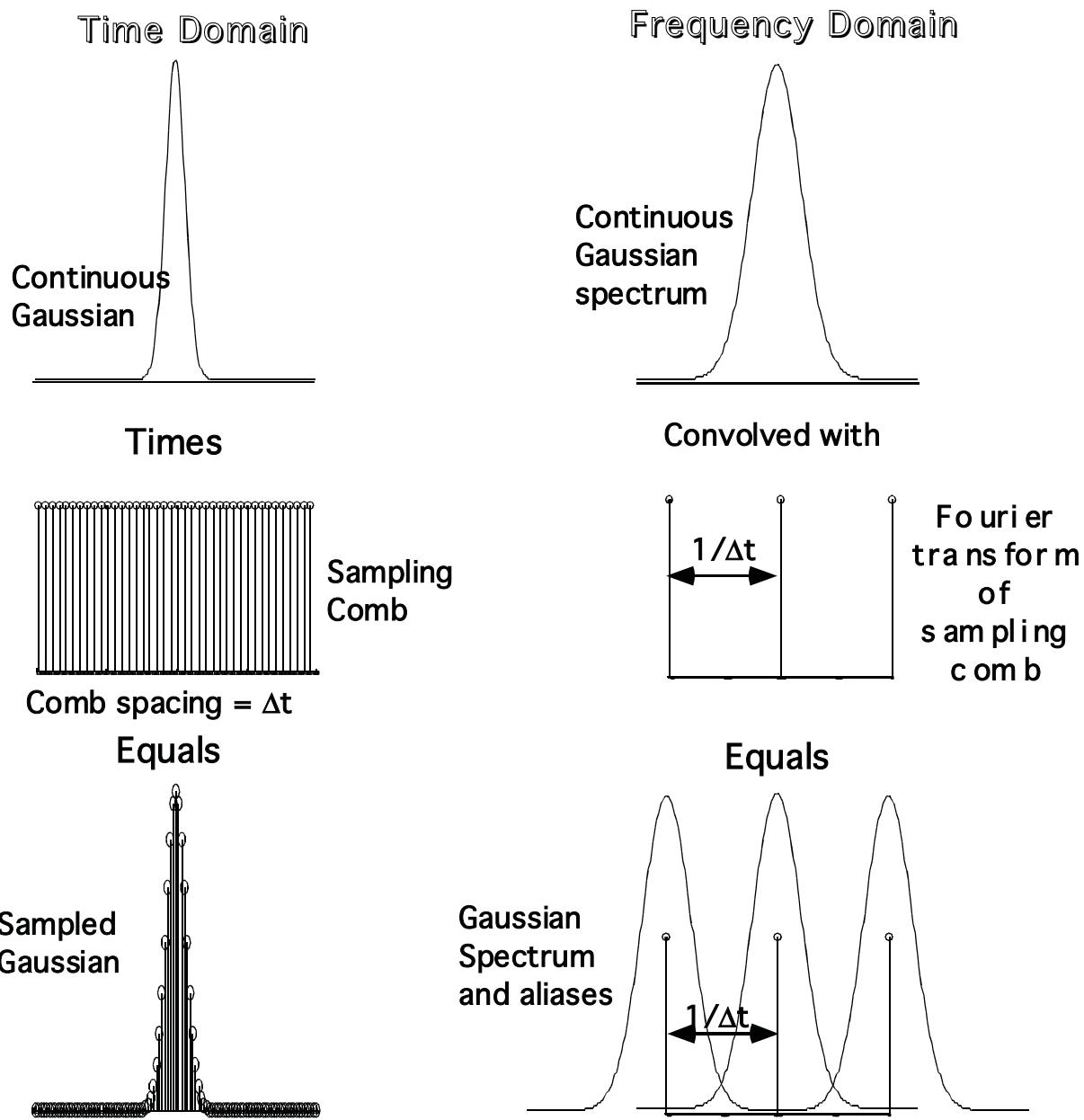
Note that multiplying complex spectra is:

$$\begin{aligned} H(\omega) &= F(\omega)G(\omega) = A_F(\omega)e^{i\phi_F(\omega)}A_G(\omega)e^{i\phi_G(\omega)} \\ &= A_F(\omega)A_G(\omega)e^{i(\phi_F(\omega)+\phi_G(\omega))} \end{aligned}$$

That is we can view it as multiplying the amplitude spectra and adding the phase spectra.

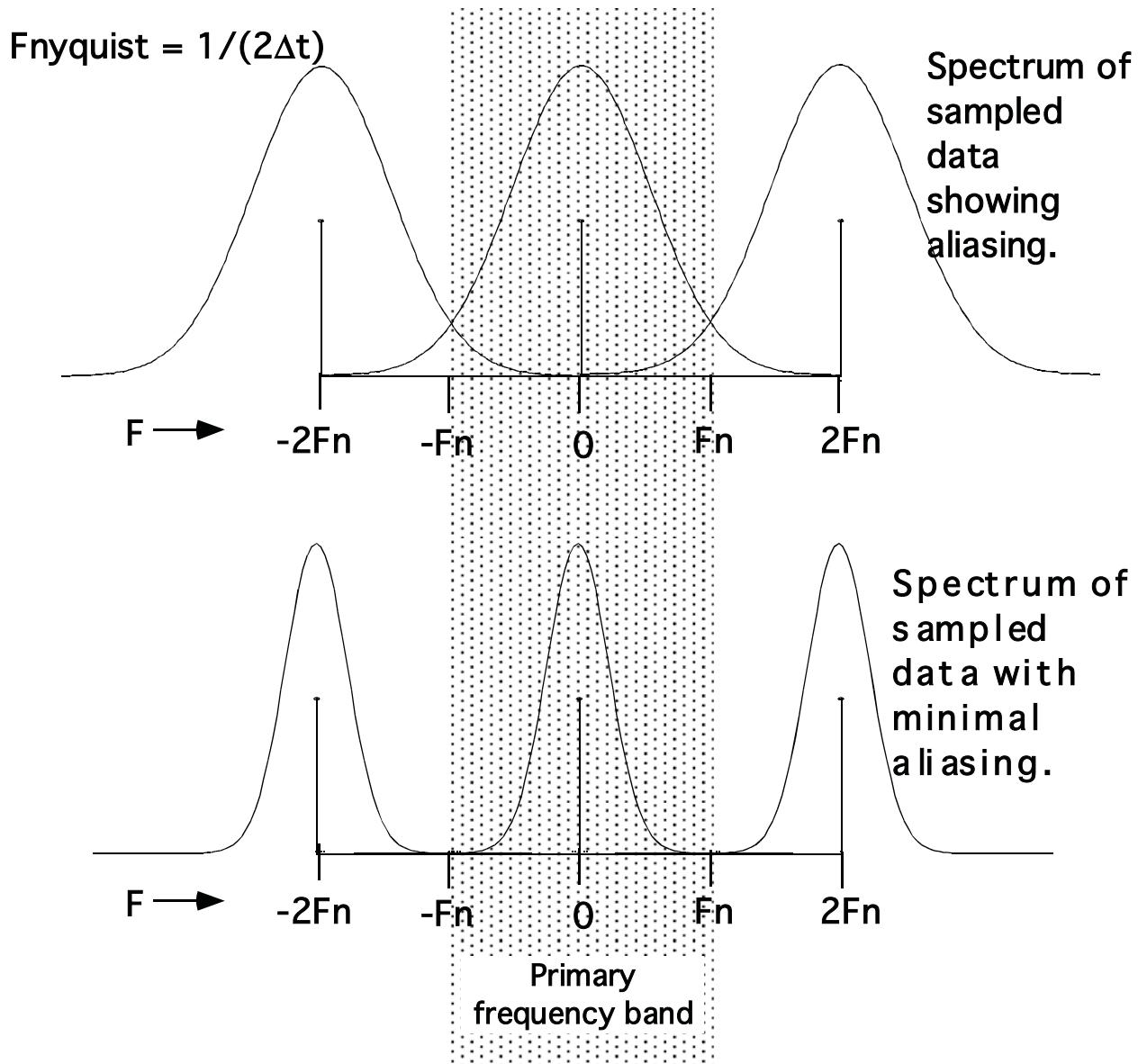
Sampling

The analytic analysis of continuous signals is most useful for gaining a conceptual understanding of signal processing. In actual practice; however, the vast majority of work is done with discretely sampled functions. The process of sampling a continuous function in time can be viewed as a multiplication by a sampling comb.



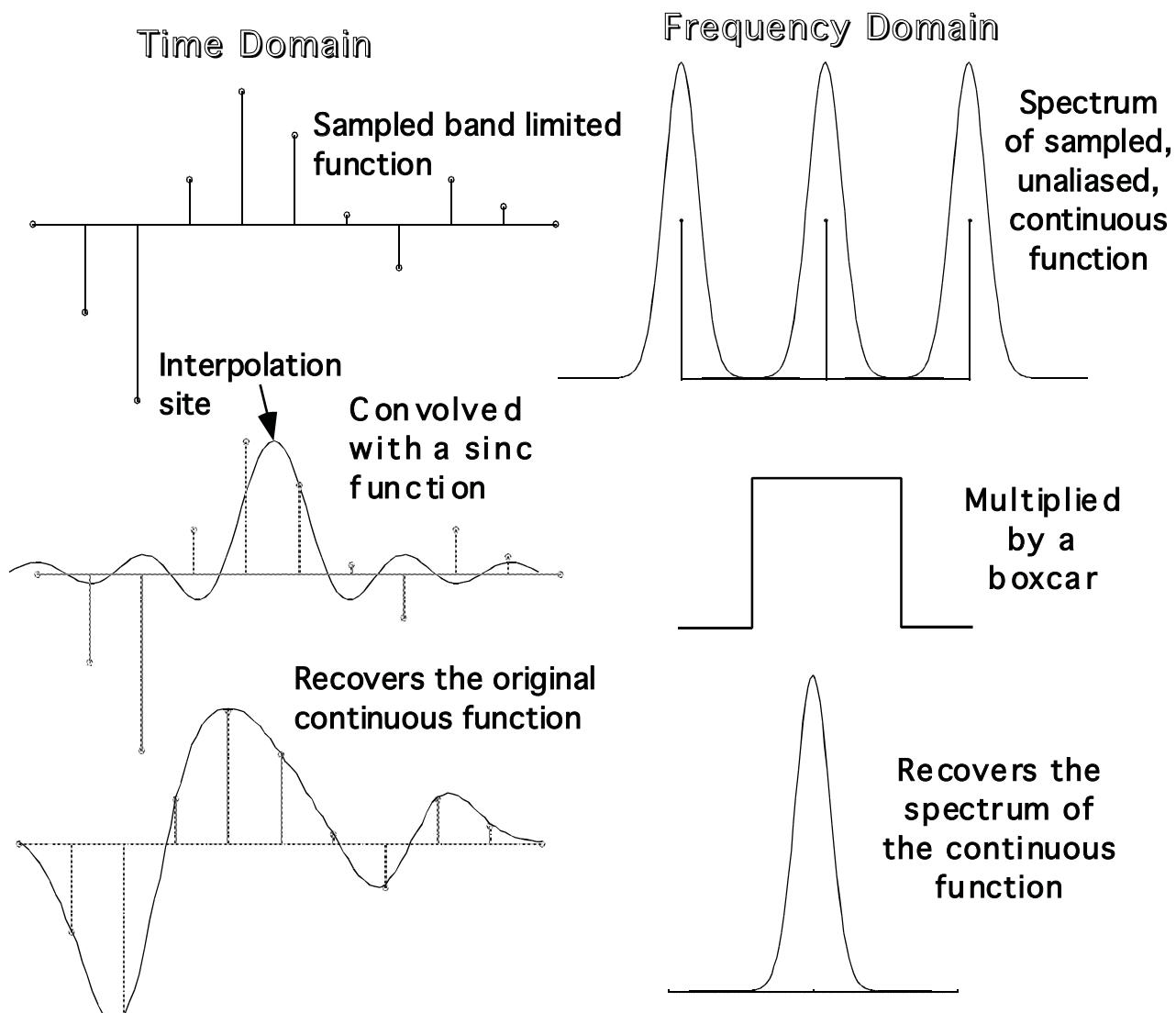
Sampling

So we have seen that sampling in the time domain causes the replication of the continuous spectrum in the frequency domain. The spacing between these spectral aliases is $1/\Delta t$ and it is customary to restrict our attention to the primary frequency band lying between $-1/(2\Delta t)$ and $1/(2\Delta t)$. The frequency $F_n = 1/(2\Delta t)$ is called the Nyquist frequency and is the limiting frequency of the sampled data.



Sampling

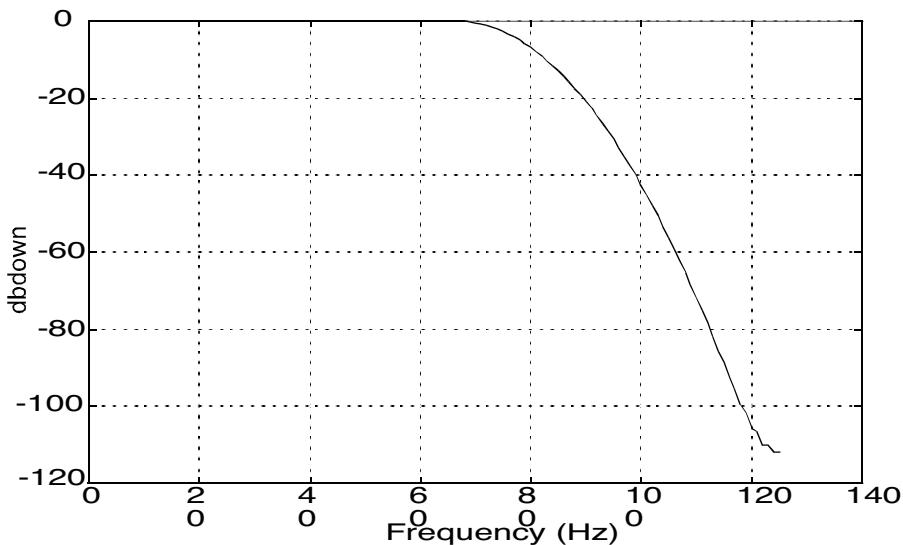
The unaliased sampling of any continuous signal requires that the signal have its power restricted to a frequency band: $-f_{\max} < f < f_{\max}$. Such signals are said to be band limited. A band limited signal can be digitally sampled, without aliasing, with a sample size of $\Delta t = 1/(2f_{\max})$. It is a fundamental theorem (The Sampling Theorem, Papoulis, Signal Analysis, p 141, 1984) that such a bandlimited, continuous, signal can be exactly recovered from its digital samples by a process known as "sinc function interpolation".



Sampling

In order to minimize aliasing, raw analog seismic data is passed through an analog antialias filter prior to digitization. A typical antialias filter has an amplitude spectrum which begins to roll off at 50% to 60% of fnyquist and reaches very large attenuation (>60db) at fnyquist.

Here is the spectrum of an antialias filter for use prior to sampling at .004 sec.



Rule of thumb: Sample your data such that the expected signal frequencies are less than half fnyquist.

Common sampling rates and their Nyquist frequencies	sample rate	Nyquist
	.008 s	62.5 Hz
	.004 s	125 Hz
	.002 s	250 Hz
	.001 s	500 Hz

Aliasing is also a possibility when resampling seismic data. If the new sample interval is more coarse than the old, then an antialias filter should be applied.

The Discrete Fourier Transform

The great utility of the continuous Fourier transform to decompose functions into fundamental complex sinusoids can be applied directly to discretely sampled time domain functions. Consider a function $h(t)$ which is zero everywhere except at N times defined by $t=k\Delta t$, $k=0,1,2 \dots N-1$, where it takes the values h_k . This function can be written with the dirac delta function as:

$$h(t) = \sum_{k=0}^{N-1} h_k \delta(t - k\Delta t)$$

If we now take the Fourier transform of $h(t)$ we have:

$$H(\omega) = \int_{-\infty}^{\infty} \left[\sum_{k=0}^{N-1} h_k \delta(t - k\Delta t) \right] e^{-i\omega t} dt = \sum_{k=0}^{N-1} h_k \int_{-\infty}^{\infty} \delta(t - k\Delta t) e^{-i\omega t} dt$$

$$H(\omega) = \sum_{k=0}^{N-1} h_k e^{-i\omega k\Delta t}$$

Here we have an analytic expression for the Fourier transform of the h_k samples which is defined for all ω . We have already seen that the phenomenon of aliasing limits the usable frequency band to $-\pi/\Delta t \rightarrow +\pi/\Delta t$. Furthermore, linear algebra tells us that N frequencies in this band should suffice to determine the N h_k . So we are lead to consider sampling the frequency domain at $\omega_v = 2\pi v/(N\Delta t)$, $v = 0, 1, 2 \dots N-1$.

$$H_v = \sum_{k=0}^{N-1} h_k e^{-i2\pi v k / N}$$

The Discrete Fourier Transform

Discrete exponentials have a well known orthogonality property such that:

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{-2\pi v(k-\bar{k})/N} = \delta_{k\bar{k}} = \begin{cases} N, & \text{if } k = \bar{k} \\ 0, & \text{otherwise} \end{cases}$$

Using this, it is not difficult to show that the h_k samples can be recovered from the H_v by:

$$h_k = \frac{1}{N} \sum_{v=1}^{N-1} H_v e^{i2\pi v k / N} \quad \text{Inverse DFT}$$

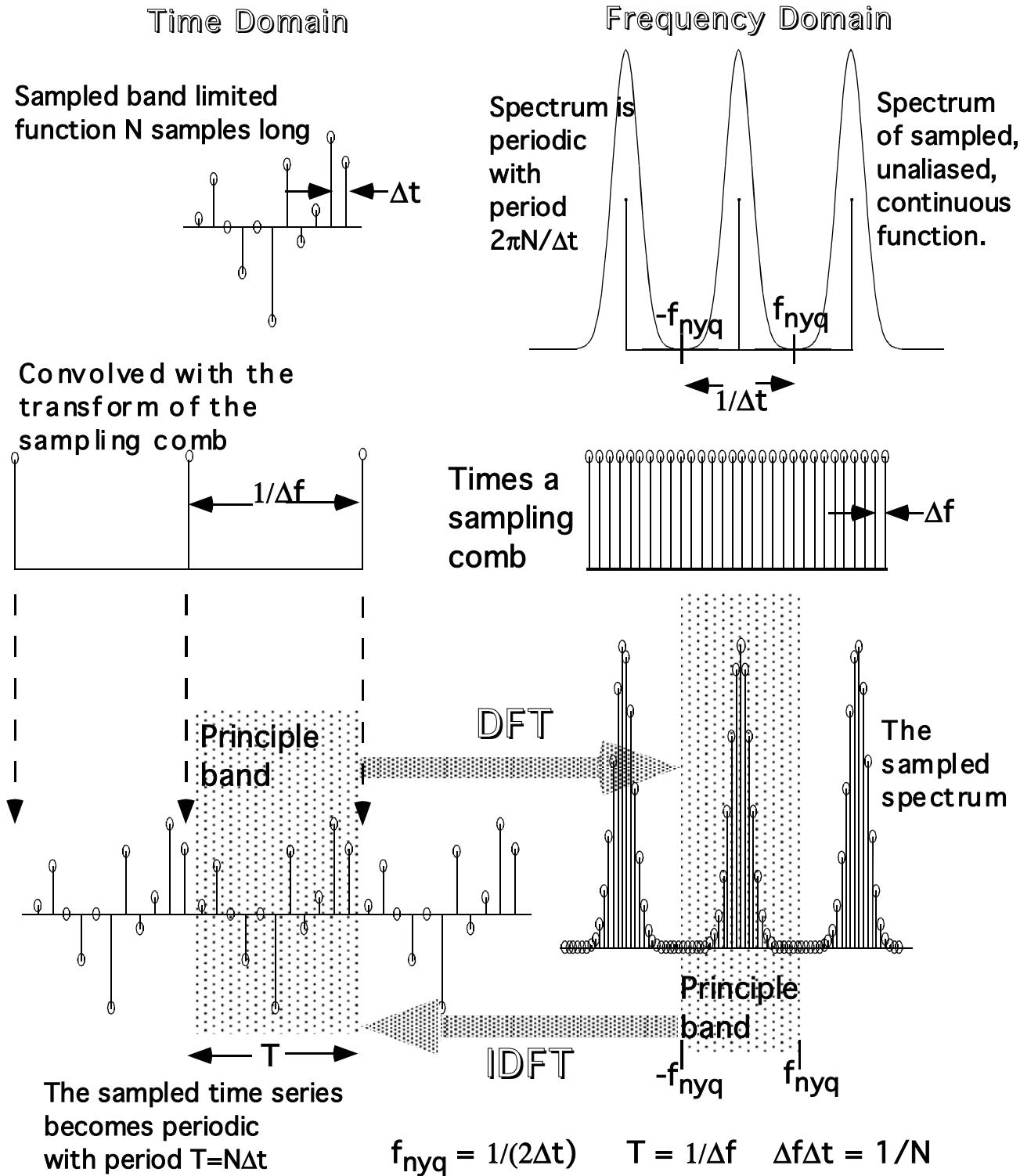
This result together with:

$$H_v = \sum_{k=0}^{N-1} h_k e^{-i2\pi v k / N} \quad \text{Forward DFT}$$

form the discrete Fourier transform pair. They are the direct analog to the continuous Fourier transform relations. Like the FT, the DFT is complete in that the h_k are exactly recoverable from their spectrum, the H_v .

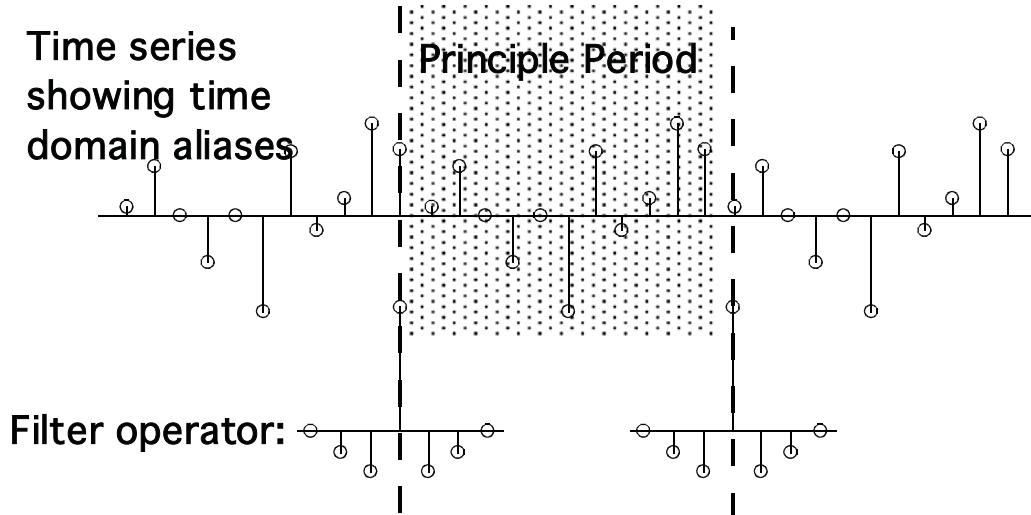
The Discrete Fourier Transform

Here is a pictorial representation of the development of the DFT from the continuous case:

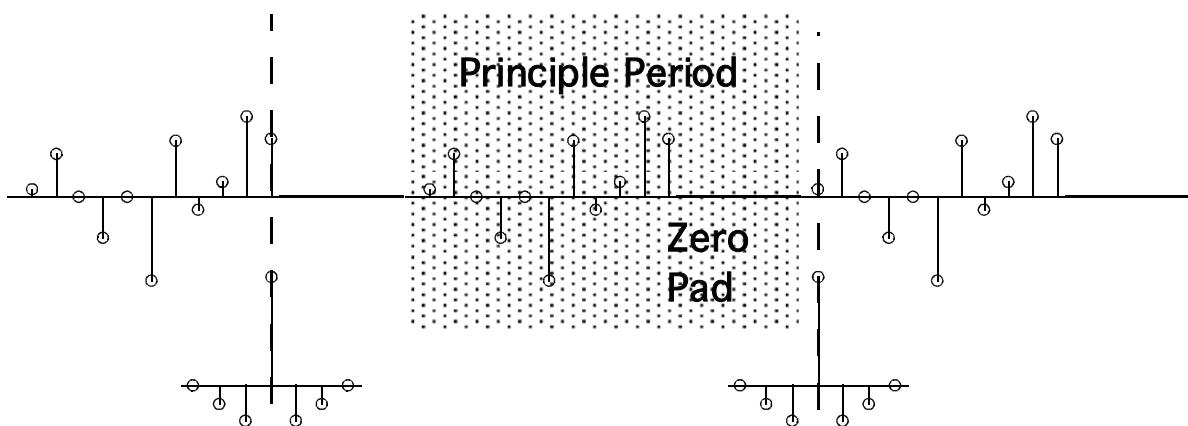


The Discrete Fourier Transform

The sampling of the spectrum of a discrete time series causes that series to become periodic with period $T = N\Delta t$. This has signal processing consequences that are apparent when we consider applying a filter with the DFT and the corresponding convolution.

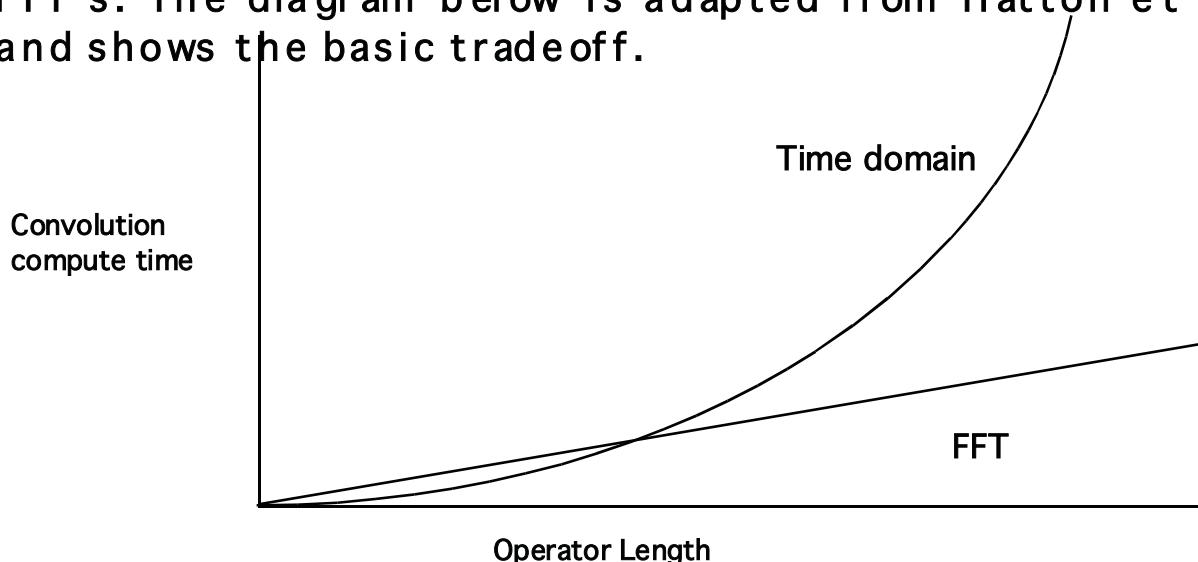


The convolution operation that duplicates multiplication with the DFT is called circular convolution. Note that the filter operator placed on the last sample of the principle period appears to "wrap around" and affect the first sample. To avoid this problem, it is common to pad the time series with a length of zeros chosen with the length of the filter operator in mind.



The Fast Fourier Transform

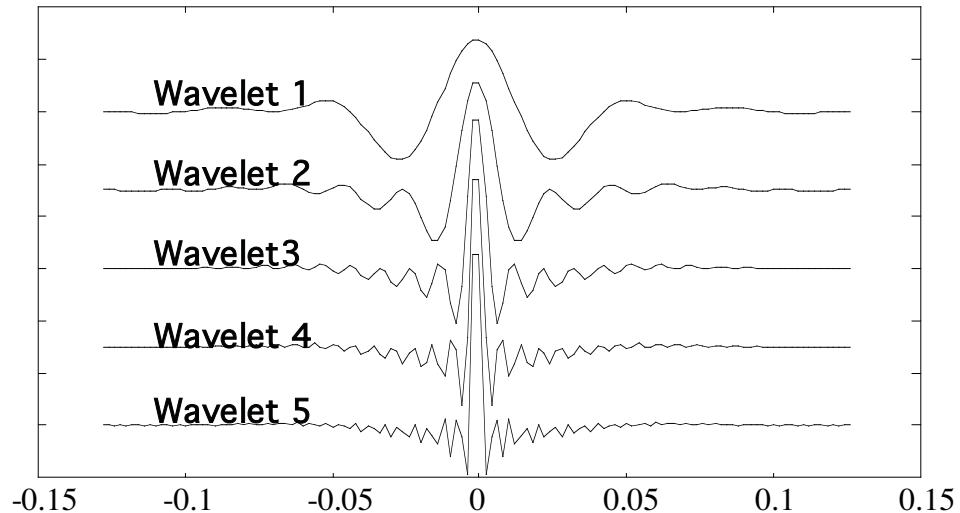
The fast Fourier transform (FFT) is nothing more than a clever way of calculating the DFT which gets impressive performance results. The convolution of an N length operator in the time domain requires on the order of N^2 floating point operations. The same computation in the frequency domain with the FFT requires roughly $N \log(N)$ operations. However, we must be careful with this statement because, generally, the two N's are not the same. This is because the FFT algorithm requires that the time series length be a "magic number" which is usually a power of 2. (Also the two time series being convolved must be the same length.) This is achieved by attaching a zero pad to the time series. Thus if N is the length of the time domain operator and if N2 is the first power of 2 greater than N, then we must compare N2 to $N2 \log(N2)$. (Often even this is not enough because the zero pad must be long enough to avoid operator wrap around.) The bottom line is that short operators (less than ~64 points) are often applied faster with convolution while long operators are MUCH faster with FFT's. The diagram below is adapted from Hatton et al. and shows the basic tradeoff.



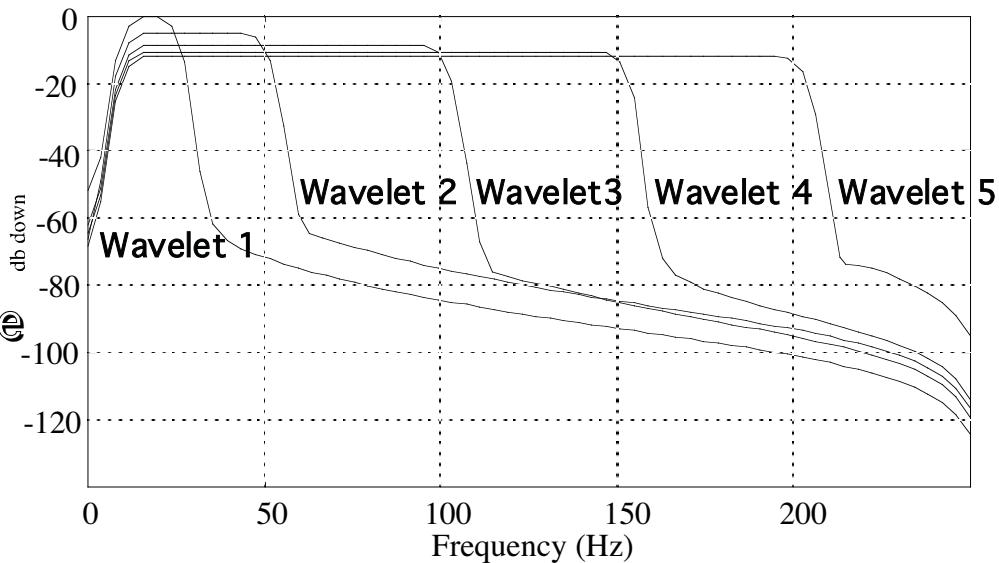
Filtering

We have seen that convolution with a waveform suppresses and possibly phase shifts some frequencies relative to others. This filtering action is often exploited to enhance signal and suppress noise. Here we see a comparison of five different zero phase filters in both the time and frequency domains. The inverse relationship between temporal width and frequency bandwidth is readily apparent.

Five
Generic
Wavelets



Their
Fourier
Amplitude
Spectra



The Z Transform

The periodicity or circularity inherent in both time and frequency is nicely captured by a powerful methodology known as the Z transform. Consider the time series, [1 -.5 -.3 0 .1 0], where it is assumed to start at $t=0$ and increment by Δt . We represent this series in the Z domain by a polynomial in z :

$$\begin{aligned} H(z) &= 1z^0 - .5z^1 - .3z^2 + 0z^3 + .1z^4 \\ &= 1 - .5z^1 - .3z^2 + .1z^4 \end{aligned}$$

So we see that the exponent of z gives the sample number and hence determines the sample time ($n\Delta t$). Note also the following:

- Negative times correspond to negative exponents of z
- Multiplication by z^n delays the time series by n samples if n is positive and advances it by n samples for negative n .

The great utility of the Z transform lies in its ability to represent discrete convolution and the DFT as operations with polynomials. It is not difficult to show that the convolution of two time series, f and g , can be realized by simply multiplying their Z transforms and reading off the result. (See Waters (p 133) for a proof.)

$$F(z) = f_0 + f_1 z^1 + f_2 z^2 + \dots \quad G(z) = g_0 + g_1 z^1 + g_2 z^2 + \dots$$

$$\begin{aligned} H(z) &= F(z)G(z) = \\ &f_0 g_0 + (f_0 g_1 + g_0 f_1) z^1 + (f_0 g_2 + f_1 g_1 + g_0 f_2) z^2 + \dots \\ h &= f \bullet g \end{aligned}$$

The Z Transform

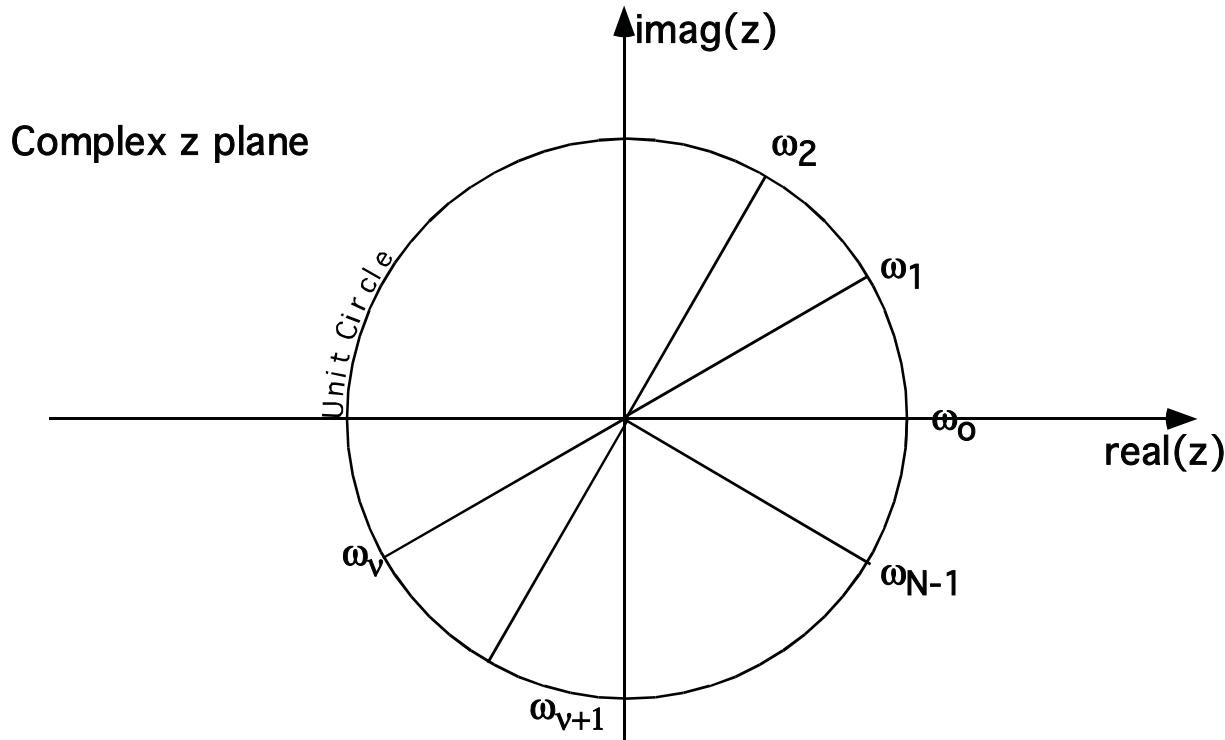
The fact that convolution is done by multiplication of Z transforms is reminiscent of the Fourier transform. In fact, if we let $z = e^{-i\omega\Delta t}$ then the Z transform becomes:

$$G(z) = \sum_{k=0}^{N-1} g_k z^k \quad G(\omega) = \sum_{k=0}^{N-1} g_k e^{-i\omega k \Delta t}$$

As with the DFT, if we now consider only discrete frequencies $\omega_v = 2\pi v / (N\Delta t)$, $v = 0, 1, 2, \dots, N-1$, then we see that the Z transform, with $z = e^{-i\omega\Delta t}$, is precisely the DFT.

$$G_v = \sum_{k=0}^{N-1} g_k e^{-i2\pi v k / N}$$

The Z transform is more general than the DFT since z can be any complex number. In fact the DFT amounts to evaluating the Z transform at N discrete locations around the unit circle in the complex z plane.



The Z Transform

Consider the elemental couplet $F(z) = 1 - az$. Now if we convolve $F(z)$ with another arbitrary time series $g(z)$, then we represent this as: $H(z) = F(z)G(z)$. Suppose that only $F(z)$ and $H(z)$ are known to us and we wish to recover $G(z)$. In the z transform domain we can simply:

$$H(z) = F(z)G(z) \quad \therefore G(z) = \frac{H(z)}{F(z)}$$

So we define the inverse of any time series as:

$$F^{-1}(z) = \frac{1}{F(z)}$$

For $F(z) = 1 - az$, this gives:

$$F^{-1}(z) = \frac{1}{1-az} = 1 + az + (az)^2 + (az)^3 + \dots$$

This series, called the geometric series, is known to converge absolutely provided that $|az| < 1$. Since we are especially interested in this result evaluated on the unit circle ($|z| = 1$) then we need $|a| < 1$. It is customary to talk about the location of the "zero" of this couplet defined by:

$$1 - az_0 = 0 \Rightarrow z_0 = \frac{1}{a}$$

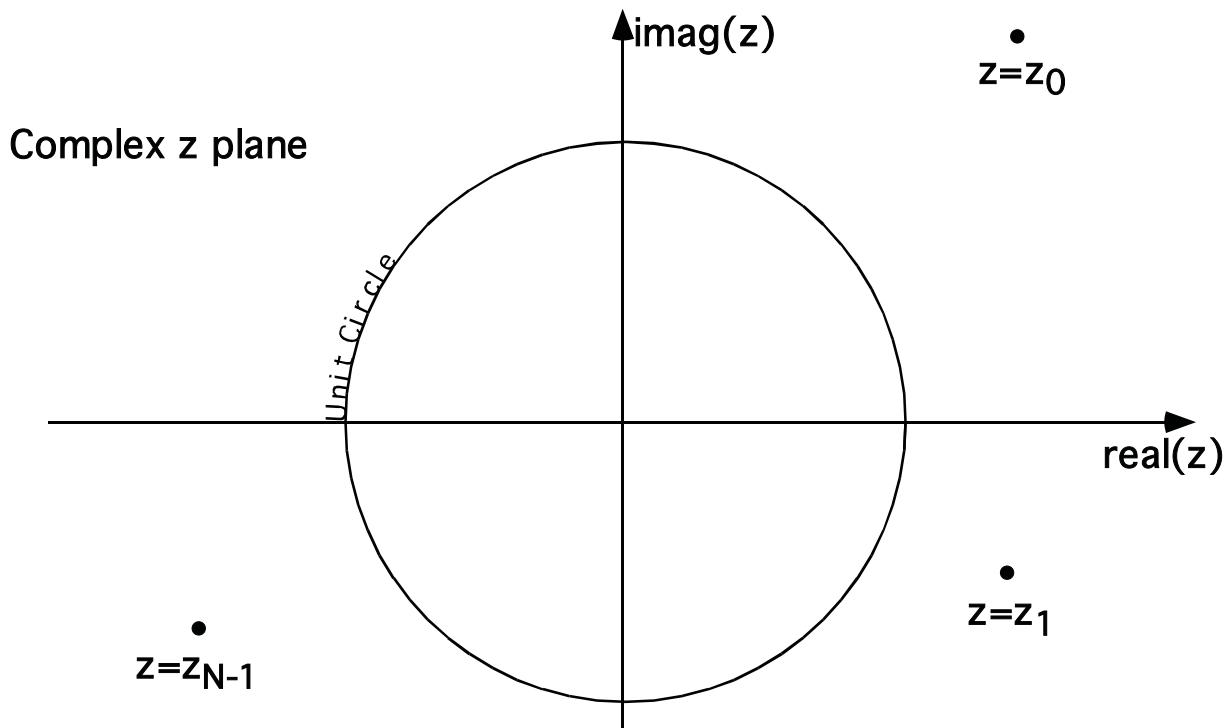
If $|a| < 1$, then we see that z_0 must lie outside the unit circle in order for the inverse to converge. Such an inverse is said to be stable (physically realizable). Note also that $F(z)$ itself is trivially stable.

The Z Transform

Any causal, stable time series with a causal, stable inverse is said to be minimum phase. Thus our elemental couplet, $1-az$, is minimum phase whenever $|a|<1$. Any more complex time series can always be factored into a set of elementary couplets.

$$G(z) = \sum_{k=0}^{N-1} g_k z^k = (z-z_0)(z-z_1)\cdots(z-z_{N-1})$$

We say that $G(z)$ is minimum phase if all its elemental couplets are minimum phase. That is equivalent to saying that all of the roots of the polynomial $G(z)$ must lie outside the unit circle in the complex z plane. If all roots lie inside the unit circle, $G(z)$ is said to be maximum phase and otherwise it is mixed phase.



A minimum phase time series has all its zeros outside the unit circle.

The Z Transform

The zeros of $F(z)$ correspond to poles for $F^{-1}(z)$. Thus, for the case of a time series whose Z transform has a denominator, we see that the stability condition requires that all poles also lie outside the unit circle. The most general time series can be written as a Z transform with both numerator and denominator such as:

$$H(z) = \frac{A(z)}{B(z)} = \frac{(z-\alpha_0)(z-\alpha_1)\dots}{(z-\beta_0)(z-\beta_1)\dots}$$

We say the corresponding time series is minimum phase if all α_i and all β_i lie outside the unit circle. The following theorem follows immediately:

The resultant of the sequential convolution of any number of minimum phase time series is also minimum phase.

Conversely:

If any time series in a sequence of convolutions is not minimum phase, then the resultant is not minimum phase.

Though these statements seem iron clad, keep in mind the unstated assumption that all these time series have the same sample rate. Thus the resampling of a time series is an operation which lies outside the scope of these theorems.

Crosscorrelation

Given two signals, r and s , the crosscorrelation provides a numerical characterization of their similarity.

The calculation of $S \otimes r$

Zero lag:

s0	s1	s2	s3	s4	s5	...
----	----	----	----	----	----	-----

Multiply aligned samples and sum:

r0	r1	r2	r3	r4	r5	...
----	----	----	----	----	----	-----

$$C_0 = s_0 r_0 + s_1 r_1 + s_2 r_2 + \dots$$

First positive lag:

s0	s1	s2	s3	s4	s5	...
----	----	----	----	----	----	-----

Multiply aligned samples and sum:

r0	r1	r2	r3	r4	r5	...
----	----	----	----	----	----	-----

$$C_1 = s_0 r_1 + s_1 r_2 + s_2 r_3 + \dots$$

First negative lag:

s0	s1	s2	s3	s4	s5	...
----	----	----	----	----	----	-----

Multiply aligned samples and sum:

r0	r1	r2	r3	r4	r5	...
----	----	----	----	----	----	-----

$$C_{-1} = s_1 r_0 + s_2 r_1 + s_3 r_2 + \dots$$

Crosscorrelation

The general form for the crosscorrelation of s and r can be written:

$$\rightarrow c_j = \sum_k s_k r_{k+j}$$

Or, for continuous signals:

$$c(\tau) = \int_{-\infty}^{\infty} s(t)r(t+\tau)dt$$

Properties of crosscorrelations:

- If either s or r is an infinite length random signal, then $c_j=0$ for all j .
- The maximum of c defines the "lag" at which s and r are most similar when aligned.
- A crosscorrelation can be computed by time reversing s and convolving. Can you prove this?
- The autocorrelation is a special case of crosscorrelation when $r=s$.

Autocorrelations

The autocorrelation, ϕ , of a signal, s , is a characterization of its self similarity. It can be computed as follows:

Zero lag:

The signal s	s0	s1	s2	s3	s4	s5	...
--------------	----	----	----	----	----	----	-----

Multiply aligned samples and sum:

A copy of s	s0	s1	s2	s3	s4	s5	...
-------------	----	----	----	----	----	----	-----

$$\phi_0 = S_o^2 + S_1^2 + S_2^2 + S_3^2 + \dots$$

First positive lag:

s0	s1	s2	s3	s4	s5	...
----	----	----	----	----	----	-----

Multiply aligned samples and sum:

s0	s1	s2	s3	s4	s5	...
----	----	----	----	----	----	-----

$$\phi_1 = S_o S_1 + S_1 S_2 + S_2 S_3 + \dots$$

Second positive lag:

s0	s1	s2	s3	s4	s5	...
----	----	----	----	----	----	-----

Multiply aligned samples and sum:

s0	s1	s2	s3	s4	s5	...
----	----	----	----	----	----	-----

$$\phi_2 = S_o S_2 + S_1 S_3 + S_2 S_4 + \dots$$

Autocorrelations

The general form for the autocorrelation of s can be written:

$$\phi_j = \sum_{k=0}^{\text{length}(s)} s_k s_{k+j}$$

Properties of the autocorrelation:

- $\phi_0 \geq \phi_j$ for all j . The zero lag is always largest.
- If s is an infinite length random sequence, then ϕ_0 gives the sum of squares of the sequence and all other ϕ_j are zero.
- The Fourier transform of the autocorrelation gives the power spectrum (squared amplitude spectrum) of the signal, s .
- The autocorrelation has no phase information.

The autocorrelation is often normalized such that $\phi_0=1$:

$$\phi_j = \frac{\sum_{k=0}^{\text{length}(s)} s_k s_{k+j}}{\sum_{k=0}^{\text{length}(s)} s_k^2}$$

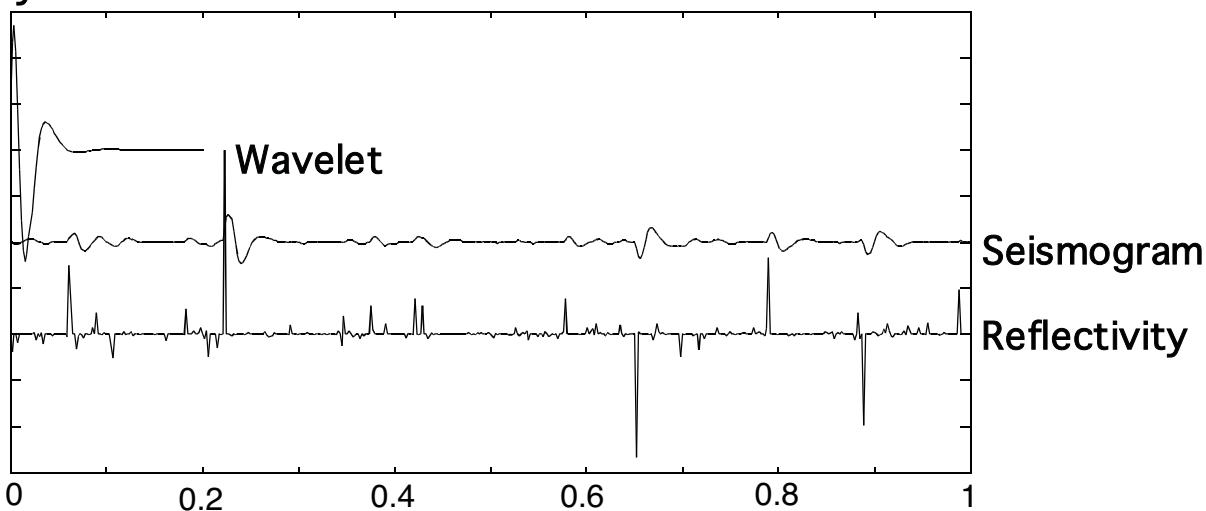
Spectral Estimation

The general problem of estimating amplitude or power spectra of an unknown signal embedded in noise or other unwanted signals is called spectral estimation. It arises in many contexts in seismic data processing but most notably in deconvolution theory. Two sample problems:

- Given a small number of lags of a possibly infinite autocorrelation, estimate the power spectrum of the underlying physical process.
- Given a small portion of a possibly infinite time series, estimate the amplitude spectrum of the underlying physical process.

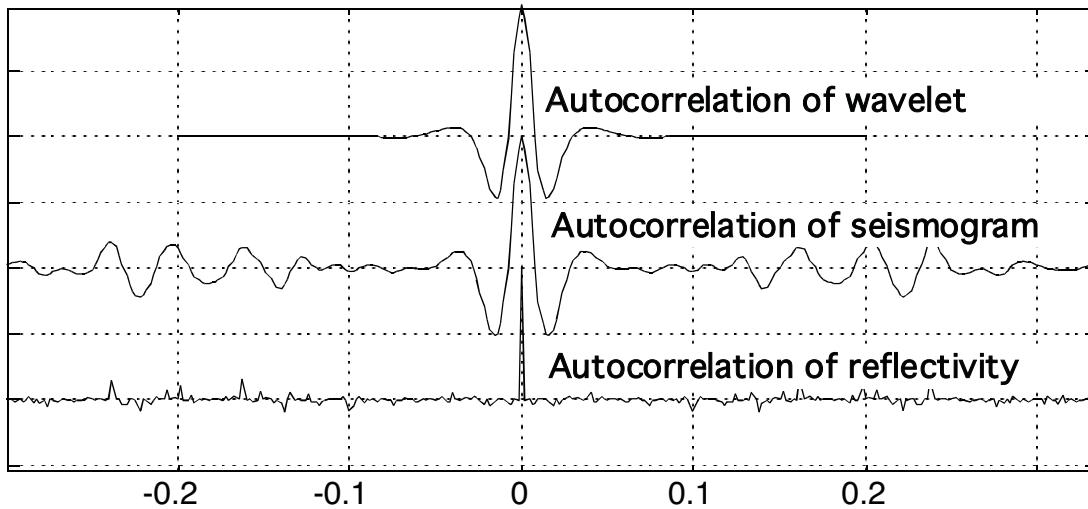
These two problems arise repeatedly, and in a variety of contexts, in seismic data processing theory. However, they are essentially similar differing only in the nature of the input: e.g an autocorrelation or a general time series.

We shall consider two approaches: the windowed DFT, and the maximum entropy spectrum (Burg spectrum). Consider the construction of an elementary seismogram by convolution:

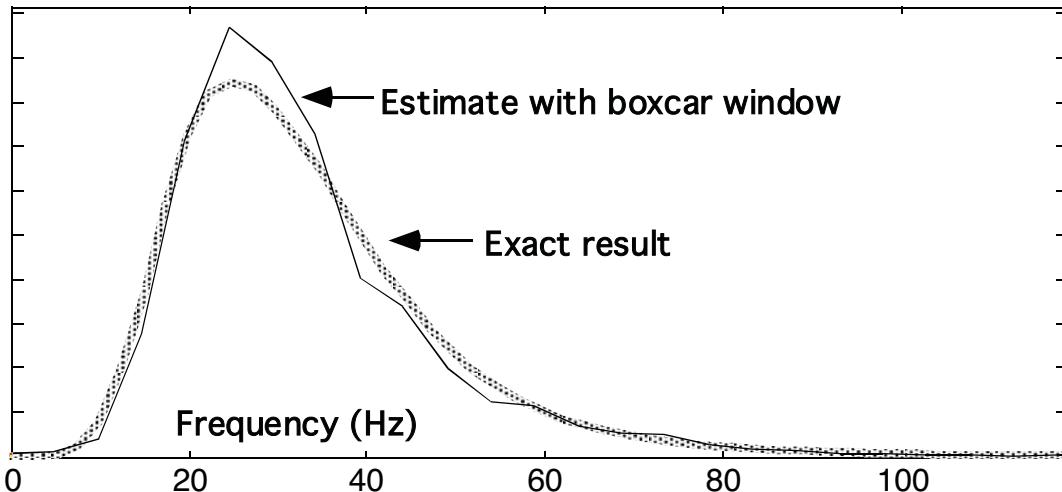


Spectral Estimation

If we compute the autocorrelations of these three functions, we obtain:

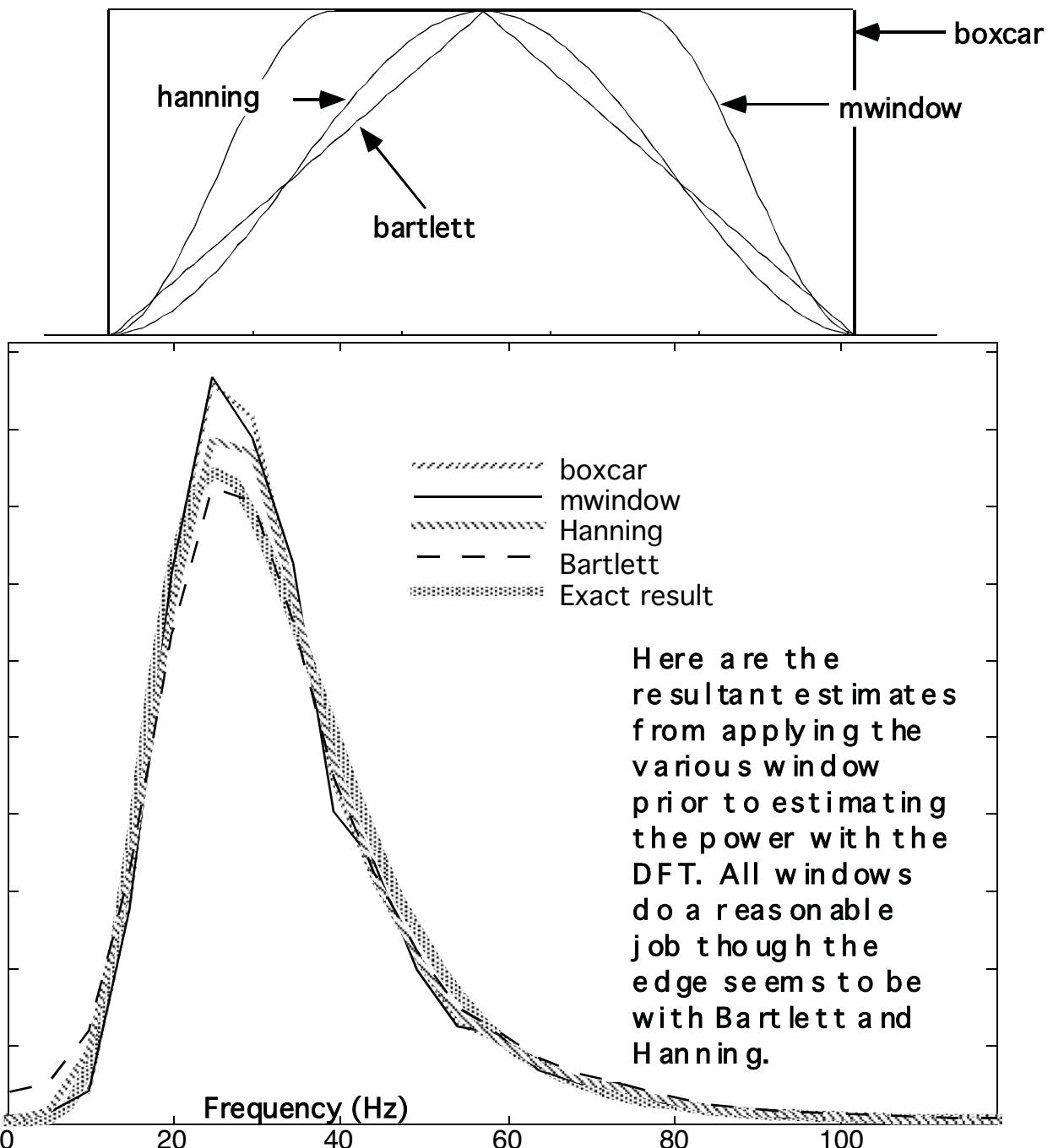


We see from this result that the autocorrelation of the seismogram is quite similar to that of the wavelet. Thus it is reasonable to ask if we can estimate the wavelet power spectrum from the central lags of the autocorrelation of the seismogram. Furthermore, we will do this without using any direct knowledge of the wavelet. So, we will take the samples from -.1 to .1 of the seismogram autocorrelation and compute their power spectrum. If we simply truncate the autocorrelation we obtain the result shown below:



Spectral Estimation

The preceding spectral estimate is not bad but can be improved by tapering the samples near the edge of the chosen window instead of simply truncating. The method of tapering is referred to as "windowing" and a number of special windows have been devised.

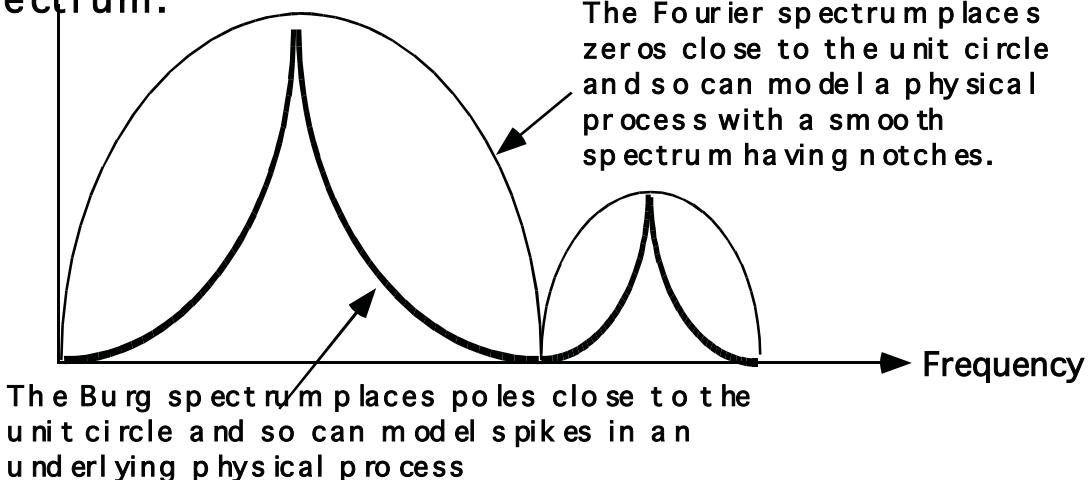


Spectral Estimation

The DFT is a polynomial in z containing no denominator terms.

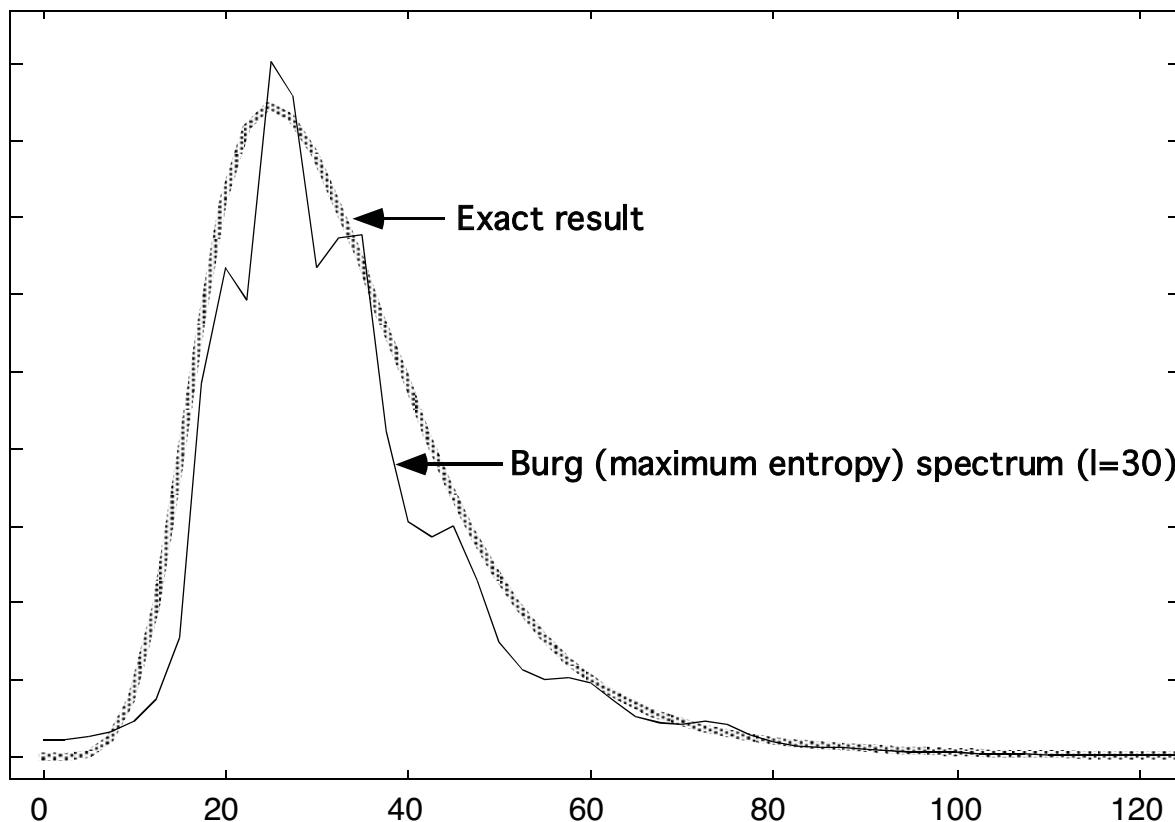
$$G(z) = \sum_{k=0}^{N-1} g_k z^k \quad z^k = e^{-i2\pi k/N}$$

Consequently, the DFT spectral estimate contains only zeros (no poles) in the z plane and is sometimes called an all-zeros estimate. An alternative estimate was developed by J.P. Burg (see Claerbout, 1976, Fundamentals of Geophysical Data Processing) which seeks to produce a spectral model using a Z transform with only denominator terms. This mathematical development of the Burg spectrum , also called the maximum entropy spectral estimate or all-poles estimate, is beyond the scope of this presentation. Nevertheless, this intuitive concept of the Burg technique helps us understand its basic behavior. As an all-poles estimate, it is very effective at modeling spectra which have isolated spikes but less so for smooth spectra. Furthermore, Burg developed the method using prediction operators to predict the time series outside of the truncation range so that the concept of a window does not apply to the Burg spectrum.



Spectral Estimation

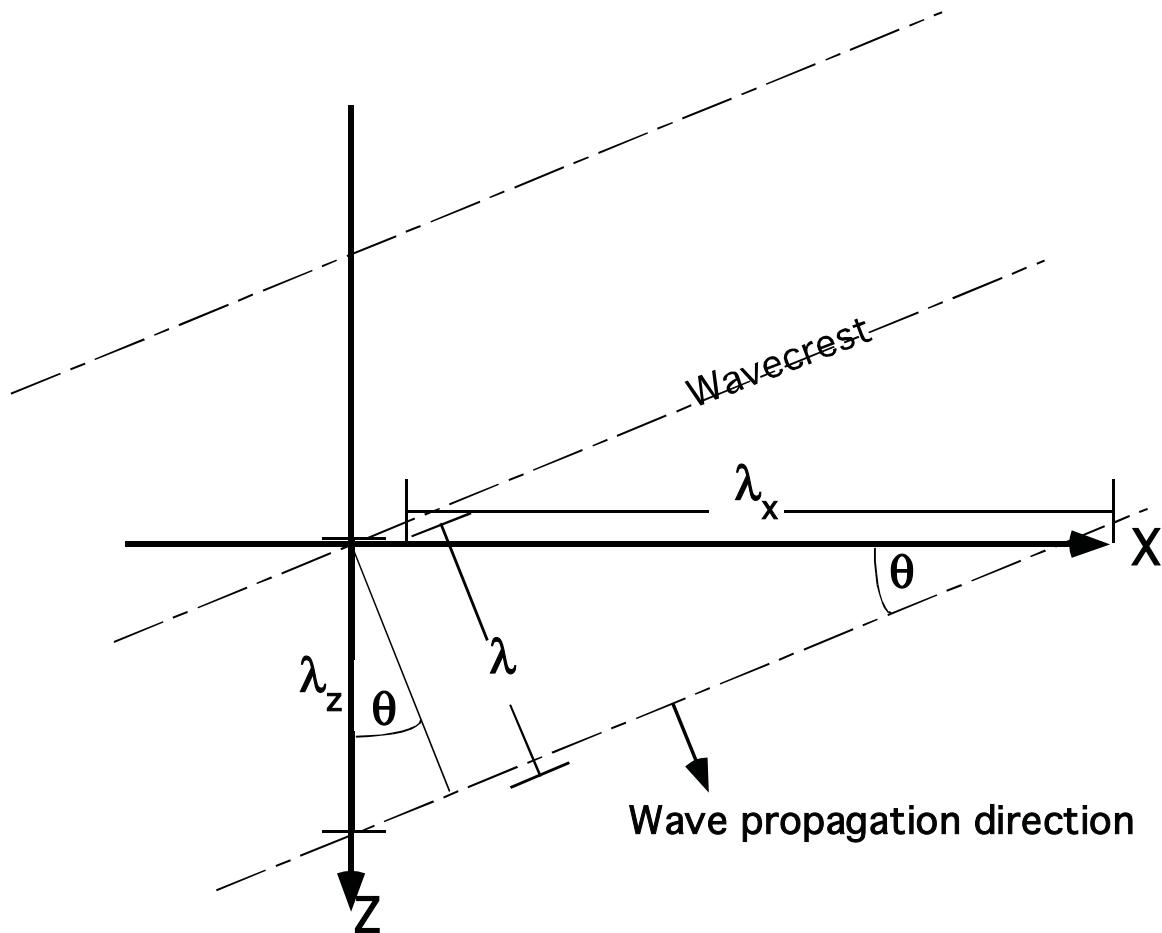
As might be expected from the preceding discussion, the Burg spectrum does not do a good job in this case:



However, this does not mean that the Burg spectrum is without merit. Hatton et al. (pages 36-38) give an excellent analysis showing the superiority of the Burg technique over the DFT in the case of the resolution of two closely space spectral peaks. Furthermore, as we shall see, the Burg technique leads to a very effective deconvolution method.

Wavelength Components

Consider a series of planar wavefronts propagating as shown below.



The distance between wavefronts, measured perpendicular to them, is defined as the wavelength, λ . We can also speak of the wavelength "components" in the various coordinate directions. For example, the horizontal wavelength, λ_x , is the distance between wavefronts measured in the x coordinate direction. Thus:

$$\lambda_x = \frac{\lambda}{\sin(\theta)} \quad \text{and} \quad \lambda_z = \frac{\lambda}{\cos(\theta)}$$

Wavelength Components

We see that the components of wavelength are never less than the wavelength itself. In fact, for a vertically traveling wave, λ_x is infinite. The components add as inverse squares:

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_x^2} + \frac{1}{\lambda_z^2}$$

It is often convenient to deal with vector quantities so we define the wavenumber, k , and its components as the inverse of the wavelength and its components.

$$k = \lambda^{-1} \quad k_x = \lambda_x^{-1} \quad k_z = \lambda_z^{-1}$$

$$k^2 = k_x^2 + k_z^2$$

k is the magnitude of a vector, \underline{k} , which points in the direction of wave propagation and whose components are the inverse wavelengths.

In 3-D, we have planar wavefronts instead of linear but a simple extension of this result still holds:

$$k_x^2 + k_y^2 + k_z^2 = k^2 \text{ The dispersion relation for scalar waves.}$$

Where:

$$k_y = \lambda_y^{-1}$$

Wavelength Components

This geometric relation between components of the wavenumber vector is fundamental to the study of wave propagation. It can be considered as the Fourier domain equivalent of the scalar wave equation. A fundamental result from theory is that the extrapolation of surface recorded data into the subsurface (z direction) requires knowledge of k_z . On the surface, we can measure k_x , k_y , and f , and since $f\lambda=v$, this allows k_z to be calculated from the dispersion relation:

$$k_z = \sqrt{\frac{f^2}{v^2} - k_x^2 - k_y^2}$$

Since k_z must be a real number (in order to be interpreted as an inverse wavelength) we see from this equation another fundamental result. Not all values of (k_x, k_y, f) can be considered as wavelike. In fact, we must have

$$\frac{f^2}{v^2} \geq k_x^2 + k_y^2$$

in order for a triplet of (k_x, k_y, f) to be a propagating wave.

Apparent Velocity

The wavelength components and the corresponding wavenumbers are closely related to the wave velocity and its components which are called apparent velocities. Recalling the basic relation, $\lambda f = v$, we see that the addition formula for wavelength components:

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_x^2} + \frac{1}{\lambda_z^2}$$

leads directly to:

$$\frac{1}{v^2} = \frac{1}{v_x^2} + \frac{1}{v_z^2}$$

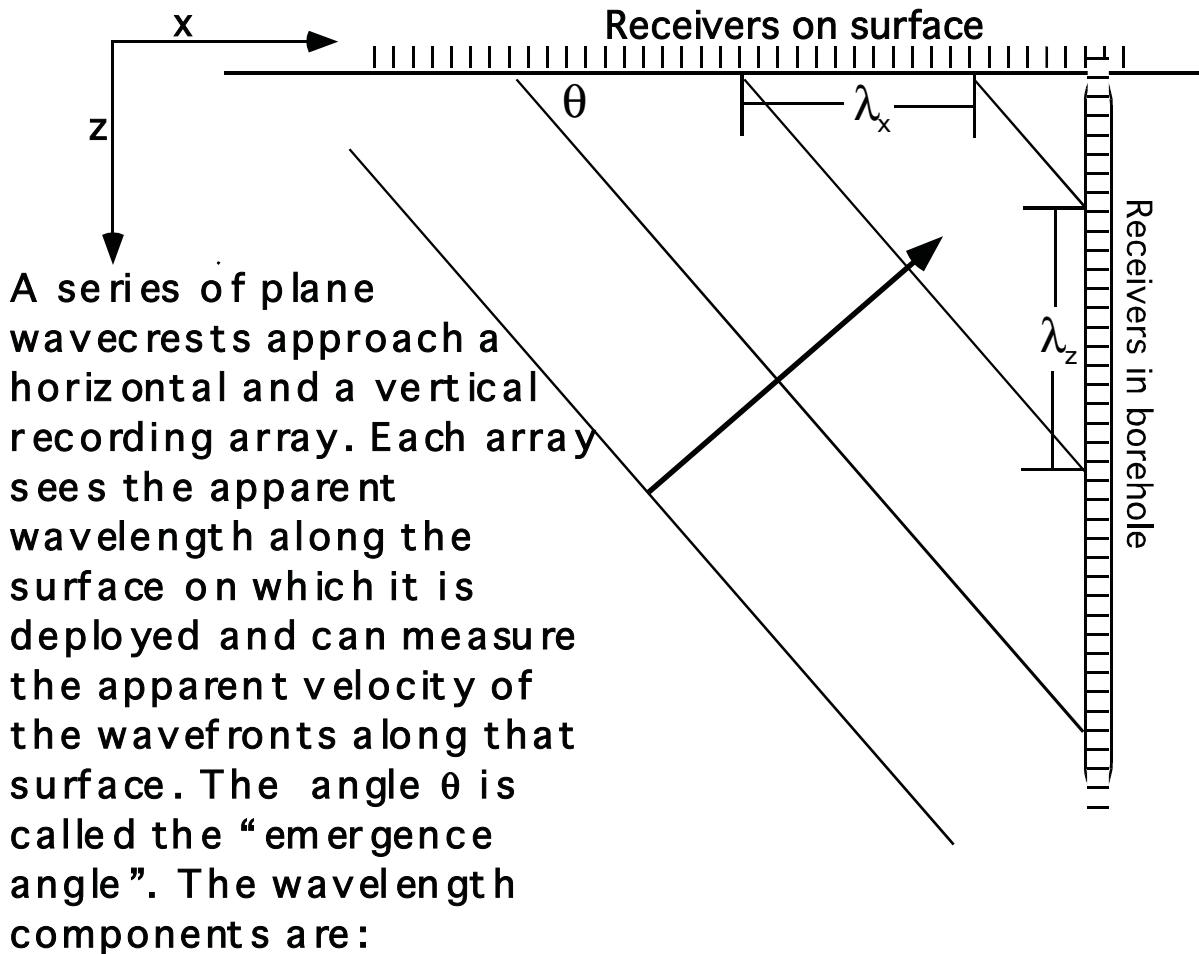
where $v_x = f\lambda_x$ and $v_z = f\lambda_z$

If we use wavenumber components, we have:

$$v = \frac{f}{k} \quad v_x = \frac{f}{k_x} \quad v_z = \frac{f}{k_z}$$

Nothing physical actually propagates at any of the apparent velocities. Rather, they are simply related to the arbitrary choice of coordinate directions and can be visualized as the wavelength along a coordinate direction divided by the time between wavecrests (i.e. the period of the waves.)

Apparent Velocity



$$\lambda_x = \frac{\lambda}{\sin(\theta)} \quad \text{and} \quad \lambda_z = \frac{\lambda}{\cos(\theta)}$$

And the apparent velocities are:

$$v_x = \frac{\Delta x}{\Delta t} = f\lambda_x = \frac{f\lambda}{\sin(\theta)} = \frac{v}{\sin(\theta)}$$

similarly $v_z = \frac{\Delta z}{\Delta t} = f\lambda_z = \frac{f\lambda}{\cos(\theta)} = \frac{v}{\cos(\theta)}$

The 2-D F-K Transform

The f-k transform is a fundamental tool which essentially allows the direct computation of wavenumber components and frequency for a multidimensional wavefield. In 2-D, it can be written:

$$\varphi(k_x, f) = \iint_{-\infty}^{\infty} \psi(x, t) e^{2\pi i (k_x x - f t)} dx dt$$

and in 3-D

$$\varphi(k_x, k_y, f) = \iiint_{-\infty}^{\infty} \psi(x, y, t) e^{2\pi i (k_x x + k_y y - f t)} dx dy dt$$

The inverse transforms are mathematically similar:

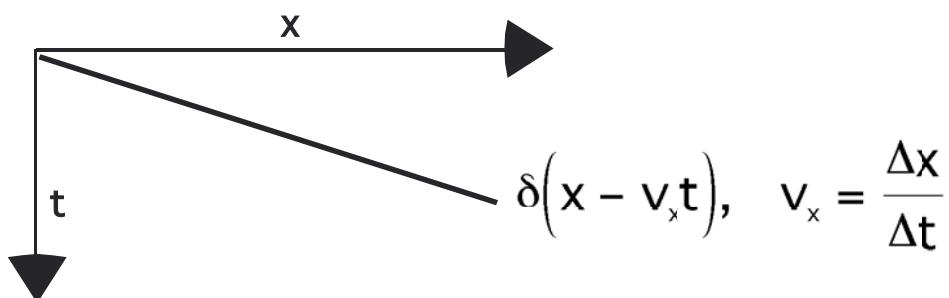
$$\psi(x, t) = \int \int_{-\infty}^{\infty} \phi(k_x, , f) e^{-2\pi i (k_x x - f t)} dk_x df \quad \text{2-D}$$

$$\psi(x, y, t) = \int \int \int_{-\infty}^{\infty} \phi(k_x, k_y, f) e^{-2\pi i (k_x x + k_y y - f t)} dk_x dk_y df \quad \text{3-D}$$

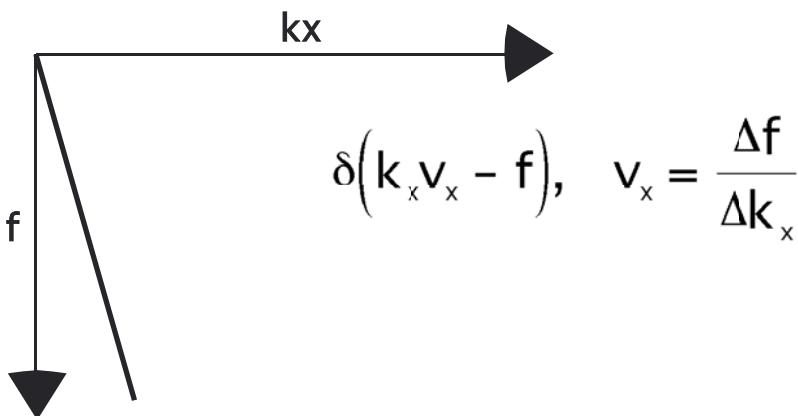
These inverse transforms have the physical interpretation of presenting a wavefield as a superposition of individual Fourier components or "plane waves".

The 2-D F-K Transform

A few f-k transforms are known analytically. Perhaps the most important is the transform of a single linear event. Using the mathematics of Dirac delta functions, a seismic wavefield consisting of an isolated linear event can be written:



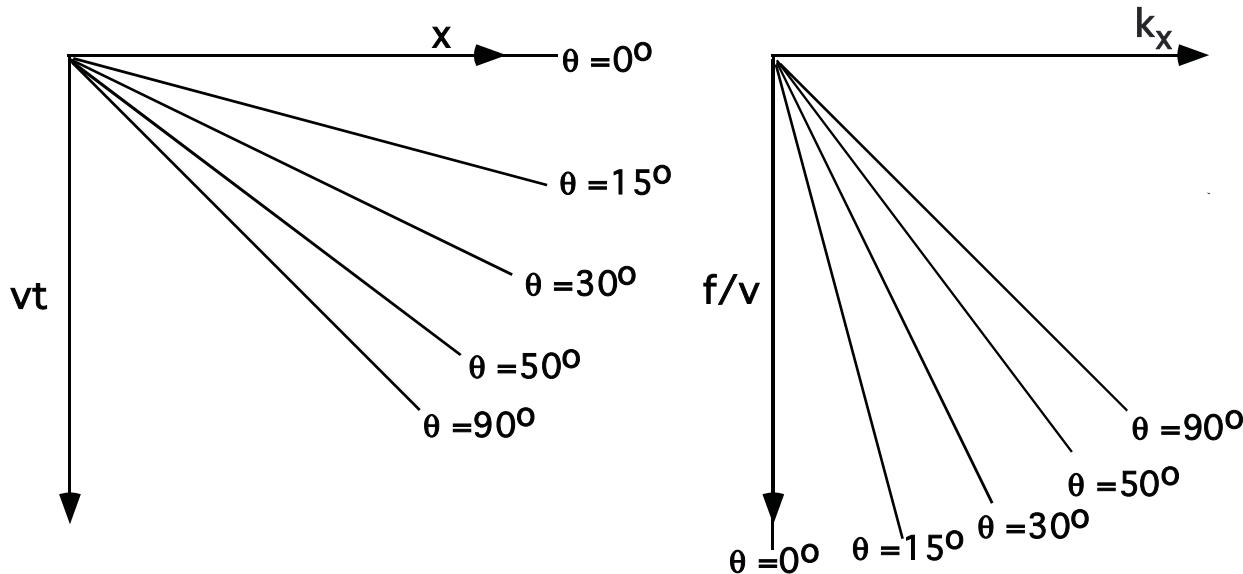
$$\begin{aligned}\varphi(k_x, f) &= \iint_{-\infty}^{\infty} \delta(x - v_x t) e^{2\pi i (k_x x - f t)} dx dt \\ &= \int_{-\infty}^{\infty} e^{2\pi i (k_x v_x - f)t} dt = \delta(k_x v_x - f)\end{aligned}$$



- Horizontal events in (x, t) are vertical in (k_x, f) and vice-versa.
- All events in (x, t) with the same apparent velocity, v_x , are collected into a single linear event in (k_x, f) . The different events are distinguished by their phase spectra but have differing phase spectra.

The 2-D F-K Transform

If we consider all possible linear events characterized by $v_x = v/\sin(\theta)$, then we have:

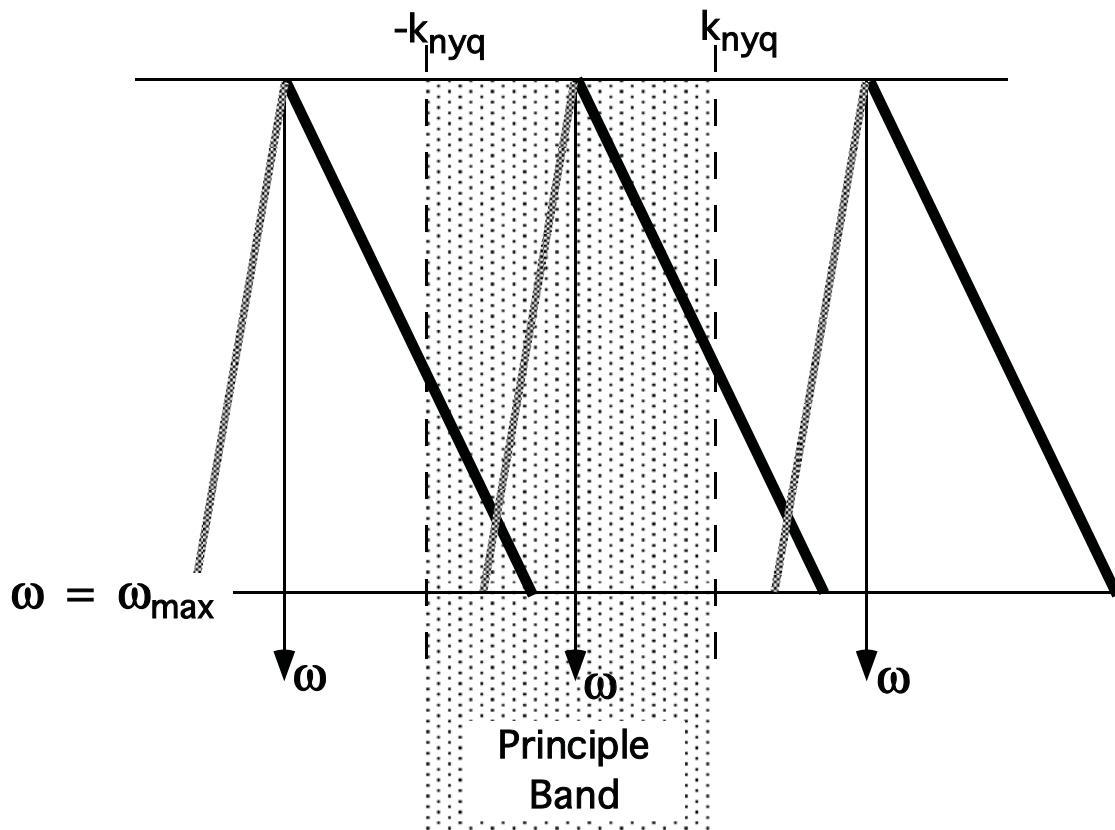


The previously encountered fact that $f/v > k_x$ is reexpressed in this analysis through the fact that a portion of (k_x, f) space is not populated. As a general rule of thumb, we see that large k_x values are only wavelike at high frequencies. This fact will turn out to be fundamental in describing the ability of seismic images to resolve small features. Small features require large k_x values which, in turn, require a large temporal frequency bandwidth.

In proceeding from analytic to discrete f-k transforms, it turns out that the implementations of the Fourier transform integrals are approximate but the forward and backward DFKT are exact inverses of each other. This fact is a great convenience in data processing and is not generally true of other transforms such as the Radon transform.

The 2-D F-K Transform

When we proceed from the continuous F-K transform to the discrete, a situation directly analogous to the 1-D case occurs. That is, the act of spatial sampling induces a periodicity in the (ω, k) domain. Unlike temporal aliasing, spatial aliasing is always present.

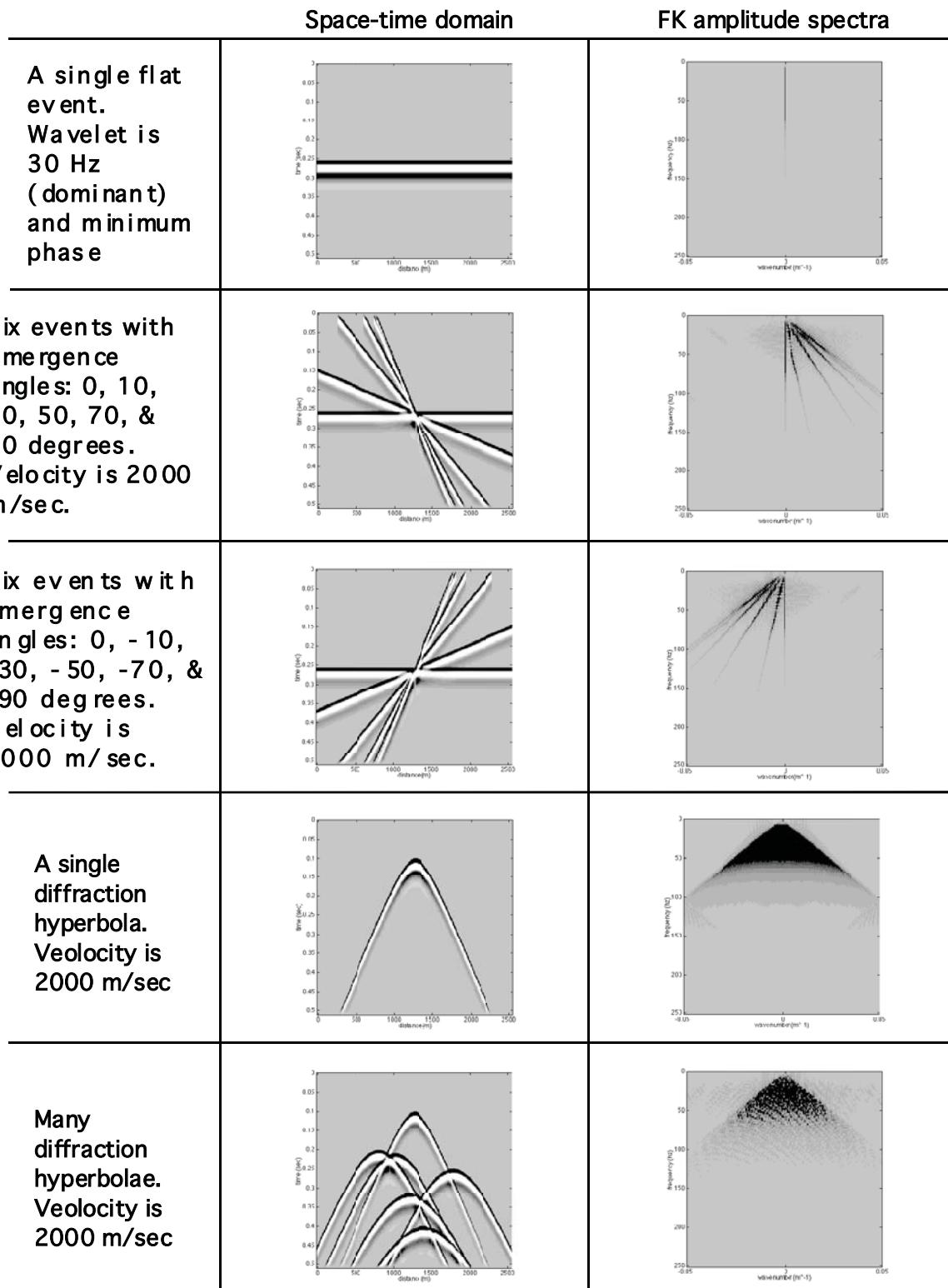


Here we see one event showing spatial aliasing and another which does not. Given a spatial sample rate of Δx and an apparent velocity v_a then all temporal frequencies higher than:

$$f_{crit} = 2\pi \omega_{crit} = v_a k_{nyquist} = \frac{v_a}{2\Delta x}$$

will be spatially aliased. For excellent illustrations of spatial aliasing see Hatton et al. pp43-45 and Yilmaz pp62-69

FK Transform Pairs

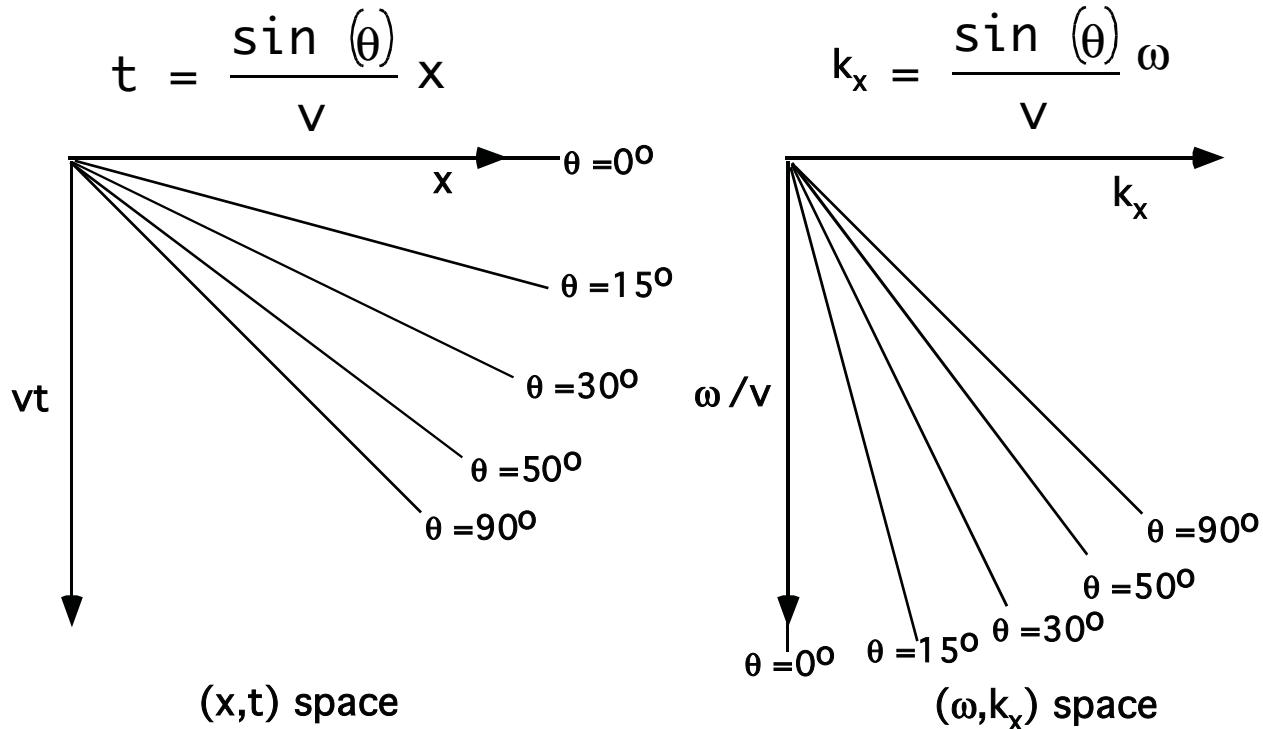


τ -p Transforms

The τ -p transform, also known as the Radon transform or slant stack, is a very useful data processing tool due to its ability to decompose a seismic matrix into events of constant horizontal slowness, p. Its close relation to the f-k transform is captured in the "projection slice theorem" which shows that the τ -p transform may be computed from an f-k transform which has been interpolated to "polar" coordinates. (See Deans, S.R., 1983, The Radon Transform and Some of Its Applications, John Wiley and Sons). Consider the expression for a forward f-k transform:

$$\phi(k_x, f) = \iint_{-\infty}^{\infty} \Psi(x, t) e^{2\pi i (k_x x - f t)} dx dt \quad (1)$$

We have seen how this expression transforms linear events in (x, t) into linear events in (k_x, f) :



τ - p Transforms

Note that $\sin(\theta)/v$ (horizontal slowness) can also be written as dt/dx or the ray parameter p . Thus, radial lines in the f - k transform are lines of constant p . This can be examined further by a substitution of variables in the f - k integral (1):

$$\phi(p, f) = \iint_{-\infty}^{\infty} \Psi(x, t) e^{2\pi i f(p x - t)} dx dt \quad \text{where} \quad p = \frac{k_x}{f} \quad (2)$$

Here p has been explicitly introduced as the ratio of k_x and f and hence is constant along radial lines in (k_x, f) space. So $\phi(p, f)$ can be regarded as a "polar coordinate" representation of $\phi(k_x, f)$. Now, consider the meaning of equation 2 for constant p by performing the t integration first:

$$\phi(p, f) = \int_{-\infty}^{\infty} \psi(x, f) e^{2\pi i f p x} dx \quad (3)$$

where

$$\psi(x, f) = \int_{-\infty}^{\infty} \Psi(x, t) e^{-2\pi i f t} dt \quad (4)$$

Then, compute the inverse Fourier transform ($f \rightarrow t$) of (3)

$$\varphi(p, \tau) = \int_{-\infty}^{\infty} \phi(p, f) e^{2\pi i f \tau} df \quad (5)$$

Now, substitute (3) into (5):

$$\varphi(p, \tau) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \psi(x, f) e^{2\pi i f p x} dx \right] e^{2\pi i f \tau} df$$

τ -p Transforms

Interchange the order of integration:

$$\varphi(p, \tau) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \psi(x, f) e^{2\pi i f(px + \tau)} df \right] dx$$

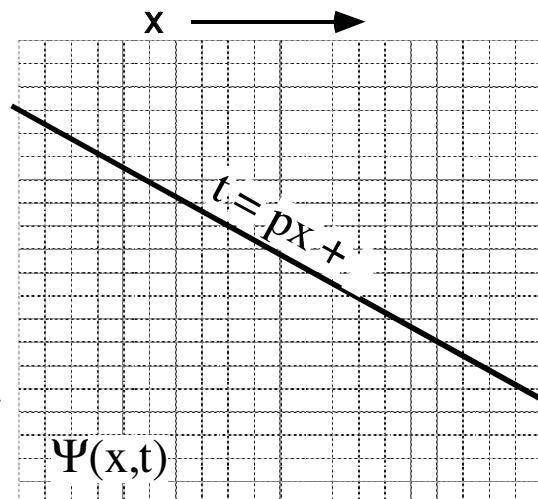
The inner integral gives $\Psi(x, px + \tau)$, so:

$$\varphi(p, \tau) = \int_{-\infty}^{\infty} \Psi(x, px + \tau) dx \quad (6)$$

Equation (6) is the conventional equation for the τ -p transform (compare with Yilmaz (Seismic Data Processing, 1987) equation 7.5). Several things can be learned from this development:

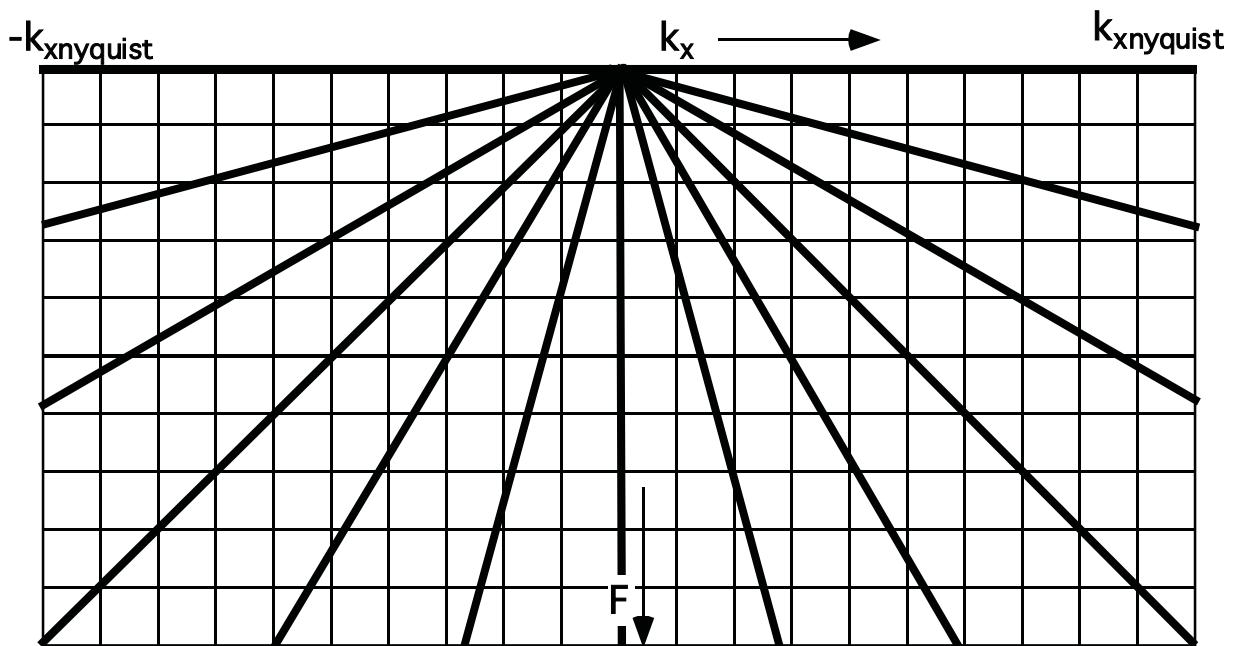
- The τ -p transform can be computed from the f-k (Fourier) transform by a coordinate change from (f, k_x) to (f, p) followed by an inverse Fourier transform from $f \rightarrow \tau$. This amounts to changing to polar coordinates in the Fourier domain.
- The τ -p transform may equivalently be computed by equation (6) which is a process known as "slant stacking"

For fixed (p, τ) , equation 6 represents a summation through the function t $\Psi(x, t)$ along a linear trajectory. Hence it is called slant stacking.



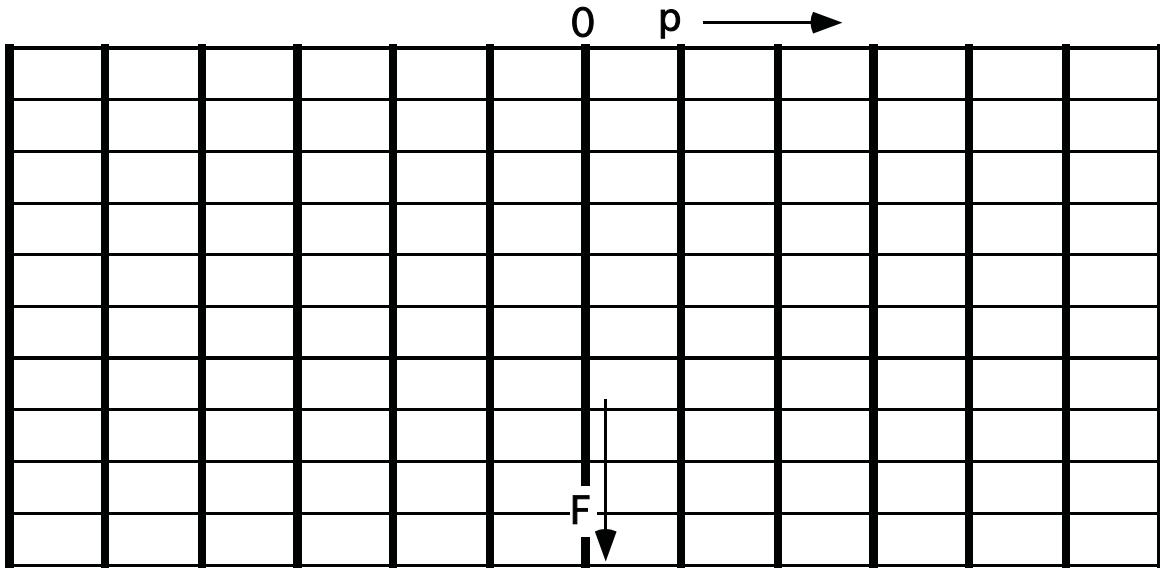
τ - p Transforms

- Since the analytic τ - p transform is computable from the 2-D Fourier transform (and vice-versa) the information content is the same in either domain. The fact that the 2-D Fourier transform is complete means that the analytic τ - p transform is also. We will see that this is not true for the digital τ - p transform.



Here we see an illustration of the representation of (f, k_x) space in both rectangular and polar coordinates. The radial lines are lines of constant p and are all shown to terminate (where possible) at the same constant f . To compute the discrete τ - p transform transform, spectral values are interpolated from the rectangular (f, k_x) grid to regularly sampled f locations on each radial line:

τ -p Transforms

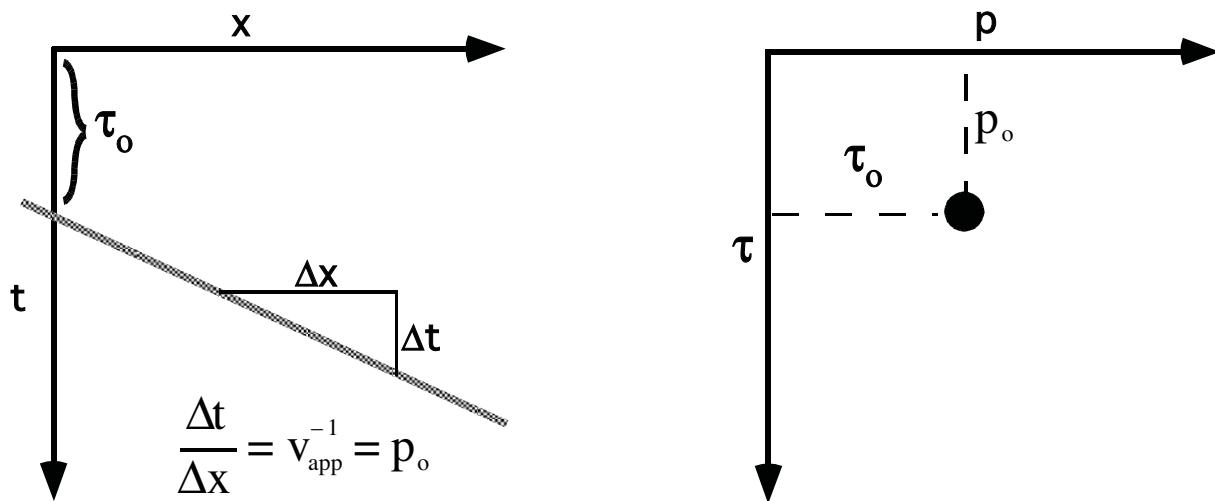


After interpolation onto radial lines, (f, k_x) space becomes (f, p) space. An inverse Fourier transform from f to τ completes the journey to (τ, p) space.

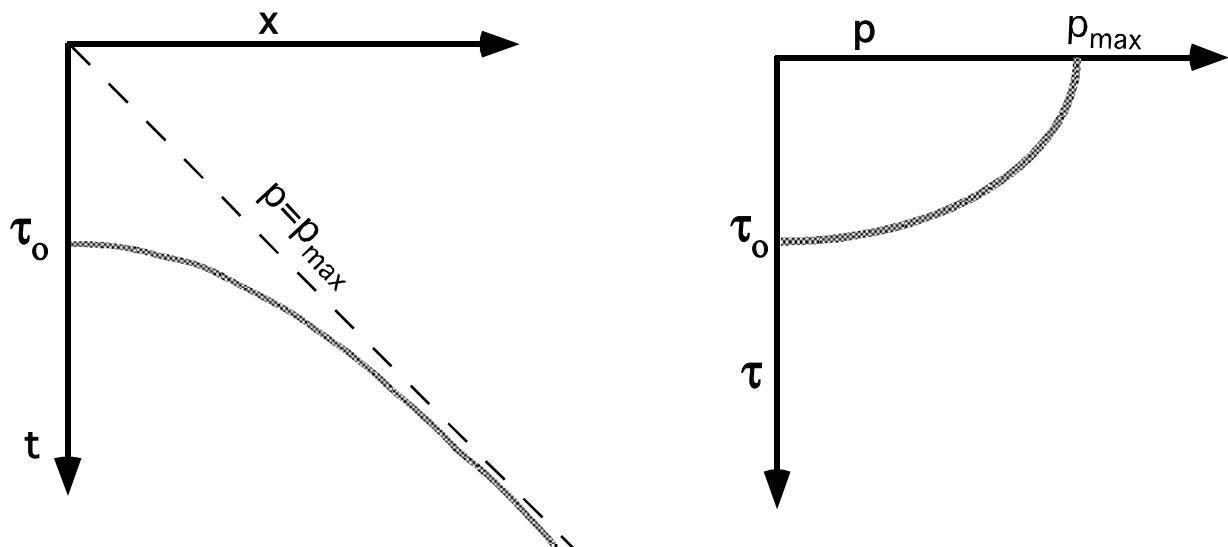
Close inspection of the figure on the previous page shows why the discrete τ -p transform has difficulty even though the analytic τ -p transform is complete. It is impossible to pick a set of discrete p values which cover the (f, k_x) grid uniformly. Either they are too far apart at the grid edges or they are too crowded near the center. In either case, it can be shown that there is always "information loss" in going to the discrete (τ, p) space and back again. Put another way, merely transforming data to (τ, p) space and back (without any τ -p processing) will always alter the data in some way. Thus the discrete τ -p transform is not complete in the same sense that the discrete (f, k_x) transform is.

Properties and uses of the τ - p Transform

The most obvious property of a τ - p transform is that it maps a linear event in (x, t) to a point in (τ, p) .

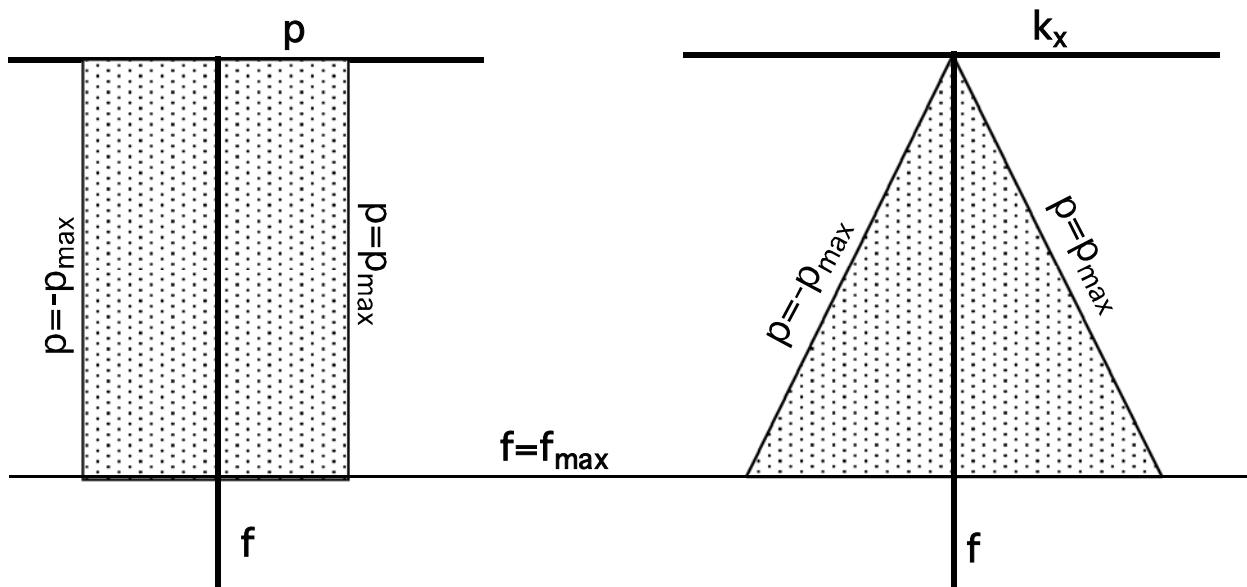


Less obviously, hyperbolae map to ellipses:



Properties and uses of the τ - p Transform

Thus we can expect that bandlimited propagating body waves in a constant velocity earth will map to a compact region of (f, p) space defined by $p_{\max} = 1/v$. In the f - k transform, this corresponds to a triangular region.



Thus we expect that apparent velocity filtering can be done in either domain by essentially muting (suppressing) that portion of the domain corresponding to the undesirable velocities.

Aliasing affects the (τ, p) transform much as it does the (f, k) transform. If the (τ, p) transform is constructed by slant stacking in (x, t) or (x, f) then it is not directly affected by horizontal aliasing. But the choice of Δp and the number of p values is a difficult one and leads directly to p aliasing.

A rule of thumb for Δp is

$$\Delta p = \frac{1}{f_{\max} (x_{\max} - x_{\min})} \approx \frac{\Delta k_x}{f_{\max}}$$

Tuner, G., 1990, Aliasing in the tau-p transform and the removal of spatial aliased coherent noise: Geophysics, 55, 1496-1503

Properties and uses of the τ - p Transform

Other uses of the τ - p Transform:

- Since a slant stack is less affected by spatial aliasing than an f-k transform, it can be used to interpolate to finer trace spacings and "unalias" data. Used in this fashion it is often called a "smart interpolator". (Yilmaz, O., 1987, Seismic Data Processing, p435.)
- It can be shown that multiples are not periodic on an offset trace in the (x,t) domain but are in the (τ,p) . (Treitel et al., 1982, Plane-wave decomposition of seismograms, Geophysics, 47, 1375-1401) This means that predictive deconvolution for multiple suppression often works better on (τ,p) gathers.
- Migration can also be done in the (τ,p) domain. (Diebold, J.B., and Stoffa, P.L., 1981, The traveltime equations, tau-p mapping, and inversion of common midpoint data, Geophysics, 46, 238-254)

Inverse τ - p Transforms

The process of reconstruction of the seismic data in (x,t) space given its τ - p transform is called an inverse τ - p transform. There are a number of ways to do this process though we shall discuss only two: Fourier methods and filtered back projection.

The Fourier method is obvious from the discussion of the forward τ - p transform. The major step is the reconstruction of the 2-D (f,k_x) transform which requires an interpolation onto a rectangular grid from a polar one. This will obviously have numerical difficulties though they are controllable. Following the interpolation, an inverse 2-D Fourier transform completes the process.

Filtered back projection avoids the (f,k_x) domain and reconstructs the image directly with a convolutional filter followed by an inverse slant stack. Consider the expression for the inverse 2-D Fourier transform:

$$\Psi(x,t) = \iint_{-\infty}^{\infty} \phi(k_x, f) e^{-2\pi i (k_x x - f t)} dk_x df \quad (1)$$

Now, letting $k_x = fp$ and converting the wavenumber integral into a p integral gives:

$$\Psi(x,t) = \iint_{-\infty}^{\infty} \left[f \phi(p, f) e^{-2\pi i f p x} \right] e^{2\pi i f t} dp df \quad (2)$$

Inverse τ - p Transforms

The term in brackets can be considered to be the product of two functions of f . Hence, it must be a convolution in time:

$$\Psi(x,t) = \int_{-\infty}^{\infty} \alpha(t) \bullet \beta(p,x,t) dp \quad (3)$$

where \bullet denotes a convolution over time and

$$\alpha(t) = \int_{-\infty}^{\infty} f e^{2\pi i f t} df \quad (4)$$

$$\beta(p,x,t) = \int_{-\infty}^{\infty} \phi(p,f) e^{2\pi i f (t-px)} df = \phi(p,t-px) \quad (5)$$

Note that $\phi(p,\tau)$ is the forward slant stack. Substitution of (5) into (3) results in:

$$\Psi(x,t) = \alpha(t) \bullet \int_{-\infty}^{\infty} \phi(p,t-px) dp \quad (6)$$

Equation (6) expresses filtered back projection from the (τ,p) domain to the (x,t) domain. Each point in (x,t) is constructed by integrating along a linear trajectory in (τ,p) , just like the forward slant stack. Unlike the forward operation, the integration is followed by a convolution which is a form of a filter.

Inverse τ - p Transforms

Another way to do the inverse transform is suggested by equation 2. Rather than convolve in the time domain with a filter operator we can do the reconstruction in the f domain. Taking a forward Fourier transform ($t \rightarrow f$) of (2) gives:

$$\mathbf{F}[\Psi(x,t)] = \psi(x,f) = f \int_{-\infty}^{\infty} \phi(p,f) e^{-2\pi i f p x} dp \quad (7)$$

To use (7) for the inversion, we first transform $\phi(p,\tau)$ to $\phi(p,f)$. Then, for each x , we multiply $\phi(p,f)$ by a p dependent phase shift and integrate over p and the result is scaled by f . After constructing $\psi(x,f)$, an inverse Fourier transform from $f \rightarrow t$ completes the process.

Least Squares τ -p and f-k Transforms

We have seen that a convenient method of implementing forward and inverse τ -p transforms is in the frequency domain:

$$\phi(p,f) = \int_{-\infty}^{\infty} \psi(x,f) e^{2\pi i f p x} dx \quad (1)$$

$$\psi(x,f) = f \int_{-\infty}^{\infty} \phi(p,f) e^{-2\pi i f p x} dp \quad (2)$$

Here (1) is the forward transform from (x,f) to (p,f) and (2) is the inverse transform. Virtually any integration can be implemented as an equivalent matrix operation for discrete data. Corresponding to (1) and (2) we have:

$$\phi_j(f) = \sum_k R_{jk} \psi_k(f) \quad R_{jk} = \exp(2\pi i f p_j x_k) \quad (3)$$

$$\psi_k(f) = f \sum_j R_{kj}^* \phi_j(f) \quad R_{kj}^* = \exp(-2\pi i f p_j x_k) \quad (4)$$

These can be written:

$$\bar{\phi} = \underline{R} \bar{\psi} \quad (5)$$

$$\bar{\psi} = f \underline{R}^{*T} \bar{\phi} \quad (6)$$

$$\bar{\phi} = \begin{bmatrix} \phi_1(f) \\ \phi_2(f) \\ \phi_3(f) \\ \vdots \end{bmatrix} \quad \bar{\psi} = \begin{bmatrix} \psi_1(f) \\ \psi_2(f) \\ \psi_3(f) \\ \vdots \end{bmatrix} \quad \underline{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & \vdots \\ R_{21} & R_{22} & R_{23} & \vdots \\ R_{31} & R_{32} & R_{33} & \vdots \\ \cdots & \cdots & \cdots & \ddots \end{bmatrix} \quad \text{etc}$$

Least Squares τ -p and f-k Transforms

Rather than compute the forward transform directly, the least squares technique uses equation (6) to pose an inverse problem for the τ -p spectrum.

$$\bar{\psi} = \underline{f} \underline{\mathbf{R}}^{*T} \bar{\phi} \quad (6)$$

$$\underline{\mathbf{R}} f^{-1} \bar{\psi} = \underline{\mathbf{R}} \underline{\mathbf{R}}^{*T} \bar{\phi}$$

$$\bar{\phi} = (\underline{\mathbf{R}} \underline{\mathbf{R}}^{*T})^{-1} \underline{\mathbf{R}} f^{-1} \bar{\psi} \quad (7)$$

Equation (7) is the standard least squares estimate of the τ -p spectrum. It is usually superior in the sense that the (x,t) domain data can be reconstructed from it with fewer artifacts. This formulation assumes that the number of p traces estimated will be no larger than the number of x traces. Even when the data is perfectly regular in x and the number of p and x traces are the same, the least squares method is usually superior because the τ -p transform is incomplete. This means that the forward and reverse τ -p processes leave artifacts in the data. The least squares approach minimizes such artifacts.

Least Squares τ - p and f-k Transforms

The more incomplete and inconsistent a transform pair are, the more the least square approach becomes useful. This means it is especially preferred for slant stacks along parabolic and hyperbolic trajectories which are incomplete even in the analytic sense.

Another example of an incomplete transform is the discrete Fourier transform for irregularly sampled data. It can also be posed as an inverse problem:

$$\begin{aligned}\bar{\psi} &= \underline{\mathbf{F}}^{*T} \bar{\phi} \\ \bar{\phi} &= (\underline{\mathbf{F}}\underline{\mathbf{F}}^{*T})^{-1} \underline{\mathbf{F}} \bar{\psi}\end{aligned}$$

where

$$\underline{\mathbf{F}} = \begin{bmatrix} F_{11} & F_{12} & F_{13} & \vdots \\ F_{21} & F_{22} & F_{23} & \vdots \\ F_{31} & F_{32} & F_{33} & \vdots \\ \dots & \dots & \dots & \ddots \end{bmatrix} F_{mn} = \exp(2\pi i k_{xm} x_n)$$

See Marfurt, et al., (1996, Pitfalls of using conventional Radon transforms on poorly sampled data: Geophysics, 61, 1467-1482) for a more complete discussion.

Methods of Seismic Data Processing

**Lecture Notes
Geophysics 557**

**Chapter 3
Amplitude Effects**

Seismic Wave Attenuation

As a seismic wave propagates through the earth, it suffers attenuation (amplitude decay) for a number of reasons:

Attenuation Mechanism #1: Geometric Spreading (or Spherical Divergence)

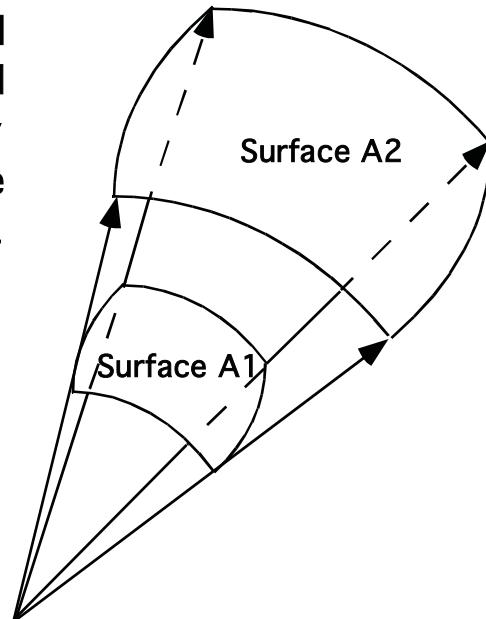
As seismic energy propagates away from a source (or focal point) the conservation of total energy requires that the energy found on the wavefront surface A₁ at some time t₁ equal that on surface A₂ at some time t₂.

$$E(t_2) = \epsilon(t_2)A_2 = E(t_1) = \epsilon(t_1)A_1$$

where ϵ is the energy per unit area. Since the displacement wave amplitude, u , is proportional to the square root of ϵ , we deduce:

$$\frac{u_2}{u_1} = \sqrt{\frac{A_1}{A_2}} = \frac{R_1}{R_2} = \frac{t_1}{t_2} \quad \text{or}$$

$$u(t) = \frac{u_0}{R(t)}$$



The proper interpretation of this result is that the wave amplitude decays as $1/R$ where R is the radius of curvature of the wavefront. In the case of a constant velocity medium, R is simply the distance travelled; however, in a layered medium, R can be shown to be proportional to $(V_{\text{rms}}^2/V_1)t$ where V_1 is the velocity of the first layer. (Newman, Geophysics, 1971, p 481-488, Hubral, P., and Krey, T., Interval Velocities from Seismic Reflection Time Measurements, 1980, Society of Exploration Geophysicists)

Seismic Wave Attenuation

Thus we deduce that the effects of spherical divergence can be approximately compensated for by applying a "gain correction" given by:

$$G(t)_{\text{spreading}} = G_0 V_{\text{rms}}^2(t)t \quad (\text{Compare with Hatton et al., page 56})$$

Attenuation Mechanism #2: Absorption (or inelastic attenuation)

In a perfectly elastic medium, the total energy of the propagating wavefield remains a constant. However, the earth is not perfectly elastic and propagating seismic waves gradually die out over time. The primary mechanism for this is the continuous conversion of a small portion of the seismic energy to heat due to irreversible anelastic behavior of rocks. It is customary to talk about the parameter Q which characterizes this energy loss:

$$Q = \frac{\text{energy}}{\text{energy loss}} \text{ per frequency cycle}$$

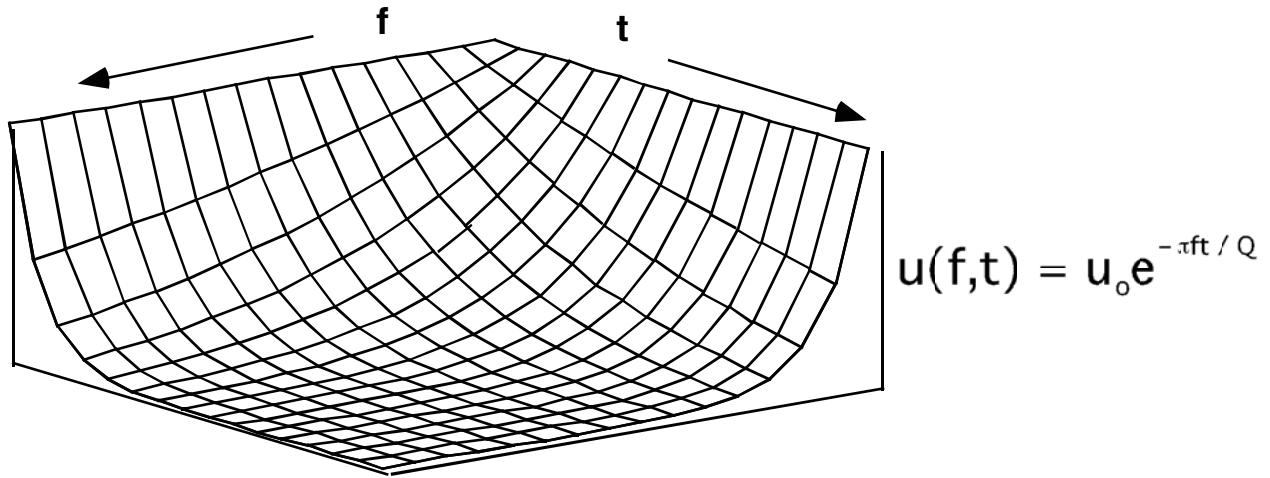
Various attenuation theories exist with the simplest being the "constant Q " theory of Kjartansson¹ and others. Most empirical evidence is consistent with the assumption that Q is independent of frequency at least over the seismic bandwidth. The constant Q theories all predict an amplitude loss given by:

$$u(f,t) = u_o e^{-\pi ft / Q}$$

¹Kjartansson, E, 1979, Constant Q-Wave Propagation and Attenuation, JGR, V84, p4737-4748

Seismic Wave Attenuation

Thus, the constant Q theory refers to a Q which is independent of frequency but predicts an attenuation which is a first order exponential in both time and frequency.



Constant Q Exponential Decay Surface

Note that $Q = \infty$ is a perfectly elastic material while $Q = 0$ is perfectly absorptive. A highly absorptive rock has a Q of 20-50 while very competent limestones and dolomites can have Q of 200 or more. The common first order correction for Q effects is to apply a simple, frequency independent, exponential gain correction. If we write:

$$e^{-\pi ft / Q} \approx e^{-\alpha t} ; \alpha = \pi f_{dom} / Q = \text{the attenuation constant}$$

Typically, the attenuation constant is expressed in db/sec which would be

$$\alpha_{db/sec} = 20\pi \log_{10}(e) \frac{f_{dom}}{Q}$$

Assuming f_{dom} of 20 and a Q of 100 leads to a "typical" value of 12 db/sec.

Seismic Wave Attenuation

Attenuation Mechanism #3: Transmission losses

In our examination of the theory of the 1-D synthetic seismogram we saw that there is a continuous amplitude decay due to transmission losses. In fact, we found that the earth's impulse response resulted in the recording of the n'th reflection coefficient at the surface multiplied by a transmission loss term:

$$\text{nth reflection coefficient recorded at surface} = R_n \text{ (Transmission losses)}$$

$$\text{where transmission losses} = \prod_{k=1}^{n-1} (1 - R_k^2)$$

This effect is highly dependent upon local geology and is difficult to estimate with any precision. It is customary to ignore it and "hope" that it is either small or included in the "db/sec" corrections already discussed.

Attenuation Mechanism #4: Mode Conversion

As waves propagate in an elastic medium, they are constantly being converted from P to S and the reverse at every impedance contrast. These mode conversions occur both upon reflection and transmission and are described by the famous Zoepritz equations (see Aki and Richards, 1980, or the CREWES Zoepritz explorer at www.crewes.org). If, as in conventional seismic, only the vertical component of ground motion is recorded, then it is rarely possible to address this effect. The solution is to record all three components of ground motion and process the data as elastic waves. This is the subject of leading edge research around the world.

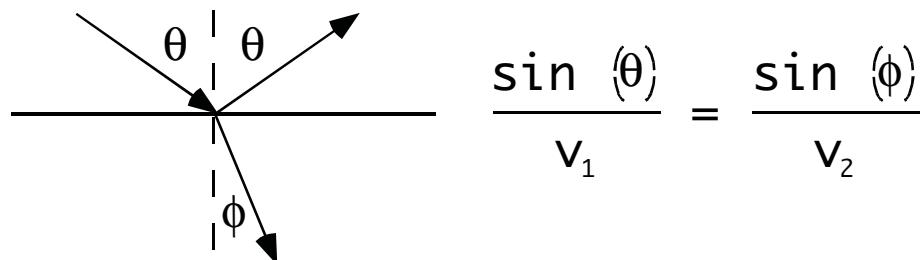
Seismic Wave Attenuation

Attenuation Mechanism #5: Scattering

Random scattering off small irregularities causes the dispersal of seismic wavefields and an apparent loss of energy. If a full 3-D wavefield has been recorded then such scatterers can be imaged by migration but the lost wavefield energy is not restored.

Attenuation Mechanism #6: Refractions and critical angles

Snell's law governs the angles of reflection and refraction when a wave interacts with an impedance contrast.

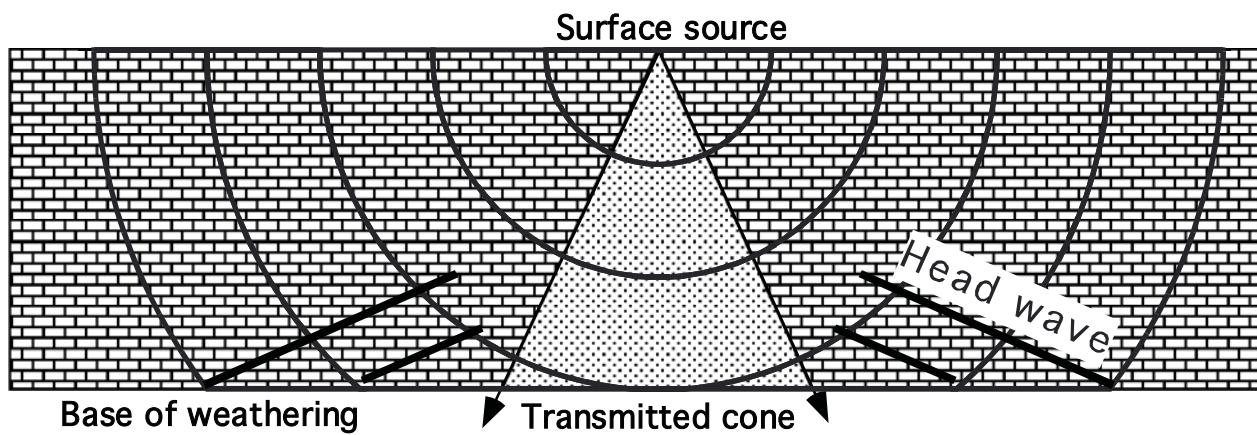


In the normal case where $v_2 > v_1$, there exists a "critical angle" of incidence beyond which no transmission occurs.

$$\sin(\theta_{\text{crit}}) = \frac{v_1}{v_2}$$

Seismic Wave Attenuation

Energy incident at or beyond the critical angle is thrown back to the surface as post-critical reflections and refractions. It is not available to illuminate deeper reflectors. This is especially noticeable in the near surface where the velocity contrast at the base of the weathering can approach 1/2 or less. Since the arcsin of 1/2 is 30 degrees, this means only a narrow cone of energy penetrates to the subsurface.



True Amplitude Processing

We have seen that a simple model for a 1-D seismogram predicts that the seismic data consists of band-limited reflection coefficients:

$$s(t) = w(t) \bullet r(t)$$

Where w is a seismic wavelet, r is the earth's reflectivity expressed as a function of 2-way vertical traveltime, s is the seismogram, and \bullet denotes convolution. Since w generally contains significant energy only over some characteristic frequency bandwidth, if we view the convolution as a multiplication in the frequency domain, we see that s is indeed a bandlimited version of r . (If w is not zero phase then there is a phase shift as well.)

There are many real earth wave propagation effects which cause the raw seismic data to deviate considerably from this model. True amplitude processing is a "holy grail" of the seismic data processing world and refers to a processing sequence which, when complete, yields data which is accurately representable as bandlimited reflection coefficients.

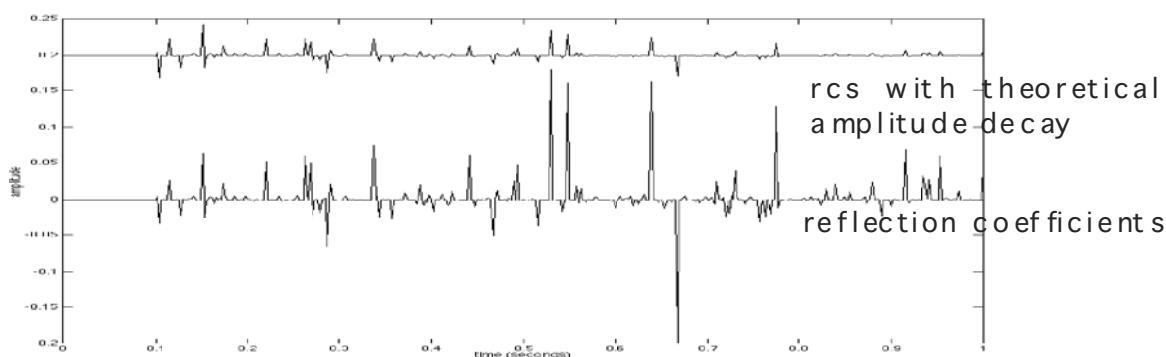
While not yet strictly possible, many data processing flows come quite close to being true amplitude processing. Generally, though not exclusively, this means the avoidance of statistical amplitude corrections like AGC in favor of deterministic corrections like spherical divergence and exponential gain.

It is not uncommon to find modern processed seismic data which is roughly proportional to well log derived reflection coefficients over limited time zones.

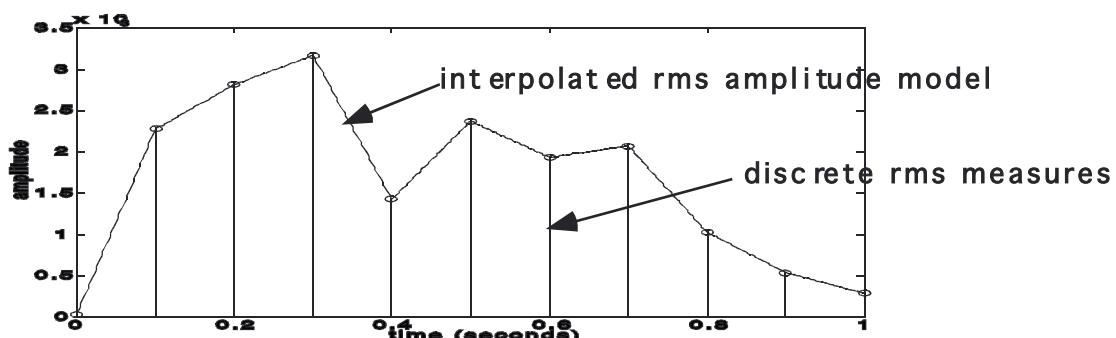
Automatic Gain Correction (AGC)

Automatic gain correction methods attempt to perform necessary amplitude adjustments to seismic data based purely on statistics of the observed amplitude decay. They should be contrasted with deterministic methods which use a physical model of one or more decay processes to determine correction factors. Generally, AGC methods are simple and efficient but tend to produce unphysical amplitude distortions. They are useful in situations where physically meaningful amplitudes are less important than "well balanced traces.

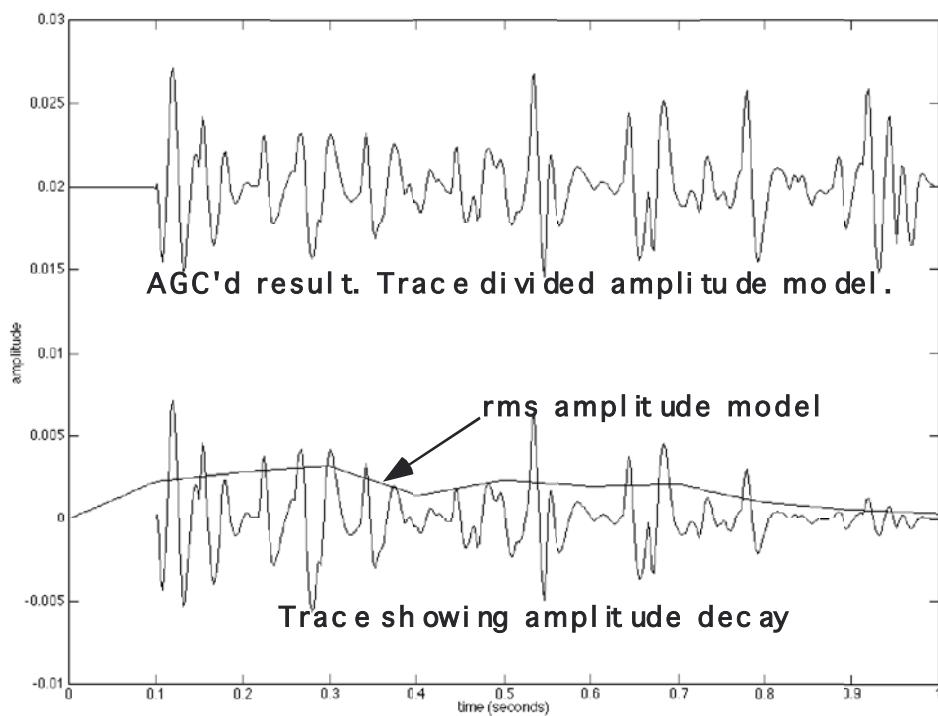
There are many AGC algorithms in common use. A simple, effective method involves the definition of a temporal window size and the measurement of the trace rms amplitude over that window. The window is then incremented and the measurement repeated. The result is a set of rms amplitude measurements at discrete times which defines an 'amplitude model' of the trace. This model is then linearly interpolated to the trace sample rate and the AGC'd trace is computed by dividing the original trace by the amplitude model.



Automatic Gain Correction (AGC)

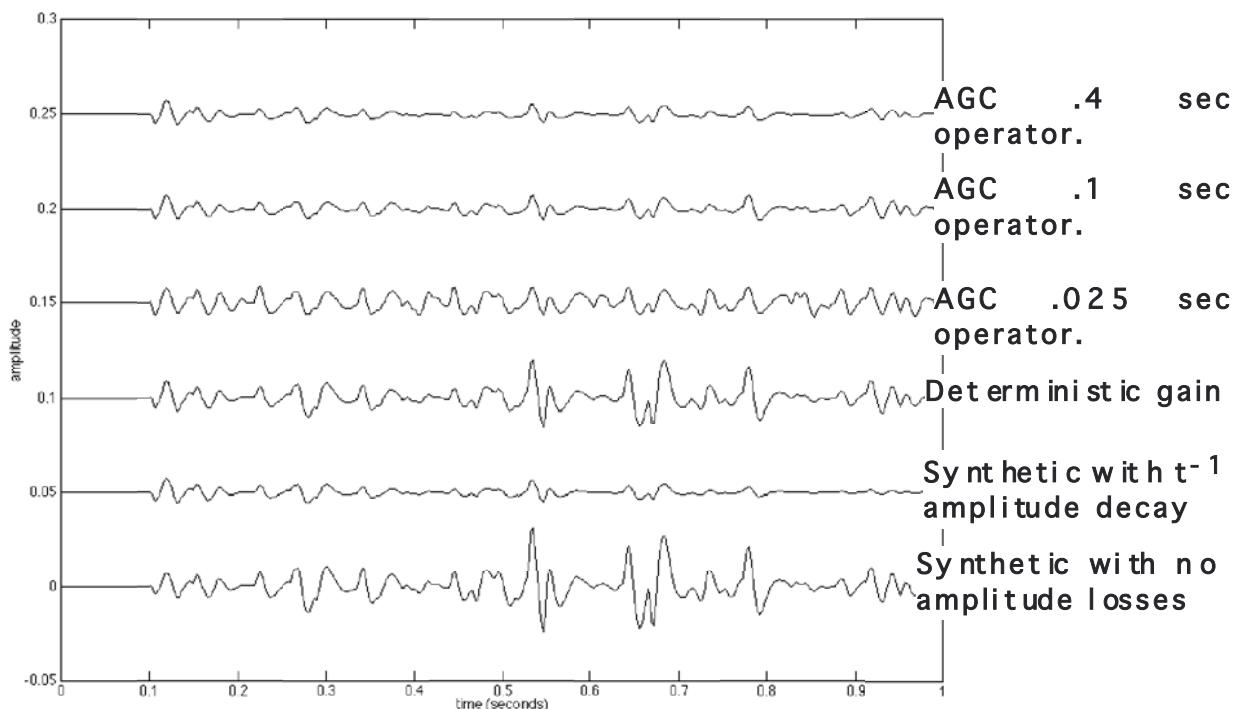


Above is the construction of an rms amplitude model from measures every .1 seconds and then interpolated. Below is the application of that model to the trace.



Automatic Gain Correction (AGC)

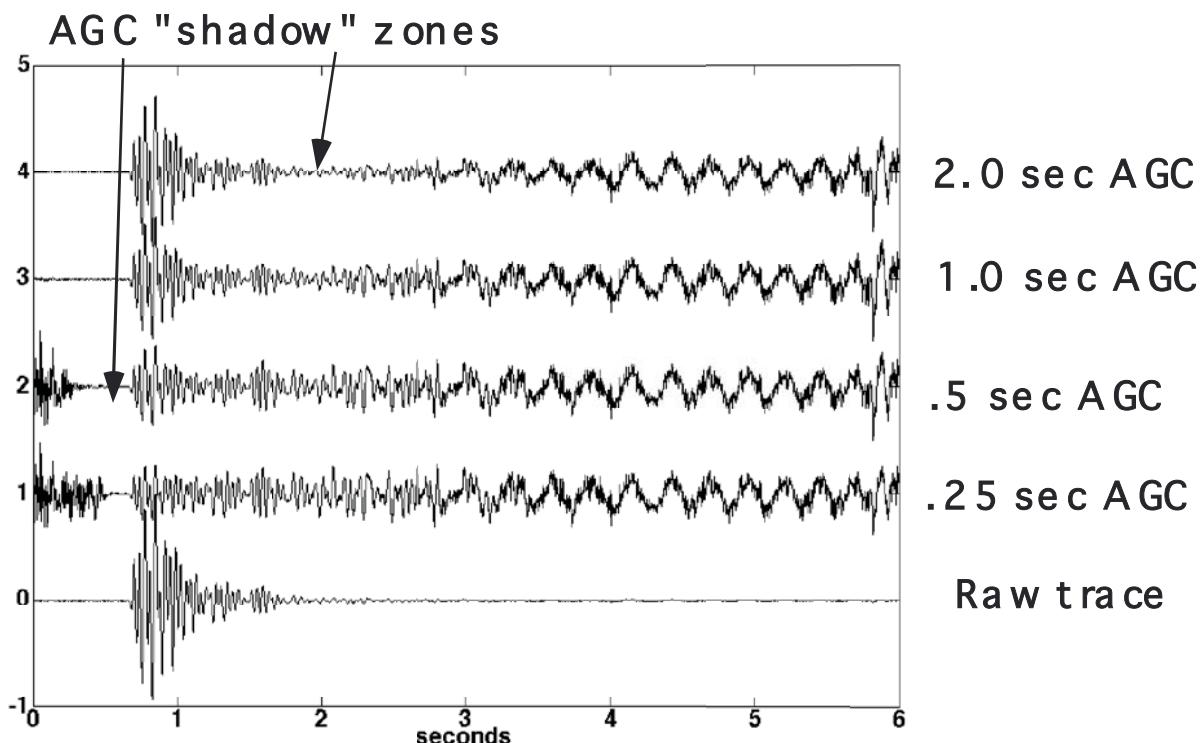
Below is a comparison of a deterministic amplitude restoration and several different AGC products. Note that the relative event 'standout' (the amplitude ratio between any two events) is best preserved by deterministic methods and secondarily by long AGC operators.



The two most common mistakes with AGC are to use it exclusively for all gain adjustments or to avoid it entirely. In the first case, AGC should be used with caution if the intended interpretation method places emphasis on reliable amplitude information. In the second case, AGC often leads to superior residual statics and velocity analyses side flows even when never used in a main flow.

Automatic Gain Correction (AGC)

A comparison of AGC operator lengths on a real, raw, seismic trace shows how the choice of operator length can drastically affect the event character. Also apparent is the effect known as an AGC shadow zone. This occurs when a package of energy (in the case the first breaks) has much higher amplitude than adjacent events. The adjacent energy tends to have a suppressed amplitude over roughly the length of the AGC operator.



A major concern when using an AGC is that serious distortions in the embedded wavelet can occur if the AGC operator length is shorter than the source waveform. This can result in a strong degradation of the performance of deconvolution algorithms. This will become more clear after the reader has studied deconvolution in the next chapter.

Trace Equalization (TE) or Trace Balancing

Traces from raw field records can often have wildly varying total (rms) power levels. There are many possible causes including: shot strength variation, geophone coupling variation, near surface geology changes, source-receiver offset, and more...

Even in cases where deterministic gain is preferred, some sort of trace balancing should still be performed. Otherwise, high rms power traces (which are often the noisiest traces), will dominate in stacking and cross-correlations.

A simple method is called trace equalization, or TE, and is usually synonymous with trace balancing. TE is a very simple process in which all traces are adjusted to have the same rms power level according to:

$$\text{output trace} = \text{input trace}/(\text{rms power of input trace})$$

A common variant of TE is to compute the rms power over a particular time zone instead of the entire trace. If the time zone varies in width, then care must be taken to normalize the rms power measures for this effect.

Caution should always be exercised when interpreting seismic plots where a trace equalization or AGC has been applied as an option in plotting. While this may be a convenience it means that the data display may not truly represent the data as stored on disk or tape. For example, data that is wildly unbalanced from trace-to-trace may appear to have good amplitude variation, leading to erroneous processing decisions.

Constant Q Effects

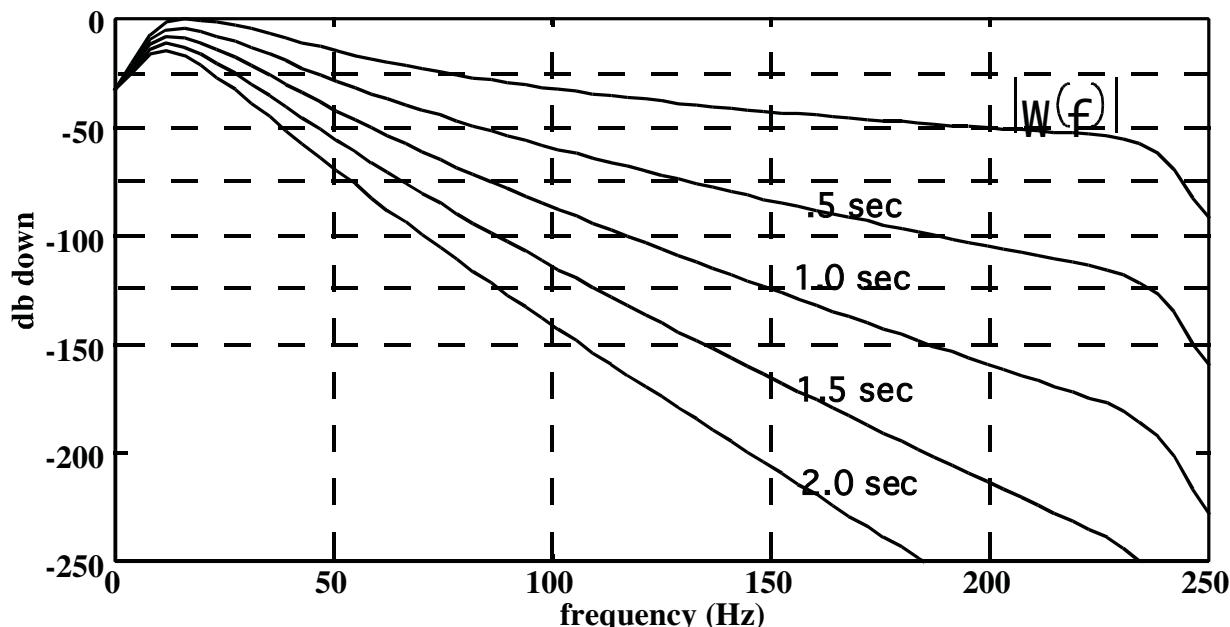
Strictly speaking, constant Q theory refers to a Q which is independent of frequency but may still depend on time. For simplicity, we will also assume time independence. Note that the attenuation can be written as:

$$\begin{aligned}\exp(-\pi ft/Q) &= \exp(-\pi fx/(vQ)) = \exp(-\pi x/(\lambda Q)) \\ &= \exp(-\pi n_\lambda/Q)\end{aligned}$$

where we have used $\lambda f = v$ and $n_\lambda = x/\lambda$ is the number of wavelengths that fit in the distance traveled. Thus, as a waveform propagates, it is constantly being attenuated with the higher frequencies being attenuated faster. If $W(f)$ is the spectrum of our source waveform, $w(\tau)$, then after propagating a time t , the amplitude spectrum of the propagating waveform has become:

$$|W_p(f)| = |W(f)| \exp(-\pi ft/Q)$$

If we assume $Q=50$, and a specific shape for $|W(f)|$, then we can compute the amplitude spectrum of the propagating waveform at any time:



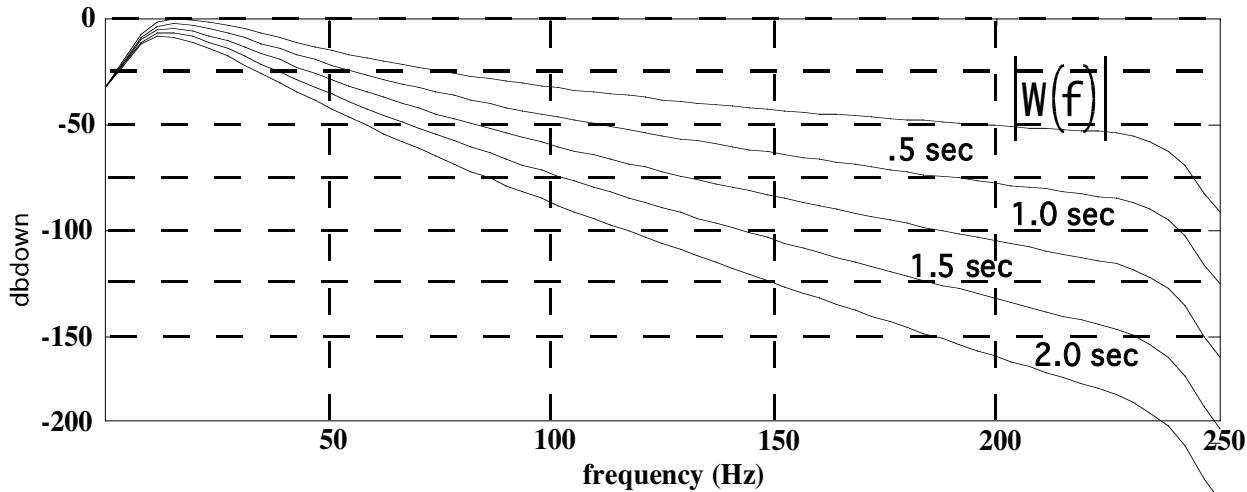
Constant Q Effects

Thus we see that seismic data must actually contain a wavelet with continuously decreasing bandwidth. This means the data signal spectrum is actually a function of time and is said to be nonstationary (or, equivalently, time-variant). Depending upon the value taken to characterize the background noise, we obtain these specific maximum signal frequency estimates (based on the preceding graph):

noise \ time	.5 sec	1.0 sec	1.5 sec	2.0 sec
100 db down	180 Hz	120 Hz	80 Hz	70 Hz
75 db down	130 Hz	80 Hz	70 Hz	55 Hz
50 db down	80 Hz	60 Hz	45 Hz	40 Hz
25 db down	45 Hz	35 Hz	30 Hz	25 Hz

Table showing predicted signal band for $Q=50$

Here is a repeat of the Q analysis for $Q=100$.



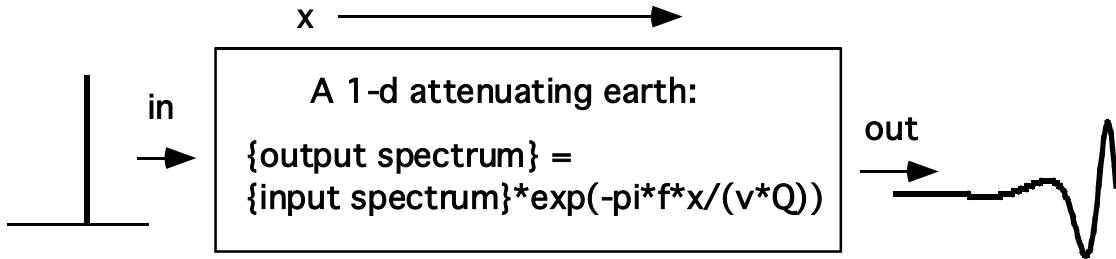
noise \ time	.5 sec	1.0 sec	1.5 sec	2.0 sec
100 db down	+200 Hz	185 Hz	140 Hz	120 Hz
75 db down	180 Hz	130 Hz	105 Hz	80 Hz
50 db down	110 Hz	80 Hz	70 Hz	60 Hz
25 db down	60 Hz	45 Hz	40 Hz	35 Hz

Table showing predicted signal band for $Q=100$

Constant Q Effects

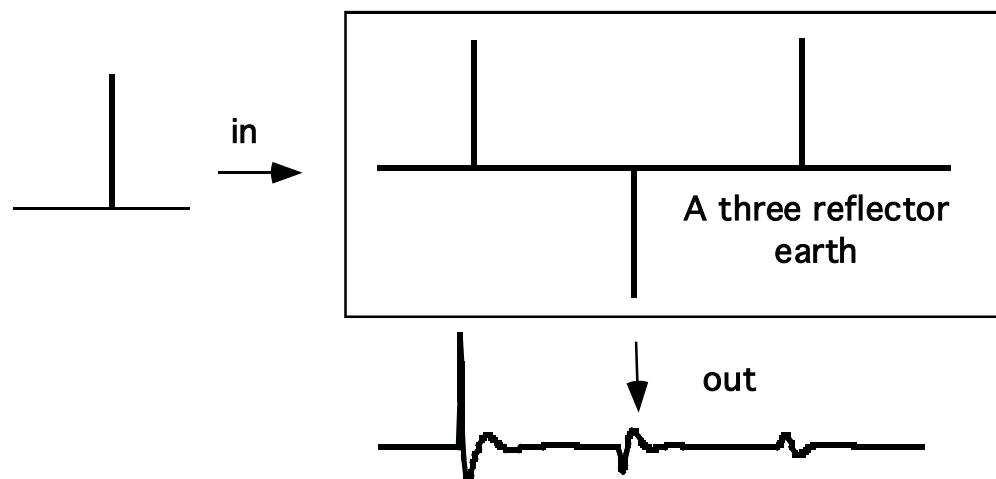
Thus far we have discussed the effects of attenuation on the amplitude spectrum of the propagating waveform but the phase effects are also dramatic.

Consider a 1-d earth with constant Q properties:



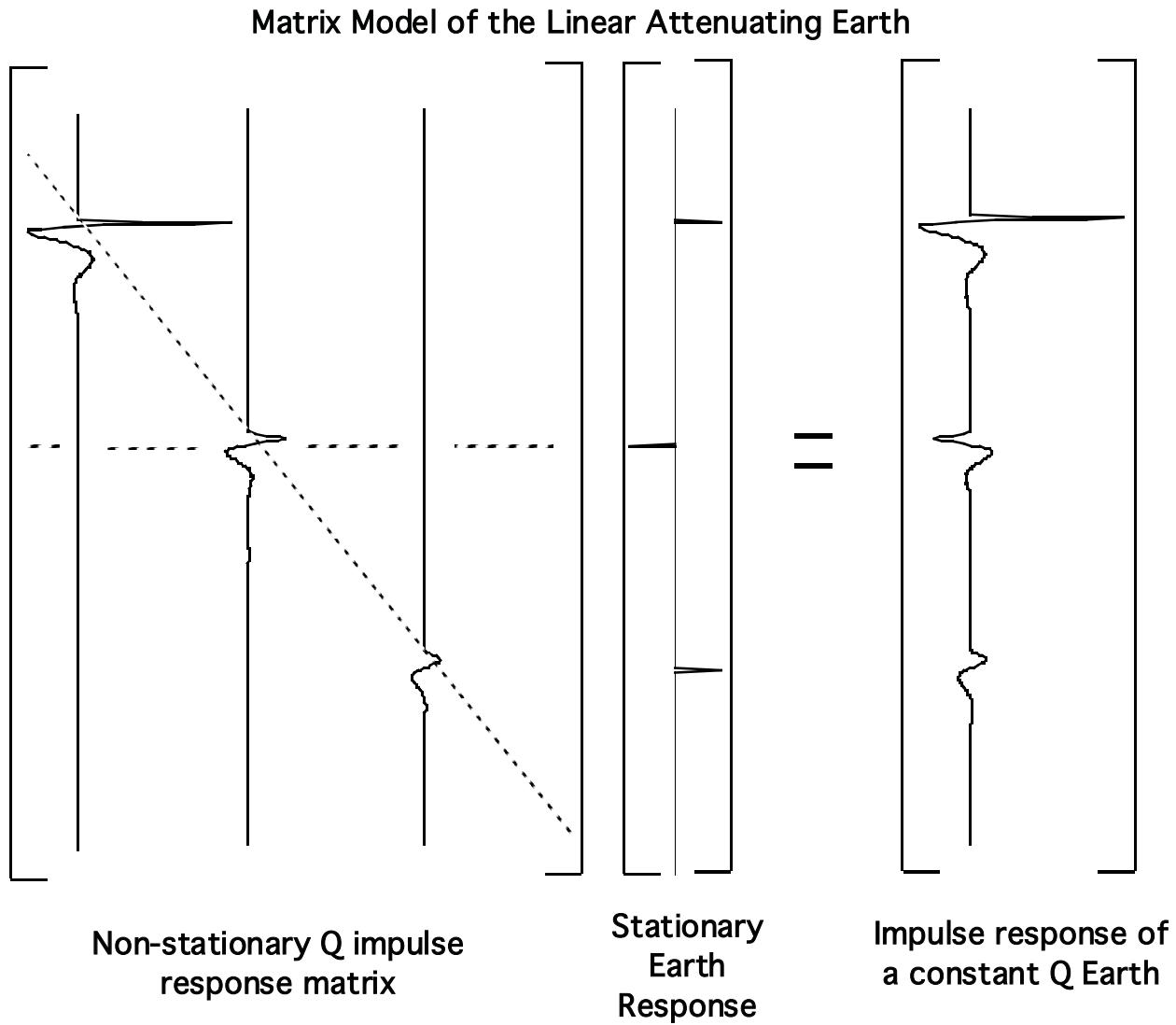
- An input impulse suffers attenuation at all non-zero frequencies
- The amount of attenuation is proportional to $x/v = t$
- The attenuation is necessarily coupled with minimum phase dispersion (Futteman, W.I., 1962, Dispersive Body Waves, JGR, 67, 5279-5291)

If the earth behaves linearly, then we can still argue that superposition holds. Thus the impulse response of an earth with reflectivity $\{r\}$ is the superposition of a set of delayed and progressively more attenuated waveforms:



Three superimposed waveforms showing increasing attenuation with increasing time.

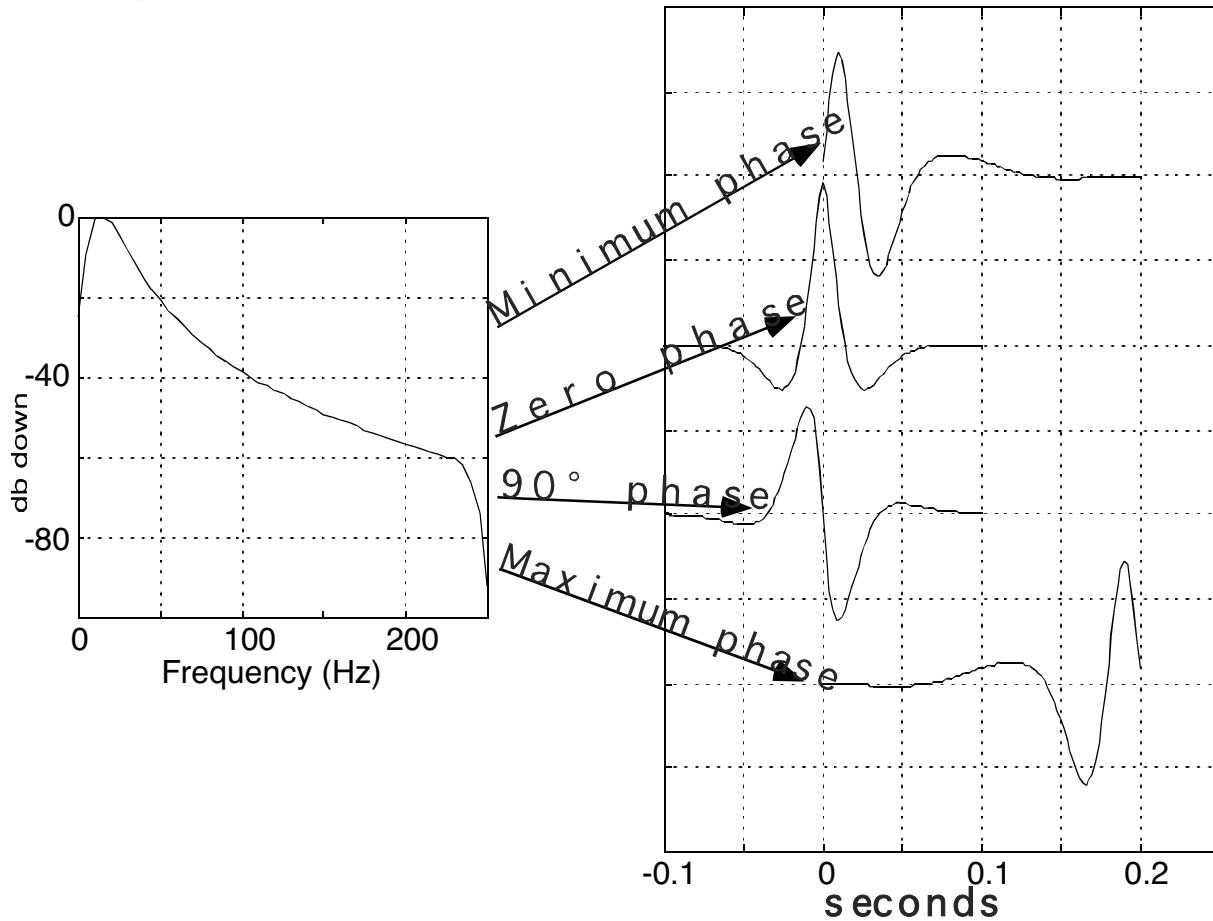
Constant Q Effects



The construction of a nonstationary multiple free synthetic seismogram is shown for a constant Q earth having 3 reflectors. The matrix multiplication shown here is performing a convolution as described on page 2-11. The convolution has been made nonstationary by changing the wavelet in each column of the convolution matrix.

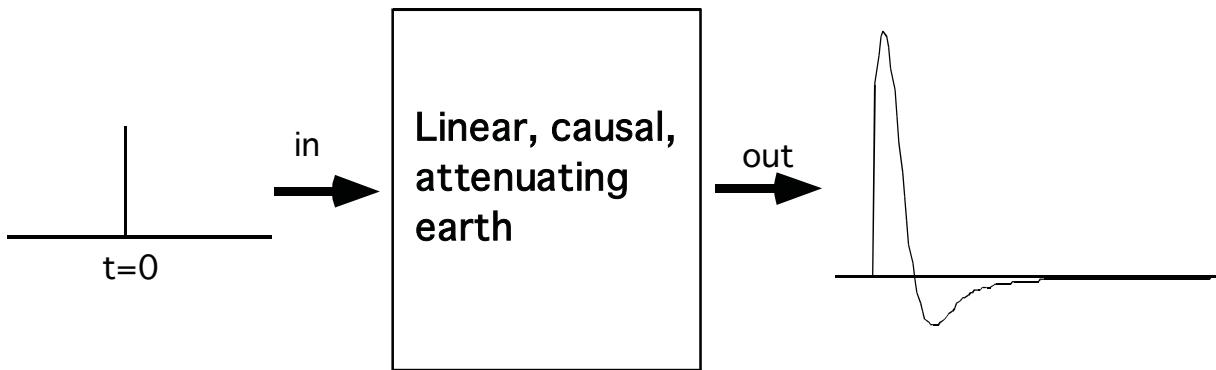
Minimum Phase, Intuitively

Infinitely many wavelets can be constructed which have the same amplitude spectrum by making different assumptions about phase.



However, only a few of these have any practical use. The minimum phase wavelet is distinguished from all others by being the most front-loaded of the "causal" wavelets.

Minimum Phase, Intuitively



Minimum phase wavelets arise naturally in the earth. Only the assumptions of causality and linearity are needed to show that attenuation in the earth is a minimum phase process. (Futterman, 1962, JGR vol 73, p 3917-3935)

The amplitude spectrum alone is sufficient to determine uniquely the minimum phase wavelet. The phase spectrum, $\phi(f)$, may be computed as:

$$\phi(f) = H(\ln(A(f)))$$

where H denotes the Hilbert Transform.

Minimum Phase, Intuitively

It is a common mistake to think that "minimum phase" refers to a particular phase spectrum which, if preserved, maintains a dataset's "minimum phaseness". We have just seen that this is not the case. Instead, minimum phase refers to a particular mathematical relationship existing between the amplitude and phase spectra so that knowledge of either one is sufficient to compute the other.

When a seismic dataset is said to be minimum phase, we generally mean that the embedded wavelet has this property and not the traces themselves. Certainly the earth's reflectivity function is not minimum phase.

True or False: If a dataset is minimum phase already, then a zero phase filter will preserve minimum phase because it does not change the phase in any way.

True or False: If the amplitude spectrum of a minimum phase dataset is changed, then the phase spectrum must also change to preserve the minimum phase relationship.

True or False: It has been proven beyond doubt that seismic data from impulsive sources is minimum phase.

True or False: All physical processes are minimum phase.

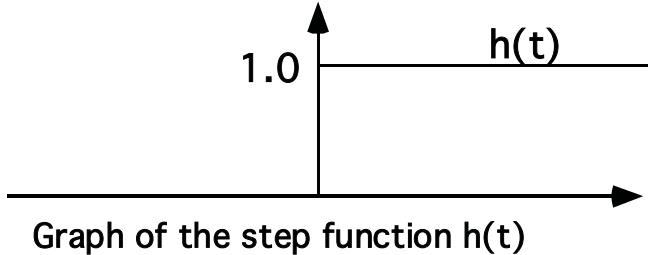
True or False: A minimum phase process can have zero amplitude over part of its spectrum.

True or False: A bandlimited process can never truly be minimum phase.

Minimum Phase and the Hilbert Transform

The concept of minimum is intimately linked with that of causality. For our purposes, a causal time series is one which vanishes for $t < 0$. The investigation of causal functions is facilitated by the following Fourier transform pair:

$$h(t) = \frac{1}{2} + \frac{1}{2}\text{sgn}(t)$$



Thus $h(t)$ is the unit causal function also called the step function.

The Fourier transform of the step function is: (See Papoulis, A., 1984, Signal Analysis, McGraw-Hill for a proof.)

$$H(\omega) = \pi\delta(\omega) - \frac{i}{\omega}$$

If $f(t)$ is any causal function, then:

$$f(t) = f(t)h(t)$$

Taking the Fourier transform of both sides of this result gives:

$$F_r(\omega) + iF_i(\omega) = \frac{1}{2\pi} \left(F_r(\omega) + iF_i(\omega) \right) \bullet \left(\pi\delta(\omega) - \frac{i}{\omega} \right)$$

Equating real and imaginary parts gives:

$$F_r(\omega) = \frac{1}{2}F_r(\omega) + \frac{1}{2\pi}F_i(\omega) \bullet \frac{1}{\omega}$$

$$F_i(\omega) = \frac{1}{2}F_i(\omega) - \frac{1}{2\pi}F_r(\omega) \bullet \frac{1}{\omega}$$

Minimum Phase and the Hilbert Transform

Thus, the spectrum of a causal function has its real and imaginary parts linked by the relations:

$$F_r(\omega) = \frac{1}{\pi} F_i(\omega) \bullet \frac{1}{\omega} \quad \text{and} \quad F_i(\omega) = -\frac{1}{\pi} F_r(\omega) \bullet \frac{1}{\omega}$$

If we write out the convolution integrals, we obtain:

$$F_r(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{F_i(\bar{\omega})}{\omega - \bar{\omega}} d\bar{\omega} \quad F_i(\omega) = \frac{-1}{\pi} \int_{-\infty}^{\infty} \frac{F_r(\bar{\omega})}{\omega - \bar{\omega}} d\bar{\omega}$$

These integrals are called Hilbert transforms and we say that the real and imaginary parts of a causal signal form a Hilbert transform pair. In our case, we actually want to relate the amplitude and phase of a causal signal to one another, not the real and imaginary parts. However, recalling that:

$$F(\omega) = A(\omega)e^{i\phi(\omega)} \Rightarrow \ln(F(\omega)) = \ln(A(\omega)) + i\phi(\omega)$$

The answer seems immediate that the phase and log amplitude spectrum are Hilbert transform pairs. However; we must ask:

- Under what circumstances can we take the log spectrum?
- Does the log spectrum still correspond to a causal time domain function?

The answer to the first question is that we can take the log so long as $A(\omega) \neq 0$. This is equivalent to saying that the time series $f(t)$ must have a stable inverse.

Minimum Phase and the Hilbert Transform

For the second question, consider the z transform of $f(t)$:

$$F(z) = \sum_{k=0}^{\infty} f_k z^k$$

We know that, since f is causal, then $F(z)$ contains no negative powers of z . Now consider the well known series expansion for the logarithm:

$$\ln(u) = (1-u) - \frac{(1-u)^2}{2} + \frac{(1-u)^3}{3} + \dots$$

which is valid in the region $0 < |u| < 2$. Since $F(z)$ contains only positive powers of z and $\ln(u)$ contains only positive powers of u , then $\ln(F(z))$ contains only positive powers of z and therefore is the transform of a causal time series. Therefore we conclude:

$$\phi(\omega) = \frac{-1}{\pi} \int_{-\infty}^{\infty} \frac{\ln(A(\bar{\omega}))}{\omega - \bar{\omega}} d\bar{\omega} \quad \ln(A(\omega)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\phi(\bar{\omega})}{\omega - \bar{\omega}} d\bar{\omega}$$

For our purposes, we state the following important result:

For a causal, stable function with a causal, stable inverse, the phase and log amplitude spectrum form a Hilbert transform pair. In particular, the phase may be computed as the Hilbert transform of the log of the amplitude spectrum. Such a function is said to be minimum phase.

$$\phi(\omega) = H(\ln A)(\omega)$$

Minimum Phase and the Hilbert Transform

We have demonstrated that a causal, stable time series with a causal, stable inverse is completely determined by either its amplitude or phase spectrum. Given one, the other can be computed. In particular, the phase spectrum is computed as:

$$\phi(\omega) = H \left(\ln(A(\omega)) \right)$$

where H denotes the Hilbert transform. We called such a waveform minimum phase. The reason for this name comes from a theorem (Robinson, E.A. and Treitel, S., 1980, Geophysical Signal Analysis, Prentice-Hall) which shows that, for all causal wavelets with the same amplitude spectrum, the minimum phase wavelet arrives the soonest with the most energy. Mathematically, this is stated by proving that the partial energies:

$$E_p = \sum_{k=0}^p f_k^2$$

are larger for the minimum phase wavelet than for any other wavelet for all p . This proof is equivalent to saying that the phase delay of the minimum phase wavelet is the smallest possible delay allowed by causality, for each frequency.

Recalling that the Hilbert transform is just a convolution with $1/\omega$, it follows that the minimum phase spectrum for any particular frequency is influenced by the amplitude spectrum at all frequencies. Put another way, a change to the amplitude spectrum at a particular frequency will change the minimum phase spectrum at all frequencies.

Minimum Phase and Velocity Dispersion

We have shown how to calculate the phase of a minimum phase wavelet given its amplitude spectrum and have also indicated that the constant Q attenuation model is minimum phase. Thus, if we represent a single propagating complex sinusoid as:

$$w(f, t, z) = A(f)e^{-i2\pi f(t-z/v)}$$

We can infer the velocity of this wave by following the motion of a point of constant phase. With complete generality, we can follow the point of zero phase by equating the phase to zero and solving for z/t . Thus we deduce its velocity to be $z/t = v$. If the same wave propagates through a constant Q medium, then we have:

$$w_Q(f, t, z) = w(f, t, z)e^{-\pi ft/Q + iH(-\pi ft/Q)}$$

or $w_Q(f, t, z) = A(f)e^{-\pi ft/Q}e^{-i2\pi f(t-z/v+\phi_Q(f))}$

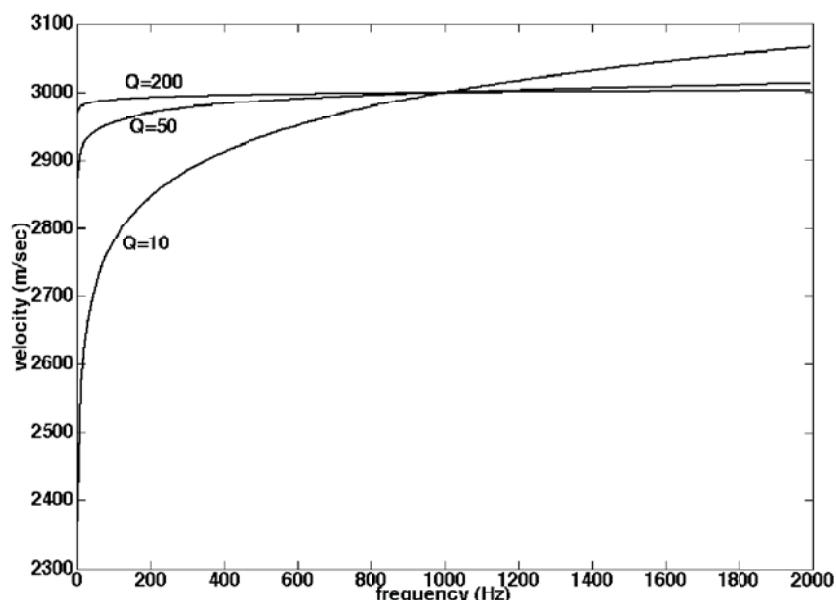
where $\phi_Q(f) \approx \frac{t}{\pi Q} \ln \left| \frac{f}{f_0} \right|$ (Kjartansson, 1979)

Then solving for the velocity:

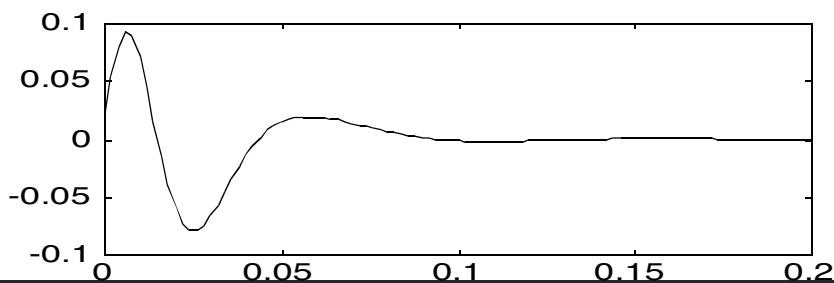
$$t-z/v + \frac{t}{\pi Q} \ln \left| \frac{f}{f_0} \right| = 0 \Rightarrow z/t = v(f) \approx v \left(1 + \frac{1}{\pi Q} \ln \left| \frac{f}{f_0} \right| \right)$$

Minimum Phase and Velocity Dispersion

Thus we see that, in an attenuating medium, velocity becomes frequency dependent, a phenomenon known as dispersion. The velocity dispersion predicted by this theory is strongest for low Q values. The figure below plots velocity versus frequency for different three different Q's. Note the nearly constant behavior for Q of 200 and the strong variation for Q of 10.

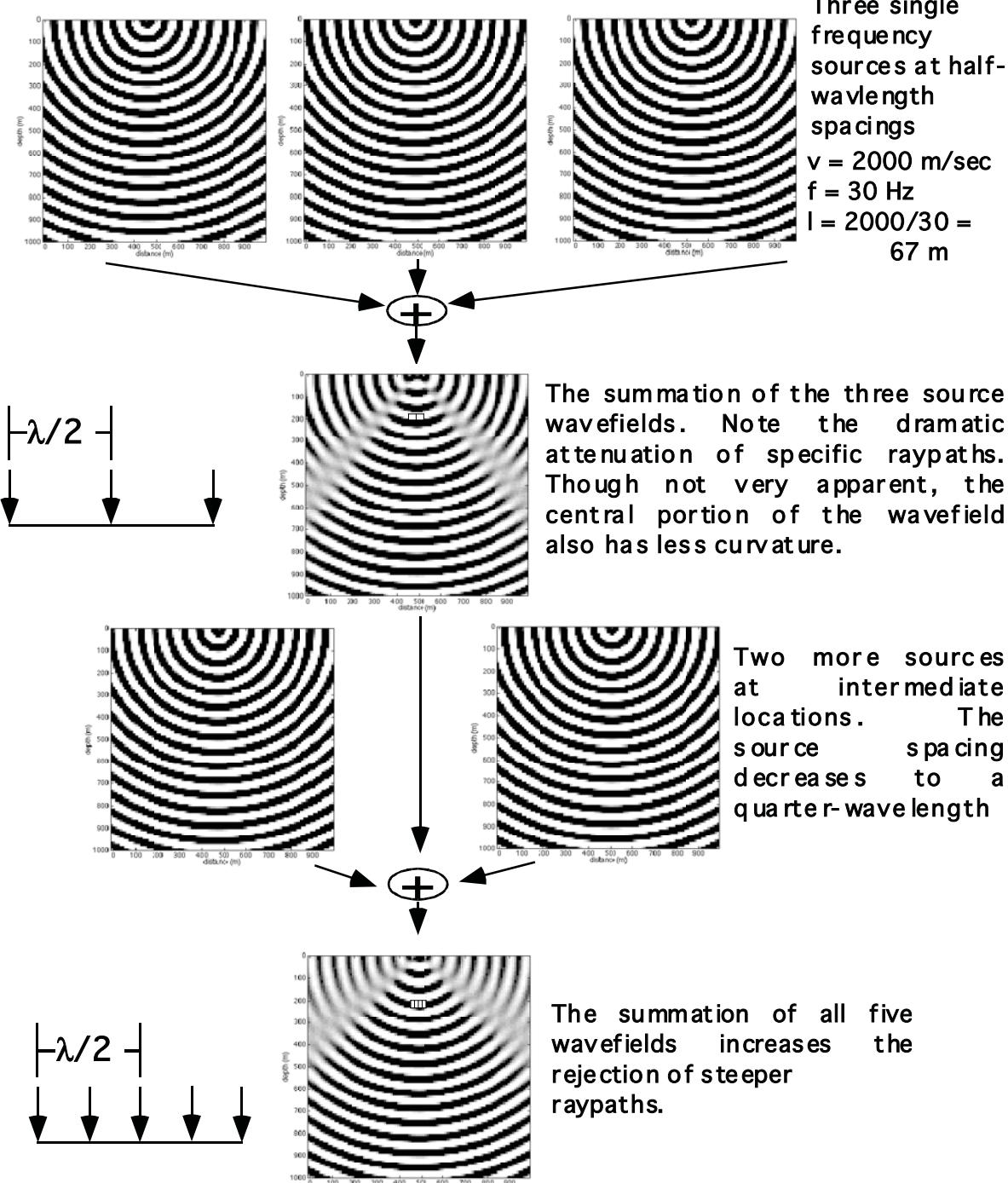


The word "dispersion" arises because a pulse will tend to spread out (disperse) as its various frequencies propagate at different velocities. It is this dispersion which leads to the characteristic pulse shape of a minimum phase wavelet. The pulse shape is strongly influenced by the near surface because it has dramatically lower Q values.

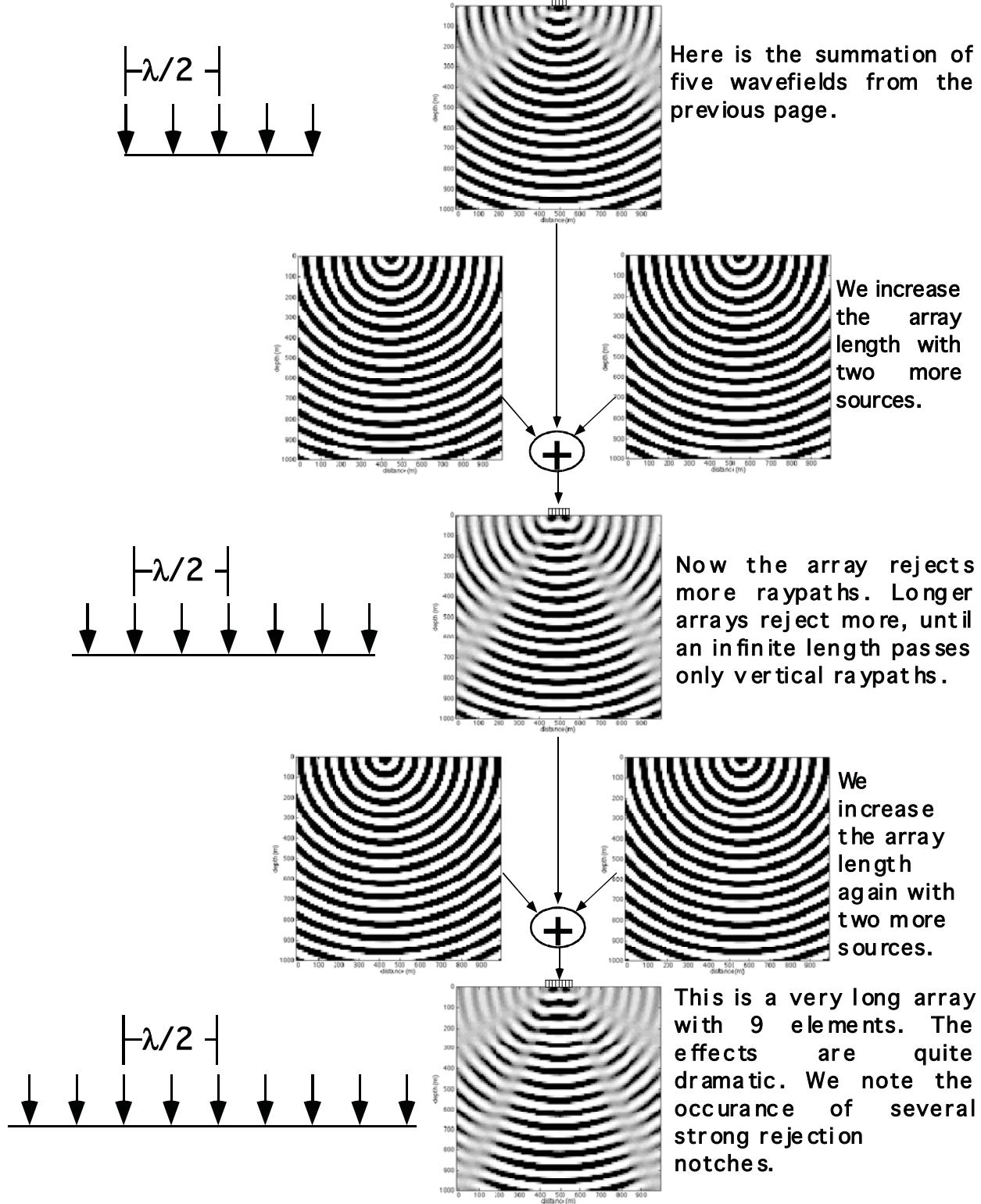


Array Theory

The use of arrays of sources and receivers is commonplace in exploration seismology. The essential details of their use are straight forward consequences of linear superposition and signal processing. Consider:

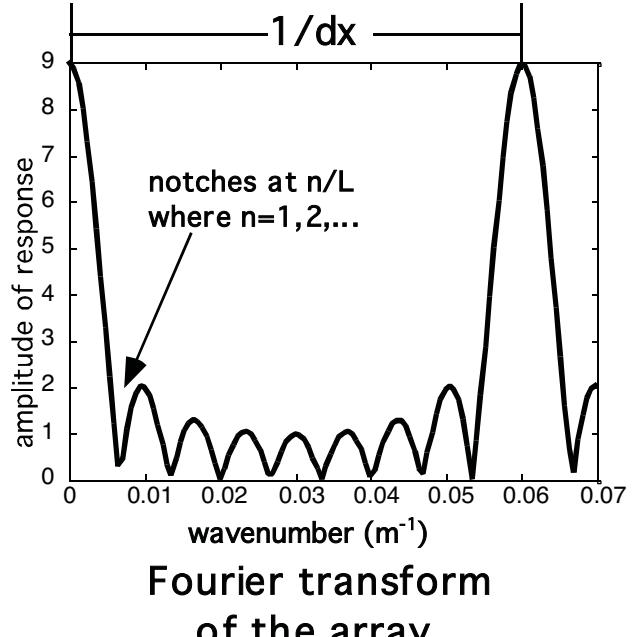
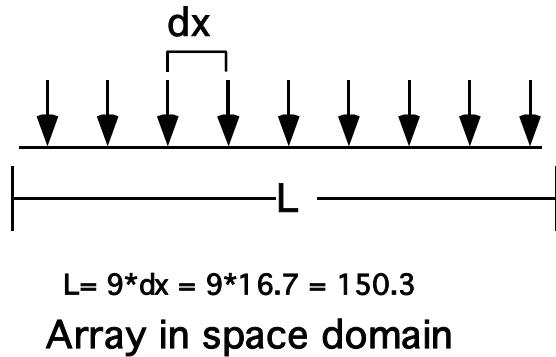


Array Theory



Array Theory

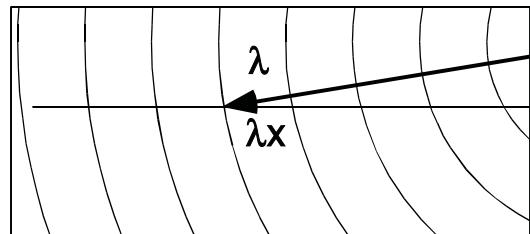
The response of an array can be analyzed by considering the idealized response of a sequence of unit spikes.



Since the array is purely a function of x , its response is purely a function of k_x . That is, it will be independent of k_z or f . However; in order to use the array response chart, we need a way to estimate k_x for an event of interest. We can do this by picking a horizontal surface of interest and measuring the horizontal apparent wavelength along it:

Since we usually don't have a monochromatic wavefield, then we usually measure apparent horizontal velocity and compute k_x from:

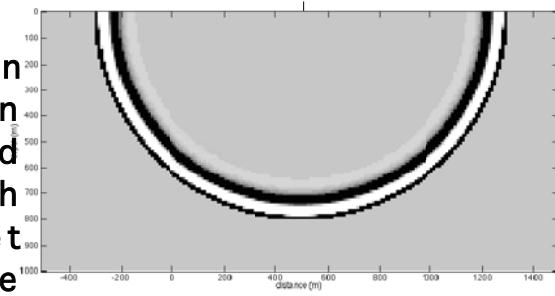
$$\frac{k_x}{f} = \frac{1}{v_a} = \frac{\sin(\theta)}{v}$$



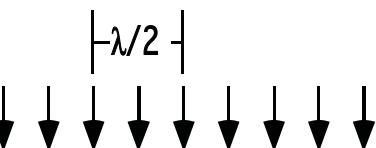
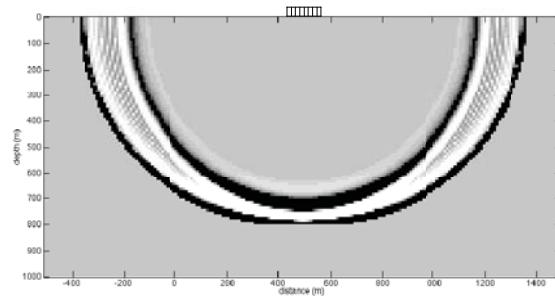
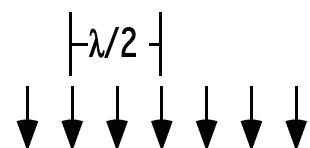
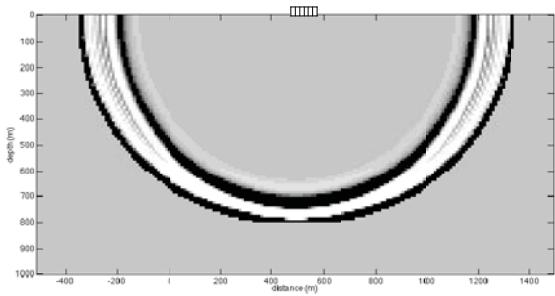
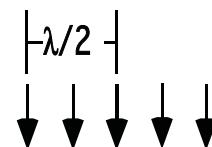
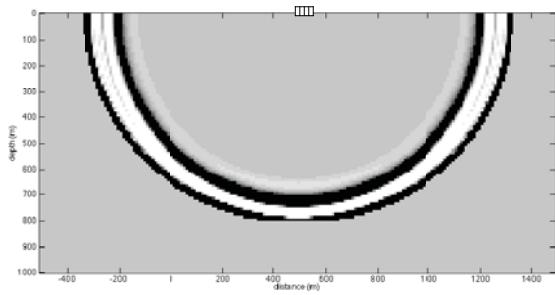
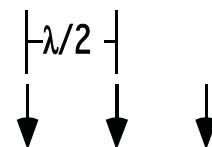
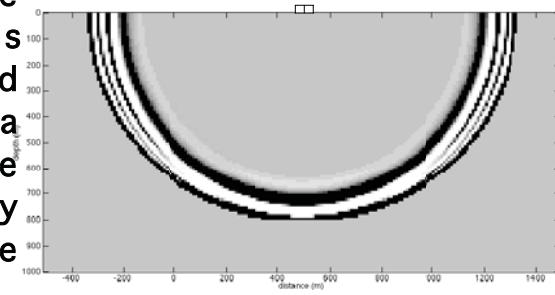
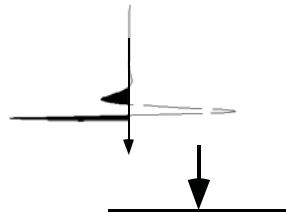
Thus, we must pick a frequency of interest to perform the analysis

Array Theory

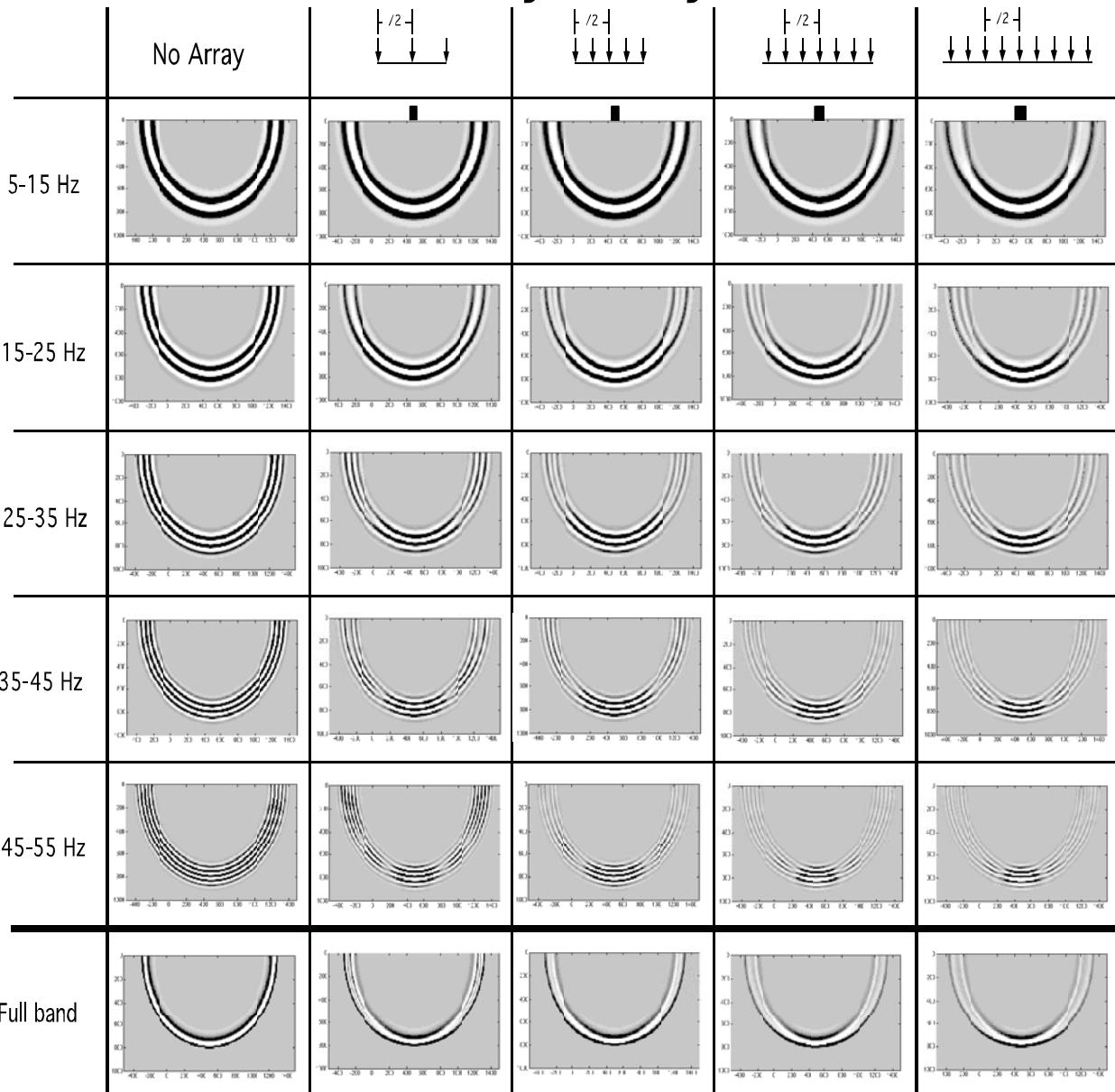
Here we see an array simulation for a broadband wavefront with the wavelet shown at the right. The same arrays as simulated previously for a 30 Hz single frequency source are shown.



Wavelet: 30 Hz,
Minimum phase

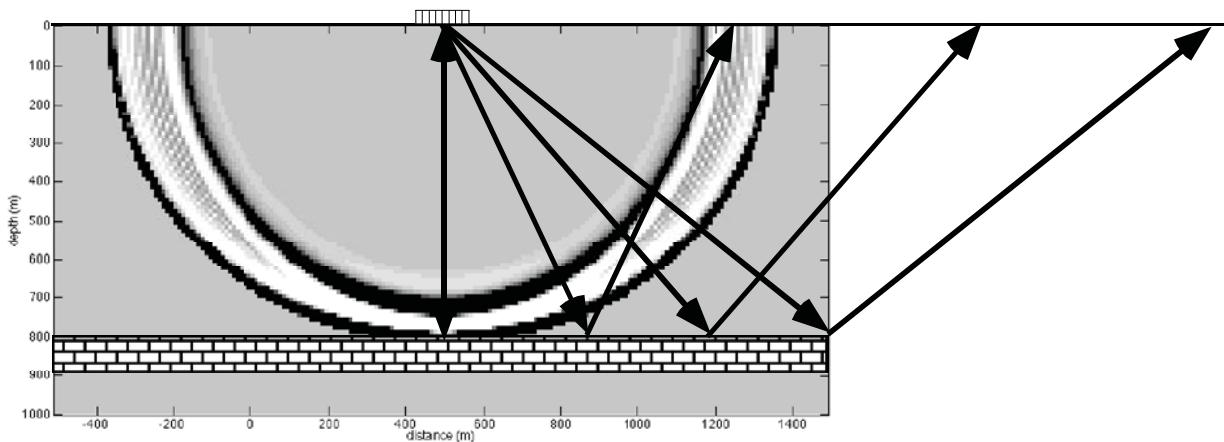


Array Theory



Here are broad-band snapshots of the simulation of an impulsive source and four different arrays. The small box at the top of each column gives the physical size of the array. Images are plotted with a slight vertical exaggeration and each wavefront is actually circular. Each source configuration is shown full-band and broken into five different sub-bands. The arrays always affect high frequencies more strongly and the longer arrays produce an undistorted waveform only for nearly vertical travel paths. The full-band images are the same as those on the [previous page](#).

Array Theory



A major effect of acquisition arrays is that they result in a variable (nonstationary) embedded wavelet. For a given reflector, the wavelet will vary with offset. For a given trace, the wavelet will vary with time. This has significant implications for deconvolution theory which assumes a stationary wavelet.

Methods of Seismic Data Processing

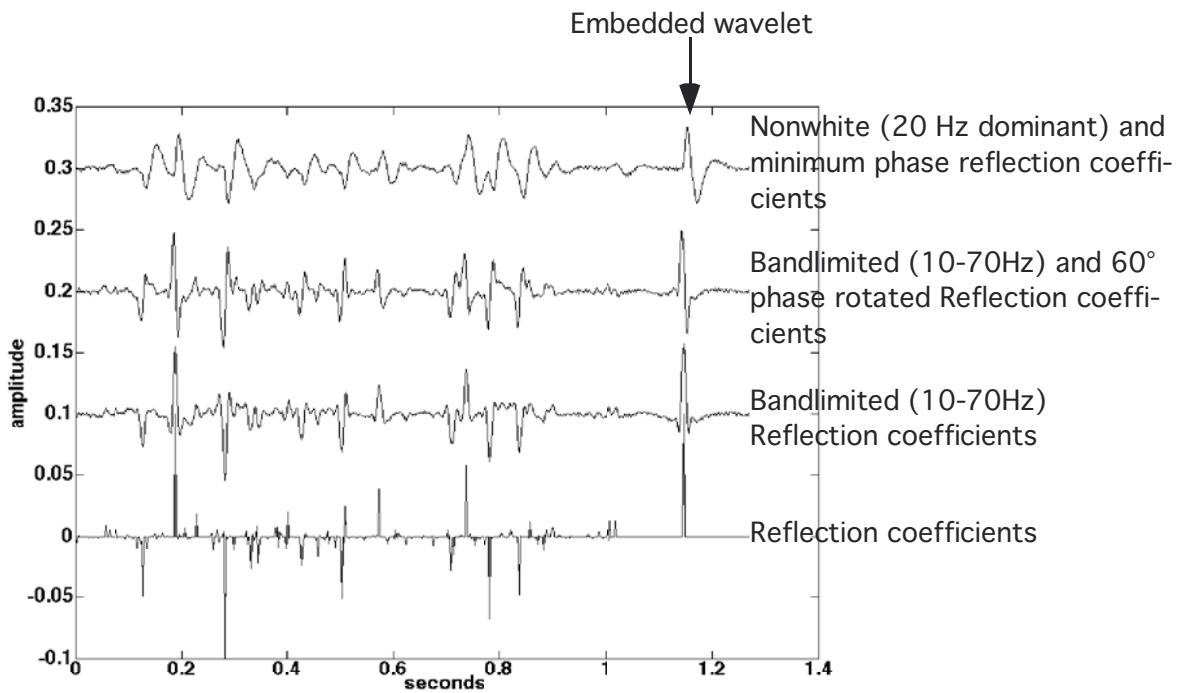
**Lecture Notes
Geophysics 557**

**Chapter 4
The Convolutional Model and
Deconvolution**

Bandlimited Reflectivity

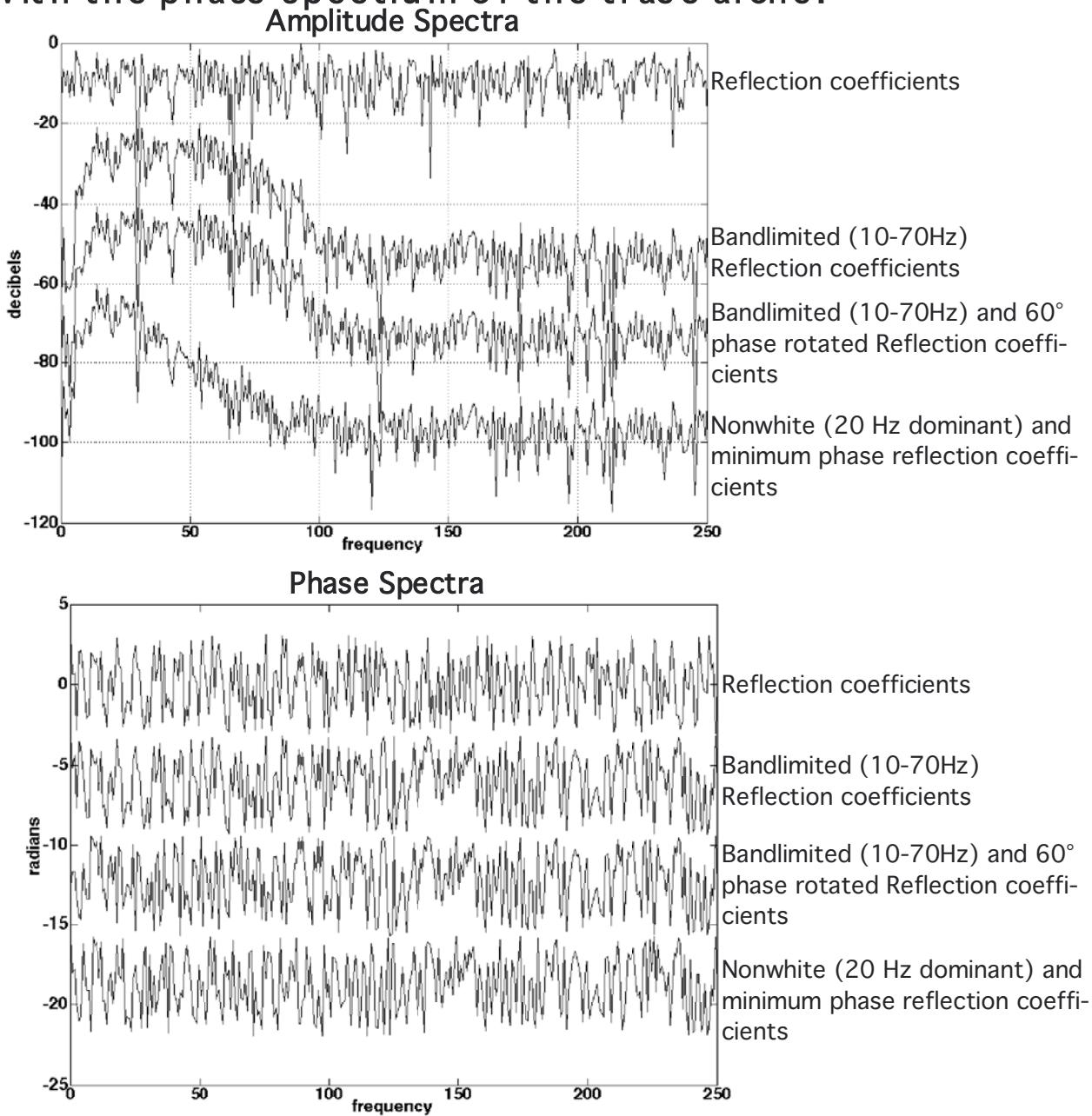
The ultimate goal of seismic data processing is to determine the earth's reflectivity as a function of position beneath the survey. Since seismic sources do not generate useful power at all frequencies, it is generally accepted that any reflectivity estimate must be "bandlimited". This means that the best possible result from fully processed seismic data is that it represents bandlimited reflectivity. We can think of this result as being the true (broadband) reflectivity convolved with a zero phase wavelet.

Even this modest goal is rarely fully realized. Deconvolution is one of our major tools for achieving this end. Shortcomings in our theory and algorithms and lack of knowledge to guide them usually means that our final estimate will have some undesired phase rotation or an incorrect amplitude spectrum. The figures below and on the next page illustrate these concepts.



Bandlimited Reflectivity

The consequence of a limited frequency band is loss of resolution. That is we cannot distinguish closely spaced reflectivity spikes. An unknown phase rotation makes it difficult to determine the precise location of a reflectivity spike or its amplitude. Note that the presence of a phase rotated wavelet cannot be detected with the phase spectrum of the trace alone.



The Convolutional Model

The majority of the theory of the deconvolution of seismic data is based on a series of simplifying assumptions concerning the nature of that data. These assumptions are usually encapsulated and referenced as "The Convolutional Model". We have already seen that, in a linear 1-D earth, we can write the the construction of a synthetic seismogram as a convolution of a source waveform and an impulse response:

where: $s(t) = I_r(t) \bullet w_s(t)$

$I_r(t)$ is the earth impulse response

$w_s(t)$ is the source waveform

$s(t)$ is the earth response to the source waveform

The assumption of linearity simply means that a linear combination of solutions to the governing 1-D wave equation is also a solution. While this is an important result from physics, for the purpose of providing a base for deconvolution theory, it is practically useless. The problem is that ALL of the physics and geology of the problem is contained in the impulse response. That is, if we consider an attenuating earth, with multiples and transmission losses, then all of these effects are contained in the impulse response. In fact, the convolutional result above, is valid in 2-D or 3-D and therefore the impulse response can also contain such effects as elastic mode conversions and spherical divergence in addition to those already mentioned. So, although this result can be proven from a very general theory, it is too general to be of use to us. Instead, we must make a number of simplifying assumptions to frame the context of deconvolution theory.

The Convolutional Model

Sheriff and Geldart (Exploration Seismology, 1995, Cambridge University Press) present the convolutional model by decomposing the earth's impulse response as:

$$I_r(t) = n_s(t) \bullet p(t) \bullet e(t)$$

where

$n_s(t)$ represents near surface effects beneath both the source and receiver

$p(t)$ represents all effects not otherwise modeled such as multiples, absorption, mode conversions, etc.

$e(t)$ is the "impulse response" (their term) of the target reflectors. "this is the signal that seismic reflection work is intended to find".

This terminology illustrates some of the typical confusion surrounding the convolutional model. Consider their definition of $e(t)$. If it is truly the impulse response of the target reflectors then it contains all multiples, absorption, mode conversions as well as primaries from that zone. This means it is NOT the signal we wish to uncover and thus their definition is self-contradictory. Also, $p(t)$ is supposed to be a convolutional operator which models a diverse range of effects without any justification that this is even possible. In fact, most of the mentioned effects are nonstationary (see 2-12 for a definition) and therefore cannot be modeled as a convolution. This is the presentation in an excellent, highly regarded reference work so it is understandable that there is a great deal of confusion surrounding the convolutional model in the industry.

The Convolutional Model

We now modify the model of Sheriff and Geldart with the intent of preserving its spirit but making it logically consistent. First we combine the source waveform and the near surface effects into an equivalent wavelet:

$$w_e(t) = w_s(t) \bullet n_s(t)$$

Next we discard $p(t)$ as containing nonstationary effects which are beyond the scope of the model and allow $e(t)$ to be an impulse response in a limited sense of the target reflectors:

$$s(t) = w_e(t) \bullet e(t) + \text{noise}(t)$$

Here we have also introduced additive, stationary, white noise. The earth's impulse response is further assumed to be:

$$e(t) = m(t) \bullet r(t)$$

where:

$r(t)$ = the earth's primary reflection series

$m(t)$ = the subset of the earth's multiple reflection response which can be modeled as a stationary process.

Thus we can write:

$$s(t) = w_e(t) \bullet m(t) \bullet r(t) + \text{noise}(t)$$

The Convolutional Model

Note that the multiple term can be equally well associated with the wavelet instead of the reflectivity so that we can write:

$$s(t) = w_m(t) \bullet r(t) + \text{noise}(t)$$
$$w_m(t) = w_e(t) \bullet m(t)$$

This result is a good starting point for deconvolution theory since it presents the seismic trace as the convolution of a wavelet with the earth's reflectivity. It is emphasized that our goal is deducing the earth's reflectivity and NOT its impulse response. The two are very different.

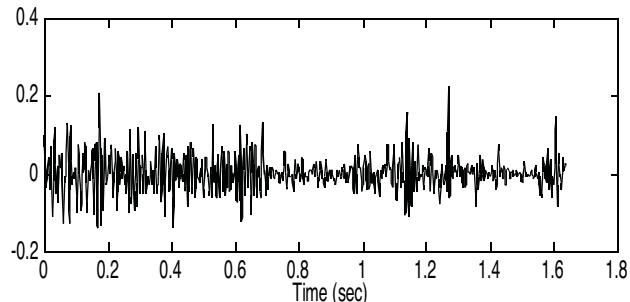
We remarked that the noise is modeled as being "stationary" and "white" in nature. Stationary in this context means that the basic features of the spectrum do not change with time. That is, if we extracted spectra from small windows ranging up and down $n(t)$ we would find essentially the same spectral shape. Gaussian or uniformly distributed noise can be shown to have this property. The convolution of two stationary signals is also stationary. An example of a nonstationary signal is the impulse response from a constant Q earth. As we have seen, the spectral response changes systematically with time.

A white spectrum is one that has constant power at all frequencies (e. g. "white noise"). An infinite length signal of Gaussian or uniformly distributed noise can be shown to have a white spectrum. Finite length noise sequences have approximately white spectra when smoothed with a short operator.

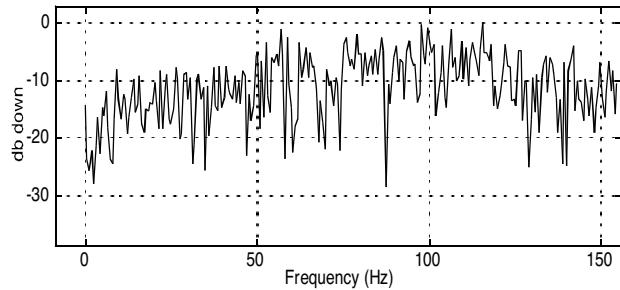
The Convolutional Model

Though not strictly part of the convolutional model, a further assumption is often made (in the context of deconvolution theory) that the reflectivity, $r(t)$, is a white and stationary time series. It can be easily demonstrated using sonic logs that real earth reflectivity does not have a white spectrum but instead shows considerable spectral color evidenced by a pronounced rolloff in power at the low frequencies.

Here we see an example of a real reflectivity (in time, computed from a sonic log assuming constant density) and its Fourier spectrum:

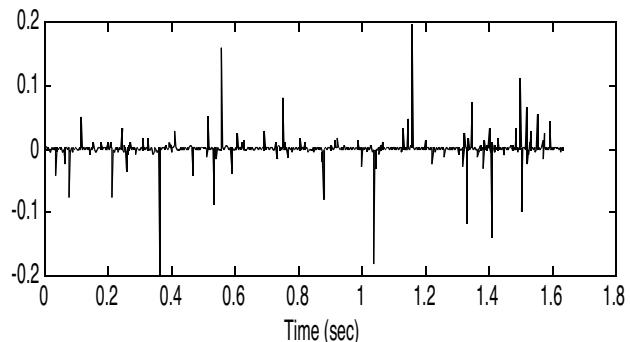


Real rcs computed from a sonic log at constant density

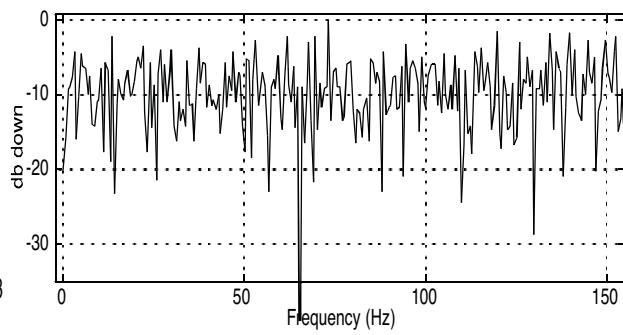


Spectrum of the real rcs. Note the 20db rolloff from 100 to 0 Hz.

Contrast this with a computer generated random reflectivity designed with a white spectrum:



Computer generated pseudo random rcs.



Spectrum of the pseudo random rcs. Note the essentially flat (white) spectrum.

The Convolutional Model

In our basic convolutional model, we assumed the effects of multiples could be treated as a convolution of the source waveform with a "multiple operator":

$$w_m(t) = w(t) \bullet m(t)$$

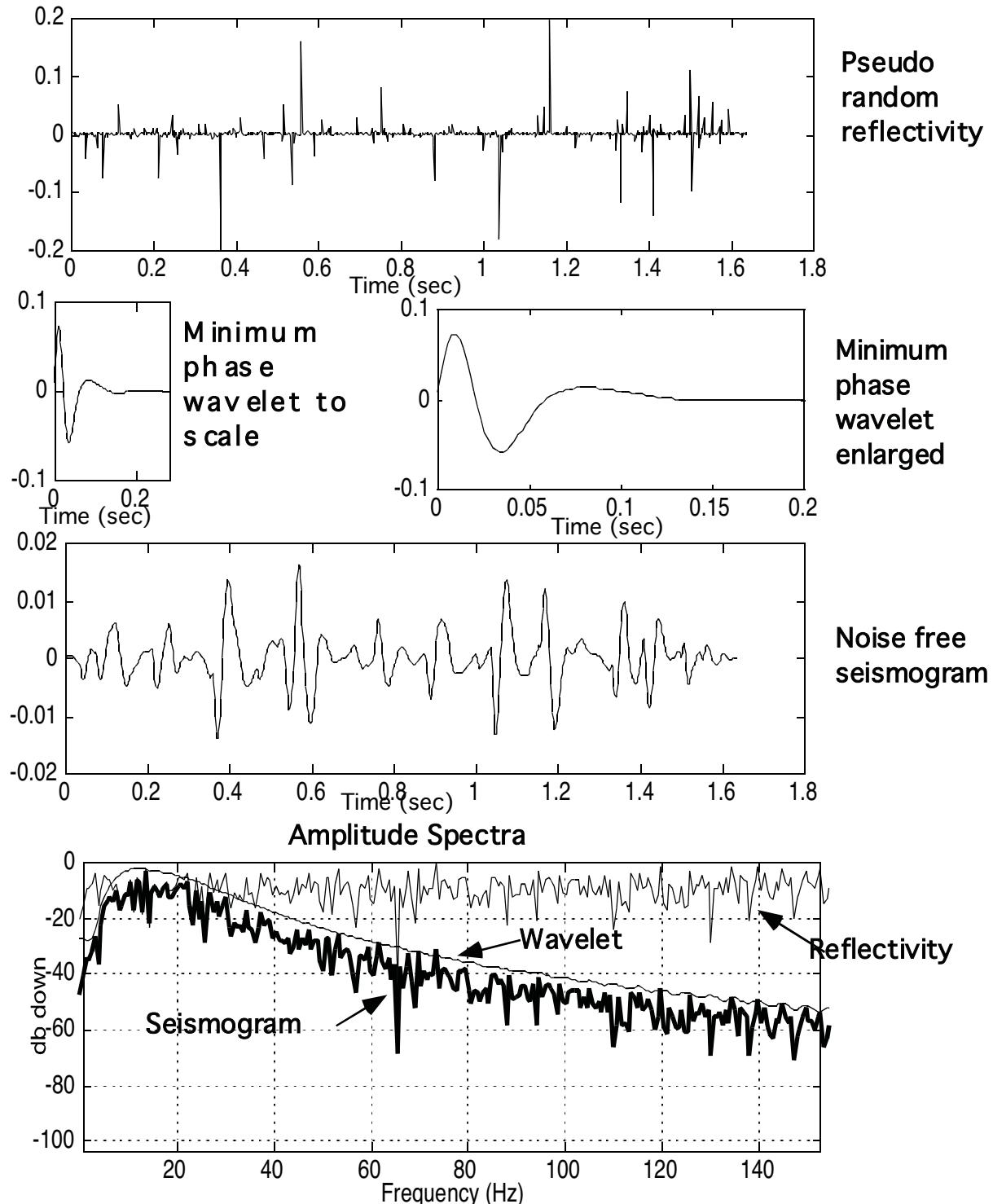
In our development of the 1-D seismogram, we examined an algorithm which is capable of generating all possible multiples. Could this operation have been performed as a convolution? The general answer to this question is "no" because the multiple train grows in length as time increases and is thus fundamentally non-stationary. However, certain classes of multiples can be modeled by a convolution including surface ghosts and water bottom multiples. In general, if we restrict our attention to the portion of an impulse response later in time than a major multiple generator, then the multiple contribution from that generating interface can be modeled as the convolution of a multiple operator with the source waveform. However, as another caveat, even water bottom multiples on far offset traces show non-periodic spacing and so violate our model.

Summary of assumptions:

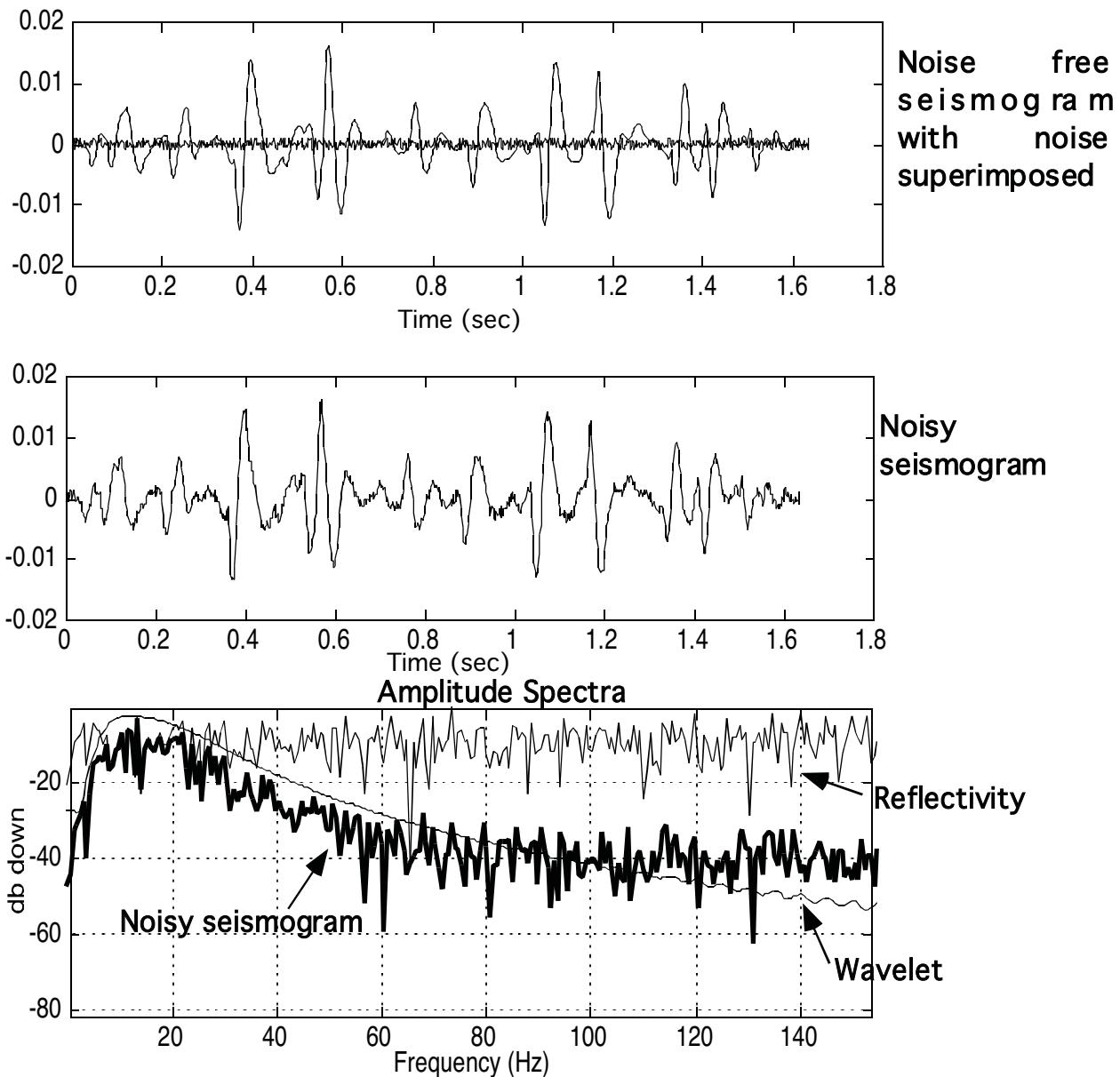
- Earth's impulse response consists of a reflectivity series possibly convolved with a multiple operator. It is also stationary.
- The effect of the source waveform may be modeled as a simple stationary convolution with the earth's impulse response.
- Any noise is additive, white, and stationary.
- Optionally, Earth's reflectivity series is white and stationary.

The Convolutional Model

Here we illustrate the steps involved in the construction of a multiple-free synthetic seismic trace using a pseudo random reflectivity:



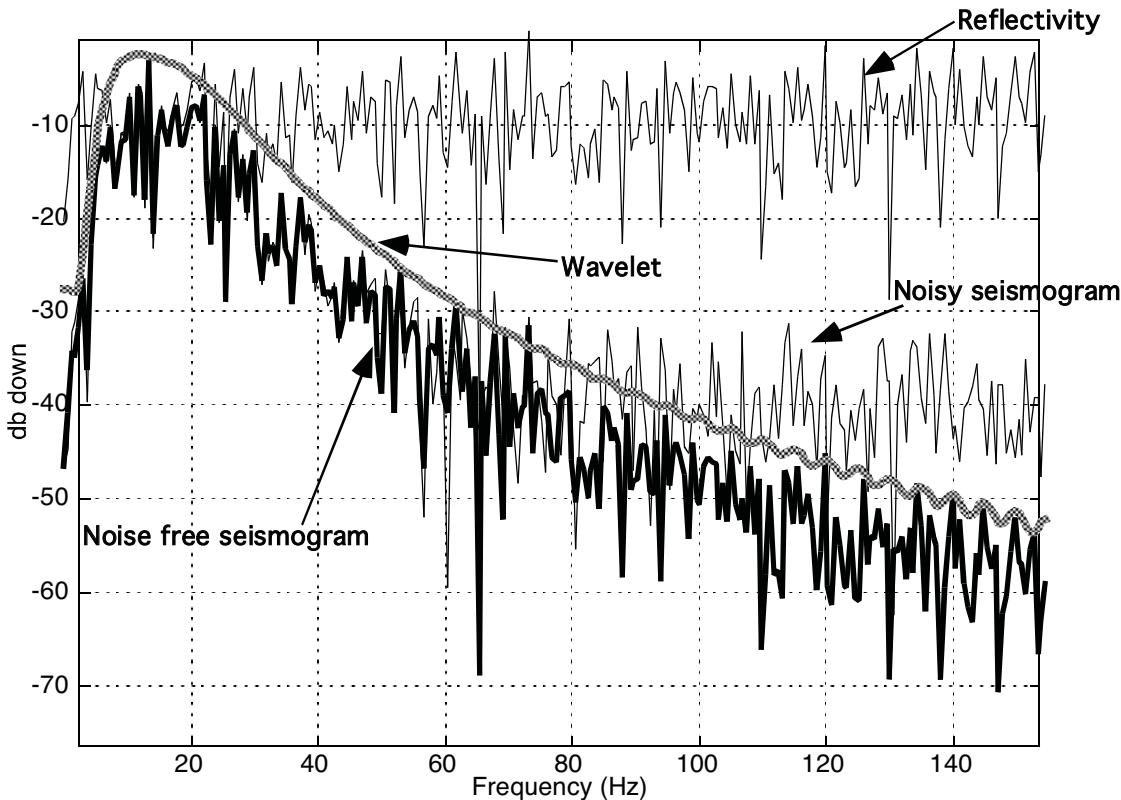
The Convolutional Model



Definition: “The embedded wavelet”. As you now know, there are many “wavelets” in exploration seismology. The phrase *embedded wavelet* refers to a wavelet derived by fitting any seismic trace to the convolutional model. That is, the embedded wavelet is whatever signal must be convolved with the reflectivity to give the trace under consideration. Even when the convolutional model is a poor fit to the data, an embedded wavelet can still be estimated in the least-square sense.

Frequency Domain Spiking Deconvolution

Perhaps the easiest deconvolution technique to conceptualize is the frequency domain method. It is suggested by the spectra we examined in our discussion of the convolutional model:



Here we see the basic idea that underlies all deconvolution concepts: The amplitude spectral shape of the seismic trace (seismogram in this case) is essentially similar to that of the unknown wavelet. Given this, all that remains is to deduce the wavelet's phase and then we can design an inverse for it. We observe that the noisy seimogram introduces a further complication in that we must restrict our attention to the signal frequency band.

Note that we are relying on the reflectivity to have a white spectrum so that we can attribute all spectral "character" to the wavelet.

Frequency Domain Spiking Deconvolution

If we can compute the amplitude spectrum of the wavelet by smoothing the amplitude spectrum of the seismic trace, then we can invoke the minimum phase assumption to completely specify the unknown wavelet. Here is the help file from the Matlab routine, deconf, which does frequency domain deconvolution:

```
%  
% [trout,specinv]=deconf(trin,trdsign,n,stab,phase)  
% [trout,specinv]=deconf(trin,trdsign,n,stab)  
% [trout,specinv]=deconf(trin,trdsign,n)  
%  
% DECONF performs a frequency domain deconvolution of the  
% input trace  
%  
% trin= input trace to be deconvolved  
% trdsign= input trace to be used for operator design  
% n= number of points in frequency domain boxcar smoother  
% stab= stabilization factor expressed as a fraction of the  
% zero lag of the autocorrelation. This is equivalent to being  
% a fraction of the mean power.  
% ***** default=.0001 *****  
% phase= 0 ... zero phase whitening is performed  
% 1 ... minimum phase deconvolution is performed  
% ***** default= 1 *****  
%  
% trout= output trace which is the deconvolution of trin  
% specinv= output inverse operator spectrum. The time domain  
% operator can be recovered by real(ifft(fftshift(specinv)))
```

- deconf algorithm
 - Compute the power spectrum of the design trace.
 - Add in the stab power.
- Convolve the power spectrum with a boxcar smoother to estimate the wavelet power spectrum.
- Compute the wavelet phase spectrum with the Hilbert transform.
- Compute the spectrum of the input trace.
- Divide the input trace spectrum by the estimated wavelet spectrum.
- Inverse FFT to give deconvolved trace.

We note that the deconvolution operator can be designed on one trace and applied to another. This is to simulate the practice of designing the operator on a segment of the trace to avoid letting such things as surface waves influence the design. The other significant parameters are: the length of a boxcar smoother, a stabilization factor, and a flag for zero or minimum phase. In order to specify n , we recall that the frequency sample size of the DFT spectrum is $\Delta f = 1/T$ where T is the trace length in seconds. Thus, a smoother of length F_{smooth} (in Hertz) will have a number of points given by:

$$n_{\text{smooth}} = \frac{F_{\text{smooth}}}{\Delta f} = T F_{\text{smooth}}$$

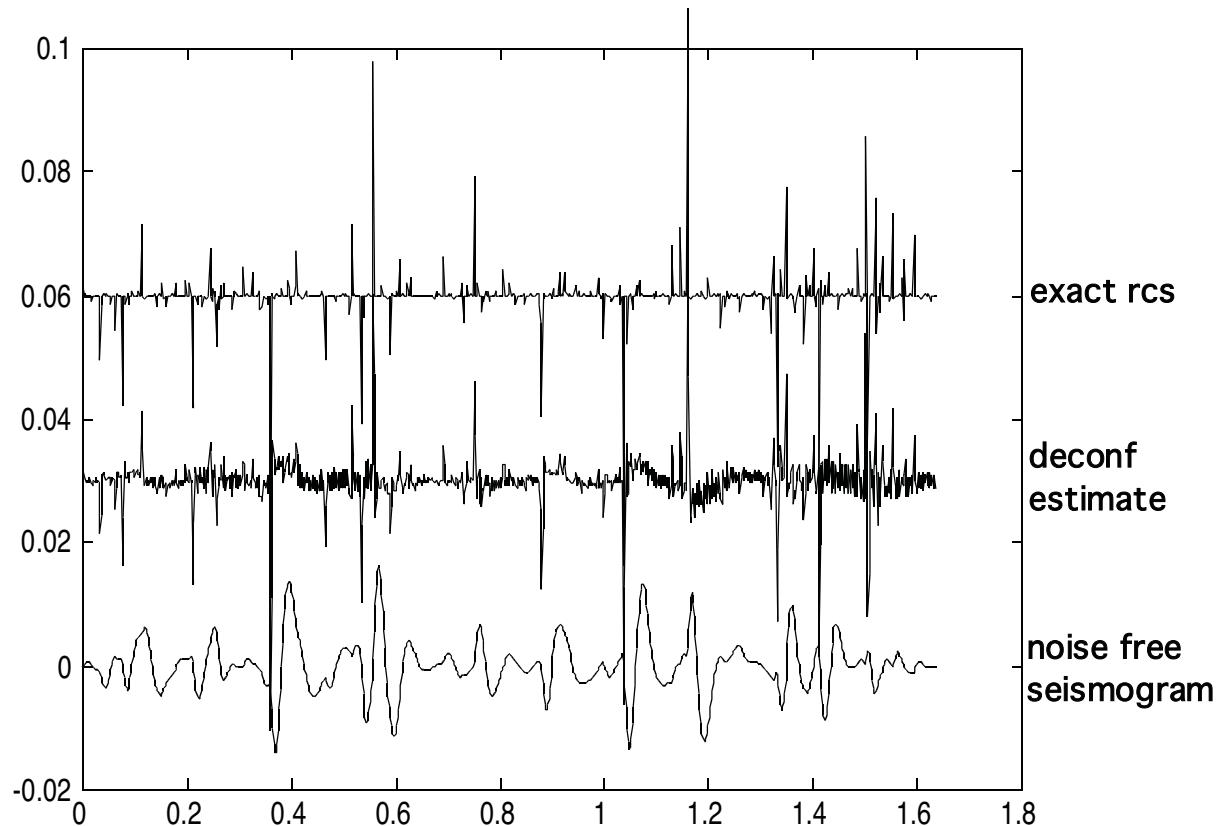
Frequency Domain Spiking Deconvolution

The stabilization factor is designed to prevent the operator design from being unduly influenced by noise and to avoid the possibility that a division by zero might occur when the spectrum is inverted. It can be thought of as white noise added to the spectrum with a certain power level. That power level is:

$$\text{stab power (db below mean power)} = 10 \cdot \log_{10}(\text{stab})$$

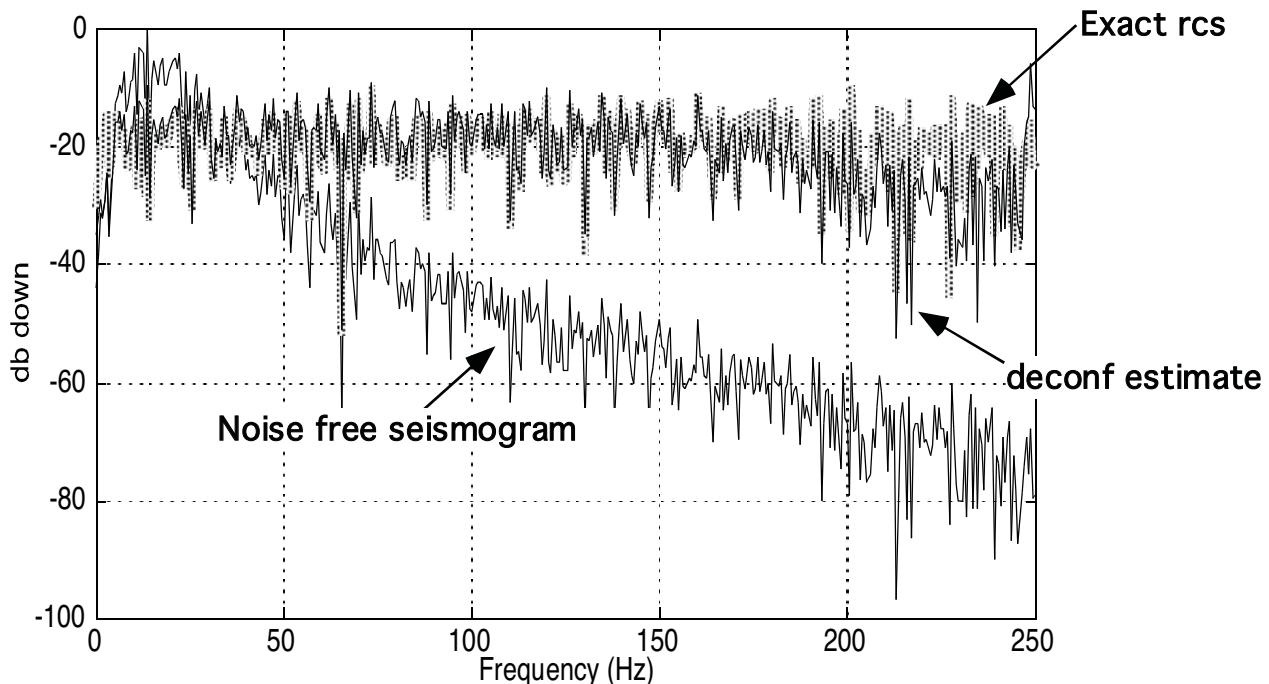
So the default stab of .0001 is a db level of $10 \cdot (-4)$ or 40 db down from mean power.

If we choose a frequency smoother of 10Hz $\rightarrow 10 \cdot 1.6 = 16$ points, and default the stab factor, then, the deconvolution of the noise free seismogram gives:



Frequency Domain Spiking Deconvolution

In the frequency domain our result looks like:

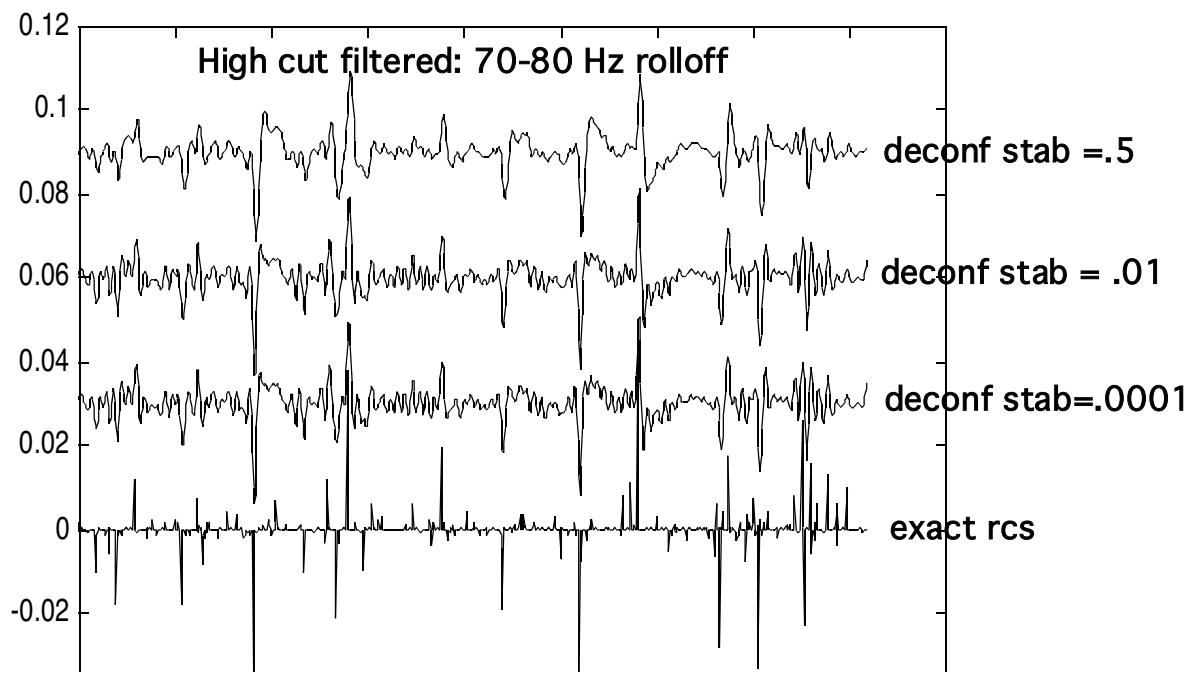
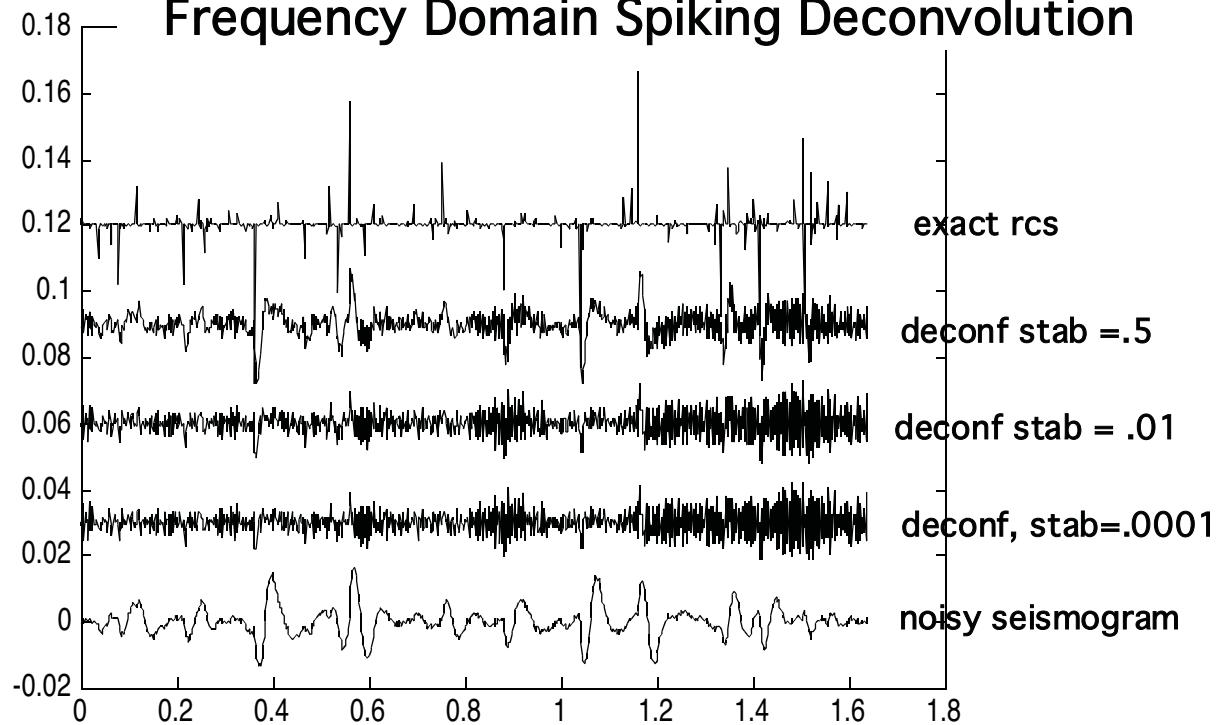


We can see that the estimate is quite good. We can be more precise about how good it is by using the Matlab function `mxcorr` which compares two time series and returns the maximum of their cross correlation and the lag at which it occurs. The results in:

`max correlation = .39 at lag of .1 samples`

If we now run the same process with the same parameters on the noisy seismogram we obtain quite a different result as shown on the next page.

Frequency Domain Spiking Deconvolution



Here are some results from maxcorr:

	Filtered results	Max corr	Lag
deconf stab=.0001		0.0603	1.3000
deconf stab = .01		0.0728	1.6000
deconf stab = .5		0.1510	3.0000

Frequency Domain Spiking Deconvolution

We can develop a simple mathematical model for deconvolution in the frequency domain. First, the convolutional model for a multiple-free seismic trace is

$$s(t) = r(t) \bullet w(t) + n(t) \quad (1)$$

where r is reflectivity, w is the wavelet, and n is additive noise. In the frequency domain, this becomes

$$S(f) = R(f)W(f) + N(f) \quad (2)$$

Generally, there will be a range of frequencies, called the signal band, over which the $R(f)W(f)$ term dominates over $N(f)$. Denoting the bounds of this frequency band by f_{\min} and f_{\max} , we can have the approximation

$$|S(f)| \approx |R(f)| |W(f)|, \quad f_{\min} \leq f \leq f_{\max} \quad (3)$$

where the vertical bars (e.g. $|S(f)|$) denote absolute values or amplitude spectra. Note that by using amplitude spectra, we are discarding the possibility of estimating the wavelet phase directly from the data.

The next step, spectral smoothing is difficult to fully justify mathematically. Denoting a smoothed spectrum by an overbar, the "white reflectivity" assumption means that

$$\overline{|R(f)|} \approx 1 \quad (4)$$

We then argue that smoothing $|S(f)|$ yields an estimate of the amplitude spectrum of the embedded wavelet. Though we know this is not precisely true, it is approximately so in many useful situations.

Frequency Domain Spiking Deconvolution

Thus we have the estimate

$$|W(f)|_{\text{est}} = |S(f)| \approx |W(f)| \quad (5)$$

The amplitude spectrum of the deconvolution operator is just the inverse of this

$$|D(f)| = |W(f)|_{\text{est}}^{-1} \quad (6)$$

Generally, this spectral division is problematic if there are frequencies where the estimated wavelet's spectrum is very small. Where it is small usually means that there was not much radiated source power and so noise is likely dominant. Since these small values are inverted, they become very important in $D(f)$. Given these considerations, it is customary to add a small constant to the estimated wavelet's amplitude spectrum prior to inversion. Then

$$|D(f)| = \frac{1}{|W(f)|_{\text{est}} + \mu A_{\max}} \quad (7)$$

where: $A_{\max} = \text{maximum}\left(|W(f)|_{\text{est}}\right)$ (8)

The constant μ is called the "white noise factor" or "stability factor" and is a small positive number usually between .01 and .000001.

Lastly, we must estimate the phase spectrum of $D(f)$. Under the minimum phase assumption and using H to denote the Hilbert transform, we have

$$\phi_D(f) = H\left(\ln\left(|D(f)|\right)\right) \quad (9)$$

Frequency Domain Spiking Deconvolution

where we use $|D(f)|$ as given by equation (7). Note that the stability factor also guards against taking the logarithm of zero in equation (9). So, we now have the amplitude and phase spectrum of the deconvolution operator and we are ready to apply it to the seismic trace. Again in the frequency domain, this is

$$S_D(f) = S(f)D(f) = S(f)|D(f)|e^{i\phi_D(f)} \quad (10)$$

If we substitute equation (3) into equation (10) we can obtain an expression for the embedded wavelet remaining after deconvolution

$$S_D(f) = [R(f)W(f) + N(f)]|D(f)|e^{i\phi_D(f)} \quad (11)$$

Neglecting the noise term, we estimate the embedded wavelet as

$$W_D(f) = W(f)|D(f)|e^{i\phi_D(f)}, f_{\min} \leq f \leq f_{\max} \quad (12)$$

Assuming a bandpass filter is applied following deconvolution, we can regard $W_D(f)$ as effectively zero outside this bandwidth. Equation (12) can be further written as

$$W_D(f) = W(f) \frac{1}{|W(f)|_{\text{est}} + \mu A_{\max}} e^{i\phi_D(f)} \approx 1, f_{\min} \leq f \leq f_{\max} \quad (13)$$

In the last step, the approximate unity follows only if the assumptions of stationary wavelet, white reflectivity, and minimum phase are approximately valid. If the first two fail then we expect a non-white amplitude spectrum for $W_D(f)$ and if the last fails then we expect a residual phase spectrum.

Finding a Wavelet's Inverse

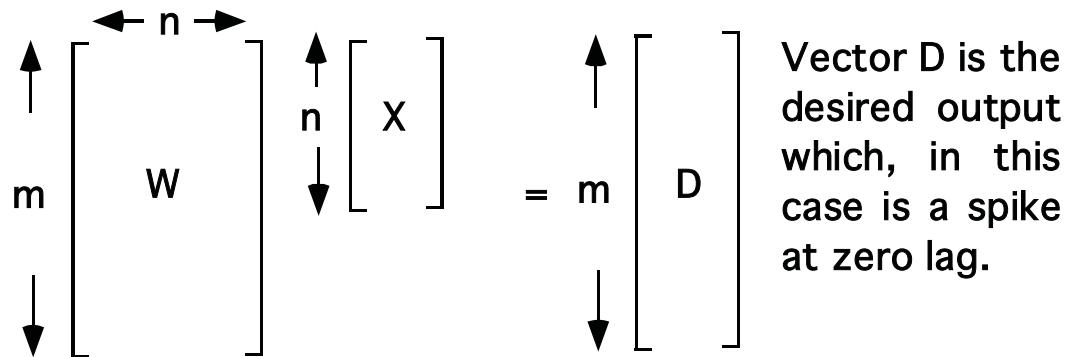
If w symbolizes a wavelet and x is its unknown inverse, then the two are related by:

$$w \bullet x = 1$$

Here, the \bullet denotes convolution and 1 is a unit vector. In matrix notation, this is written:

$$\begin{bmatrix} w_0 & 0 & 0 & 0 & 0 \\ w_1 & w_0 & 0 & 0 & 0 \\ w_2 & w_1 & w_0 & 0 & 0 \\ w_3 & w_2 & w_1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & w_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

Here we have assumed that both w and x are causal. In general such an inverse will require infinitely many terms to produce an exact result so we will look for an approximate finite length inverse. If n is the length of the inverse and m is the length of the wavelet, then the above matrix equation is:



Finding a Wavelet's Inverse

Thus we have chosen $n < m$ providing more equations than unknowns and are in a position to seek a least squares solution. The classic least squares approach is (see Hatton et al., 1986 p 31):

$$W^T W X = W^T D \quad \text{The normal equations}$$

$$X = (W^T W)^{-1} W^T D \quad \text{The estimated } X$$

$$\begin{bmatrix} w_0 & w_1 & w_2 & w_3 & \cdots \\ 0 & w_0 & w_1 & w_2 & \cdots \\ 0 & 0 & w_0 & w_1 & \cdots \\ 0 & 0 & 0 & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & w_0 \end{bmatrix} \begin{bmatrix} w_0 & 0 & 0 & 0 & 0 \\ w_1 & w_0 & 0 & 0 & 0 \\ w_2 & w_1 & w_0 & 0 & 0 \\ w_3 & w_2 & w_1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} w_0 & w_1 & w_2 & w_3 & \cdots \\ 0 & w_0 & w_1 & w_2 & \cdots \\ 0 & 0 & w_0 & w_1 & \cdots \\ 0 & 0 & 0 & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & w_0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ \vdots \end{bmatrix}$$

Multiplication by W^T does a crosscorrelation because it can be easily seen to be convolution with the time reversed wavelet. This can be seen to be:

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \phi_3 & \cdots \\ \phi_1 & \phi_0 & \phi_1 & \phi_2 & \cdots \\ \phi_2 & \phi_1 & \phi_0 & \phi_1 & \cdots \\ \phi_3 & \phi_2 & \phi_1 & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \phi_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} w_0 \\ 0 \\ 0 \\ \vdots \\ \vdots \end{bmatrix}$$

Where ϕ_j is the jth lag of the autocorrelation of w .

Finding a Wavelet's Inverse

We have seen that the process of finding the m-length causal inverse, x , of a causal wavelet, w , reduces to solving the m by m linear system:

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \phi_3 & \cdots \\ \phi_1 & \phi_0 & \phi_1 & \phi_2 & \cdots \\ \phi_2 & \phi_1 & \phi_0 & \phi_1 & \cdots \\ \phi_3 & \phi_2 & \phi_1 & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \phi_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} w_0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

Where ϕ_j is the j th lag of the autocorrelation of w .

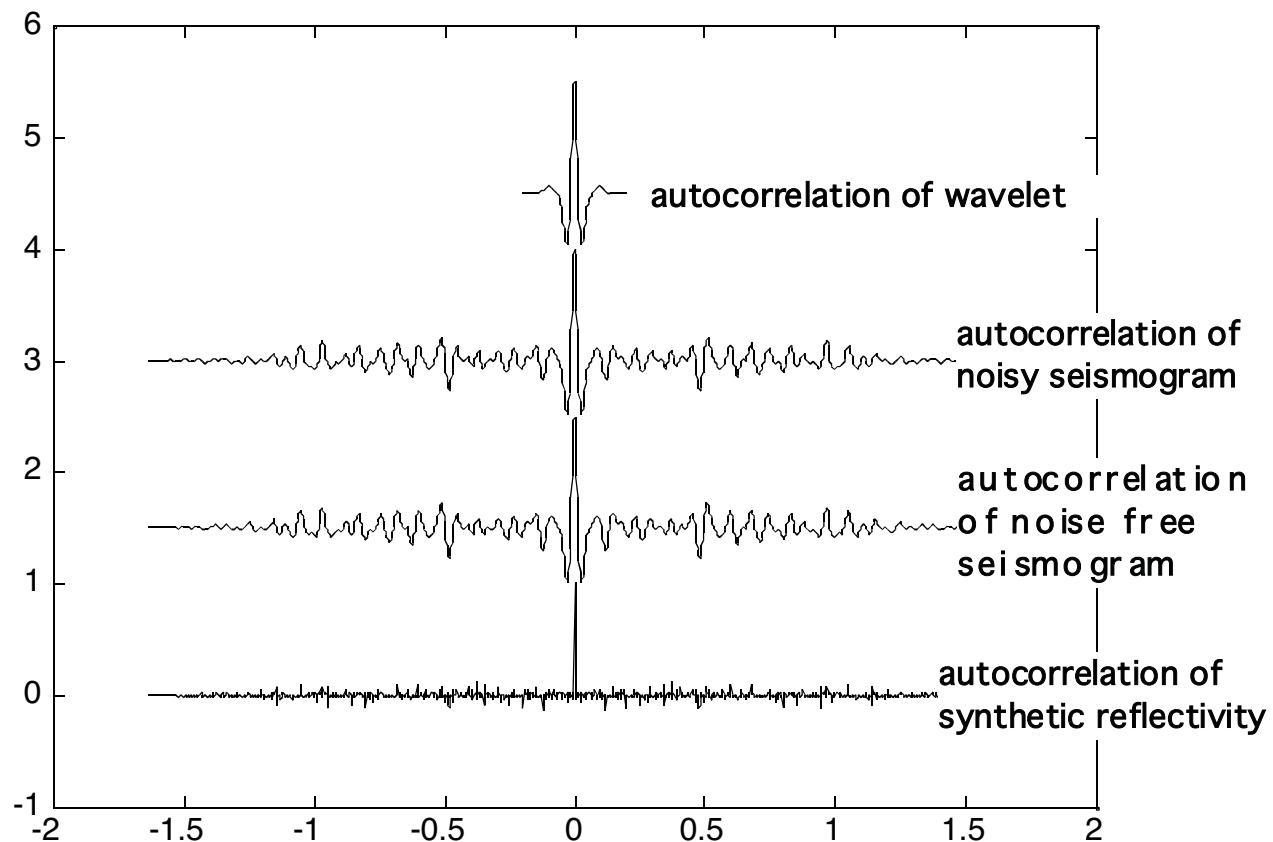
This remarkable result says that we don't need to know the wavelet itself, just m lags of its autocorrelation. And, if we are content to be off by an arbitrary scale factor, then we can replace w_0 by 1. How this is possible is a consequence of the following facts:

- A causal, stable wavelet with a causal, stable inverse IS minimum phase. (Karl, J.H., An Introduction to Digital Signal Processing, Academic Press, 1989, see pages 35-37)
- The Fourier transform of the autocorrelation is the power spectrum of the wavelet (Wiener-Khintchine Theorem). Thus the phase information is not present in the autocorrelation.

Thus, the problem of estimating the inverse to a minimum phase wavelet is reduced to one of estimating the autocorrelation of the unknown wavelet. Most techniques do so imperfectly.

Wiener Spiking Deconvolution

The original deconvolution technique, and still the workhorse of the methodology is a time domain method referred to as Wiener deconvolution. It rests on the time domain computation of the inverse of a minimum phase filter given its autocorrelation. Below we see the autocorrelations of the synthetic trace which we have been examining:



Thus we are reminded of the fact that the autocorrelation of the seismogram is very similar to the autocorrelation of the wavelet. This is a consequence of our assumption that the reflectivity is a random, white sequence and can be demonstrated mathematically as follows:

Wiener Spiking Deconvolution

Recall the expression for the convolutional model:

$$s(t) = w_m(t) \bullet r(t) + n(t)$$

which expresses the seismic trace, s , as a convolution between a wavelet with a possible multiple train, w_m , and a reflectivity, r , plus additive random noise, n . Since an autocorrelation is formed by time reversing the trace and convolving it with itself, we have:

$$\begin{aligned} A_s(t) &= s(t) \bullet s(-t) = [w_m(t) \bullet r(t) + n(t)] \bullet [w_m(-t) \bullet r(-t) + n(-t)] \\ &= w_m(t) \bullet r(t) \bullet w_m(-t) \bullet r(-t) + w_m(t) \bullet r(t) \bullet n(-t) + \\ &\quad n(t) \bullet w_m(-t) \bullet r(-t) + n(t) \bullet n(-t) \end{aligned}$$

Since the order of convolution is unimportant, the first term in this expression can be seen to be the convolution of the autocorrelations of w_m and r . The second and third terms both involve the cross correlations between two random sequences, r and n , and hence are zero while the last term is the autocorrelation of n . Thus

$$A_s(t) = A_w(t) \bullet A_r(t) + A_n(t)$$

Since r and n are both random sequences by assumption, their autocorrelations are delta functions and we obtain:

$$A_s(t) = A_w(t) + p_n \delta(t)$$

where p_n is the mean noise power. So we see that the autocorrelation of seismogram and wavelet should be equal except for the possibility of a slight increase in the zero lag power.

Wiener Spiking Deconvolution

The proof that the autocorrelation of the wavlet can be obtained from that of the seismogram relies on statistical properties that can never be exactly satisfied in practice. Therefore two problems arise: we must choose how many lags of the autocorrelation to allow into the solution, and we must guard against the possibility that the spectrum of the truncated autocorrelation might contain zeros. The first is "solved" by making the number of lags a user parameter, while the second requires the addition of a "stab" factor to the zero lag of the autocorrelation. Thus the normal equations which must be solved for the wavelet inverse are modified to:

$$\begin{bmatrix} \phi_0 + \lambda & \phi_1 & \phi_2 & \phi_3 & \cdots \\ \phi_1 & \phi_0 + \lambda & \phi_1 & \phi_2 & \cdots \\ \phi_2 & \phi_1 & \phi_0 + \lambda & \phi_1 & \cdots \\ \phi_3 & \phi_2 & \phi_1 & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \phi_0 + \lambda \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

Where ϕ is the autocorrelation of the seismic trace, λ is the stab factor, and x is the unknown inverse operator. In comparing this algorithm with frequency domain decon, it is noted that they are nearly the Fourier equivalents of one another. Windowing the autocorrelation in Wiener decon is equivalent to smoothing the power spectrum in frequency decon. The number of lags in the autocorrelation and the number of points in the frequency domain smoother are inversely related. Reasoning very loosely, we have:

$$n_{\text{lags}} \Delta t \approx \frac{1}{n_{\text{smooth}} \Delta f} \Rightarrow n_{\text{lags}} \approx \frac{1}{n_{\text{smooth}}} \frac{T}{\Delta t} = \frac{n_{\text{samps}}}{n_{\text{smooth}}}$$

Wiener Spiking Deconvolution

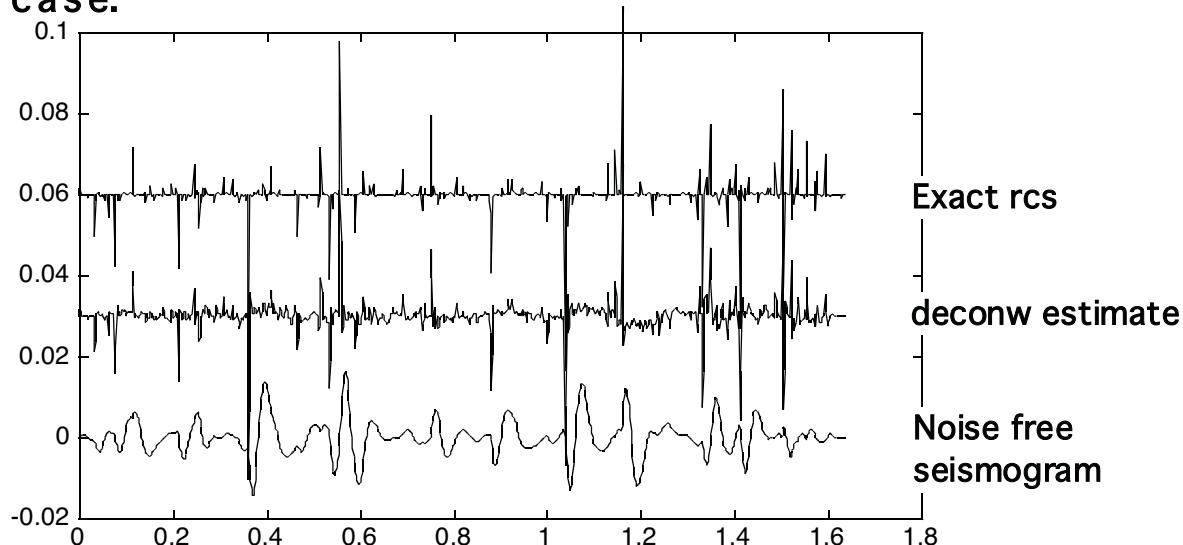
Since our synthetic has a 2 mil sample rate, it has roughly 800 samples, so we expect the 16 point smoother we used to be similar to $800/16 = 50$ lags of the autocorrelation. Here are is the help file from the Matlab routine deconw:

```
%  
% [trout,x]=deconw(trin,trdsign,n,stab)  
% [trout,x]=deconw(trin,trdsign,n)  
%  
% routine performs a Weiner style deconvolution of the  
% input trace  
%  
% trin= input trace to be deconvolved  
% trdsign= input trace to be used for operator design  
% n= number of autocorrelogram lags to use (and length of  
% inverse operator  
% stab= stabilization factor expressed as a fraction of the  
% zero lag of the autocorrelation.  
% ***** default=.0001 *****  
%  
% trout= output trace which is the deconvolution of trin  
% x= output inverse operator used to deconvolve trin
```

Algorithm:

- Compute the autocorrelation of the input seismic trace.
- Window the autocorrelation (boxcar) to only n lags
- Set up the normal equations for the wiener inverse, add the stab factor to the diagonal, and solve
- Convolve the inverse operator with the seismic trace.

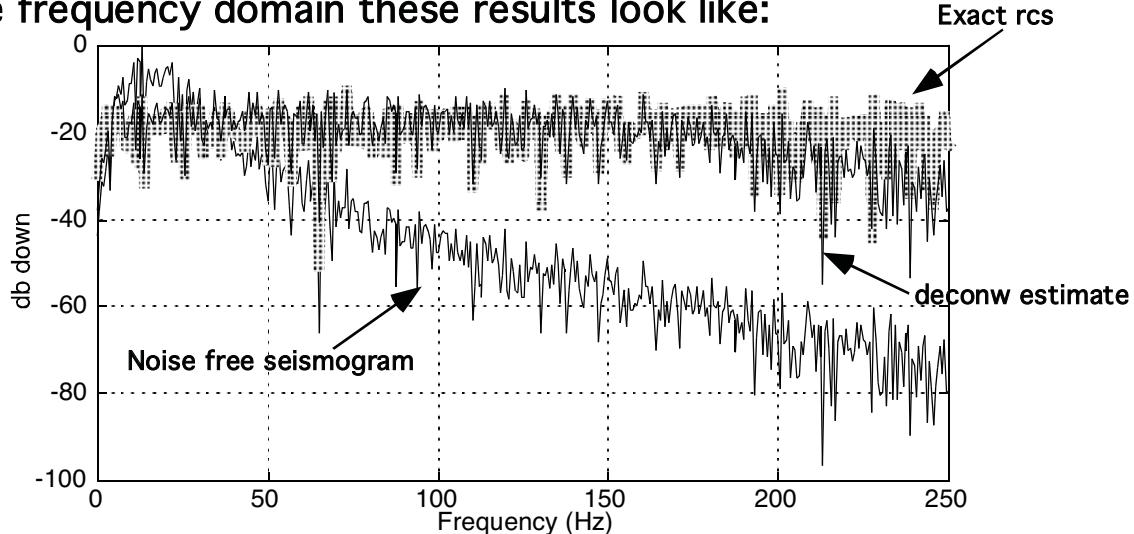
Using essentially comparable parameters to the frequency domain example, we obtain for the noise free case:



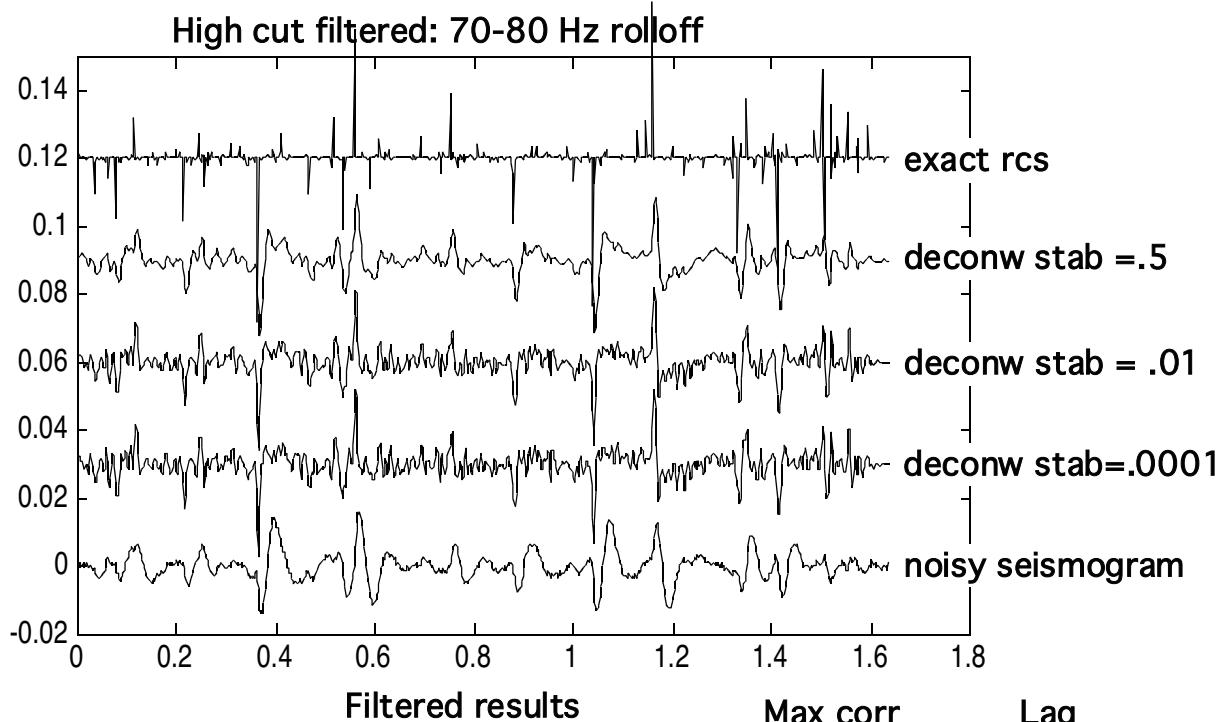
From maxcorr, we obtain a maximum cross correlation coefficient between the estimated rcs and the exact ones of .39 at a lag of .2 seconds. Very close to the result from deconf.

Wiener Spiking Deconvolution

In the frequency domain these results look like:



Here are some sample decons of the noisy trace which have already been filtered back to 70Hz.



Here are some results from maxcorr:

	Filtered results	Max corr	Lag
deconvw stab=.0001	0.1262	1.4000	
deconvw stab = .01	0.1414	1.6000	
deconvw stab = .5	0.1802	3.0000	

Prediction and Prediction Error Filters

A forward prediction filter is a linear convolutional operator which is designed to predict the next element in a sequence given the values preceding it. We can write this process using the matrix form of convolution as:

$$\begin{bmatrix} w_0 & 0 & 0 & \cdots & 0 \\ w_1 & w_0 & 0 & \cdots & 0 \\ w_2 & w_1 & w_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ w_m & \vdots & \vdots & \vdots & w_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{m+1} \end{bmatrix} \quad \text{Eqn 1}$$

Multiplying both sides of this by the transpose of the Toeplitz W matrix and forming the normal equations as we did for inverse filtering gives:

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_N \\ \phi_1 & \phi_0 & \phi_1 & \cdots & \cdots \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ \phi_N & \vdots & \vdots & \vdots & \phi_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{N+1} \end{bmatrix} \quad \text{Eqn 2}$$

Here, in contrast to the normal equations for inverse filtering, we have the signal autocorrelation appearing on both sides of the equation. The solution to these equations gives a prediction filter x , which, in practice, is used to predict values "off the ends" of the sequence on which it was designed. We might suspect that since there is no phase information going into the prediction filter design that the filter will be minimum phase and that is indeed the case.

Prediction and Prediction Error Filters

We now wish to draw a parallel between prediction filtering and the design of inverse filters. It turns out that the relationship is not with prediction filters but with a closely related filter, the prediction error filter. To derive this, note that equation 1 can be written as the following expression with z transforms:

$$w(z)x(z) = z^{-1}(w(z) - w_0)$$

Now, we can reformulate this into:

$$w(z)x(z) - z^{-1}w(z) = -z^{-1}w_0$$

Note that the left hand side is essentially the difference between the predicted values, $w(z)x(z)$, and their actual values, $z^{-1}w(z)$. Hence it is termed the prediction error. Manipulating further:

$$w(z)(z^{-1} - x(z)) = z^{-1}w_0$$

multiply by z $w(z)(1 - zx(z)) = w_0 \quad \text{eqn 3}$

The operator, $1 - zx(z)$, is called a prediction error filter of unit lag because it asserts that we can operate on $w(z)$ to yield w_0 followed by a sequence of zeros. That is, we can't possibly predict the first value in a sequence, so the error in that prediction must always be 100%, however, we assert through equation 3, that all other values can be predicted without error. Of course this won't be possible in general and we will obtain a least squares solution which minimizes the prediction error.

Prediction and Prediction Error Filters

Formally, equation 3 is identical, to within a scale factor, of the z transform expression for the design of an inverse filter for w. That is, w^{-1} must satisfy:

$$w(z)w^{-1}(z) = 1$$

If we write the matrix expression for equation 3, we have:

$$\begin{bmatrix} w_0 & 0 & 0 & \cdots & 0 \\ w_1 & w_0 & 0 & \cdots & 0 \\ w_2 & w_1 & w_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ w_m & \vdots & \vdots & \vdots & w_0 \end{bmatrix} \begin{bmatrix} 1 \\ -x_0 \\ -x_1 \\ \vdots \\ -x_N \end{bmatrix} = \begin{bmatrix} w_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Forming the normal equations as before leads to:

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_N \\ \phi_1 & \phi_0 & \phi_1 & \cdots & \cdots \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ \phi_N & \vdots & \vdots & \vdots & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ -x_0 \\ -x_1 \\ \vdots \\ -x_N \end{bmatrix} = \begin{bmatrix} w_0^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{eqn 4}$$

As expected, equation 4 is nearly identical to the normal equations for a Wiener inverse filter. Thus we make two conclusions:

- Prediction filters and prediction error filters are minimum phase.
- Spiking (Wiener) deconvolution is identical to unit lag prediction error filtering.
- Thus deconvolution removes the predictable part of the trace.

Prediction and Prediction Error Filters

Having designed a prediction filter to predict one sample ahead, it is a simple matter to design one to predict α samples ahead by modifying equation 1 to give:

$$\begin{bmatrix} w_0 & 0 & 0 & \cdots & 0 \\ w_1 & w_0 & 0 & \cdots & 0 \\ w_2 & w_1 & w_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ w_m & \vdots & \vdots & \vdots & w_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} w_\alpha \\ w_{\alpha+1} \\ w_{\alpha+2} \\ \vdots \\ w_{\alpha+m} \end{bmatrix} \quad \text{eqn 5}$$

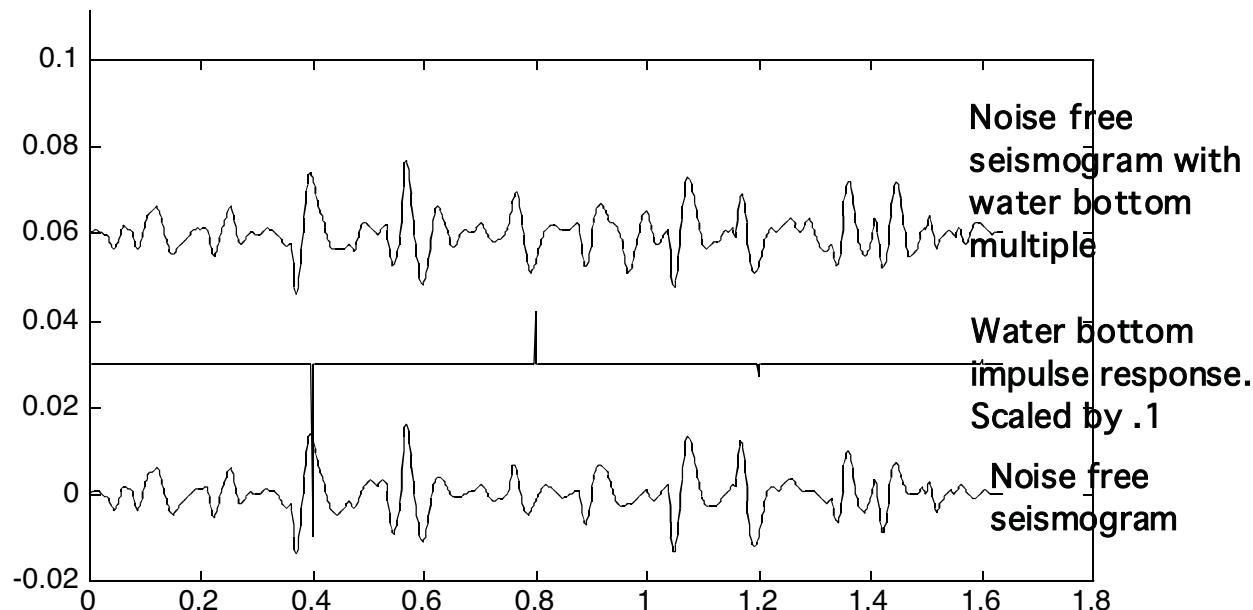
Proceeding as before, we form the normal equations equivalent to equation 2:

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_N \\ \phi_1 & \phi_0 & \phi_1 & \cdots & \cdots \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ \phi_N & \vdots & \vdots & \vdots & \phi_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \phi_\alpha \\ \phi_{\alpha+1} \\ \phi_{\alpha+2} \\ \vdots \\ \phi_{\alpha+N} \end{bmatrix} \quad \text{eqn 6}$$

The solution to equation 6 gives a $N+1$ long prediction operator which attempts to predict α samples ahead. It is called a gapped prediction operator and plays an essential role in the suppression of multiples which fit the convolutional model.

Gapped Predictive Deconvolution

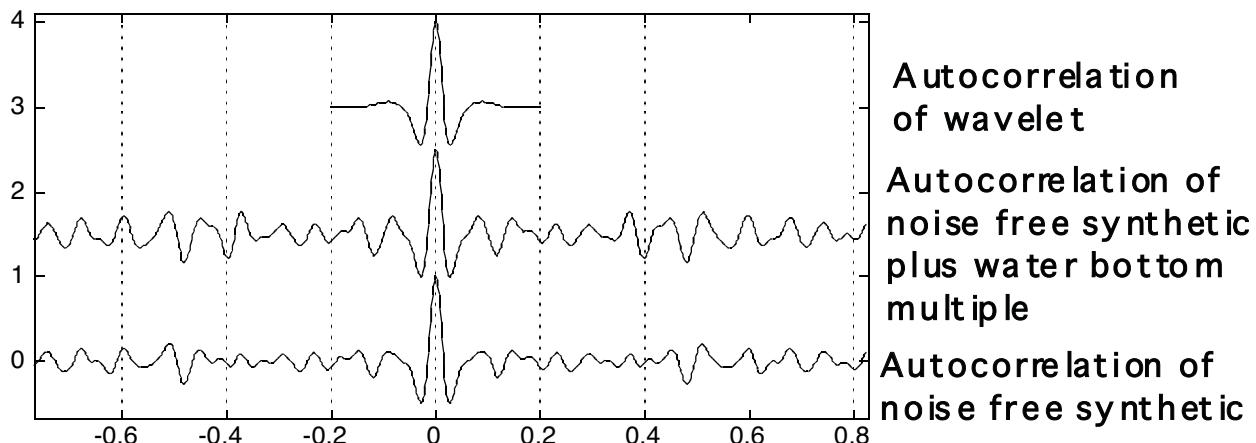
We have seen that Wiener spiking deconvolution is equivalent to unit lag prediction error filtering. A similar technique is to use prediction filters of some lag other than zero to compute the predictable part of a seismic trace and subtract it from the original trace. Thus, if the lag used is 1, we should get the same result as Wiener spiking deconvolution. This technique is most useful in attenuating multiples that fit the convolutional model. Such a multiple is the water bottom multiple which can be simulated (Backus, M.M., Geophysics, vol 24, p233-261, 1959) by convolving our seismogram with the impulse response of an ocean .2 seconds deep (2-way time) and an ocean bottom rc of .4 (huge).



It is difficult to see the effects on the seismogram but if you look closely at .4 seconds behind a major reflector, then you should see a reverse polarity image of it superimposed on the seismogram.

Gapped Predictive Deconvolution

On the autocorrelations, we see a significant new side lobe has developed at a lag of .4 seconds.



Based on these displays we are lead to select a prediction gap of 180 samples (.36 seconds) and an operator length of 50 samples (as in spiking decon). Here is the help file from the Matlab function deconpr:

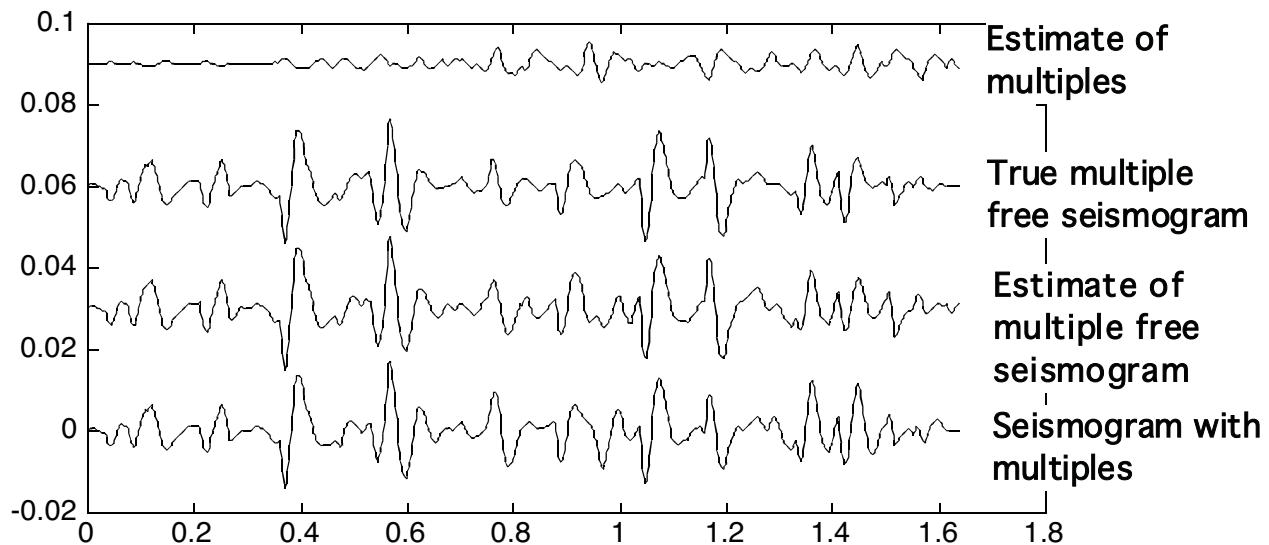
```
% [trout,x]=deconpr(trin,trdsign,nop,nlag,stab)
% [trout,x]=deconpr(trin,trdsign,nop,nlag)
%
% DECONPR performs Wiener predictive deconvolution by calling
% PREDICT to design a prediction filter, nop long with lag nlag
% and stab factor, using trdsign. The predicted part of trin,
% trinhat,
% then formed by convolving the prediction operator with trin,
% and trout is computed by delaying trinhat by nlag samples and
% subtracting it from trin. The prediction operator is returned
% in x.
%
% trin= input trace to be deconvolved
% trdsign= input trace used to design the prediction operator
% nop= number of points in the prediction operator
% nlag= prediction lag distance in samples
% stab= stabilazation factor expressed as a fraction of the zero
%      lag of the autocorrelation.
% ***** default=.0001 *****
%
% trout= deconvolved output trace
% x= prediction operator
%
% See also: Peacock and Treitel, Geophysics vol 34, 1968
% and the description of PREDICT
```

Algorithm:

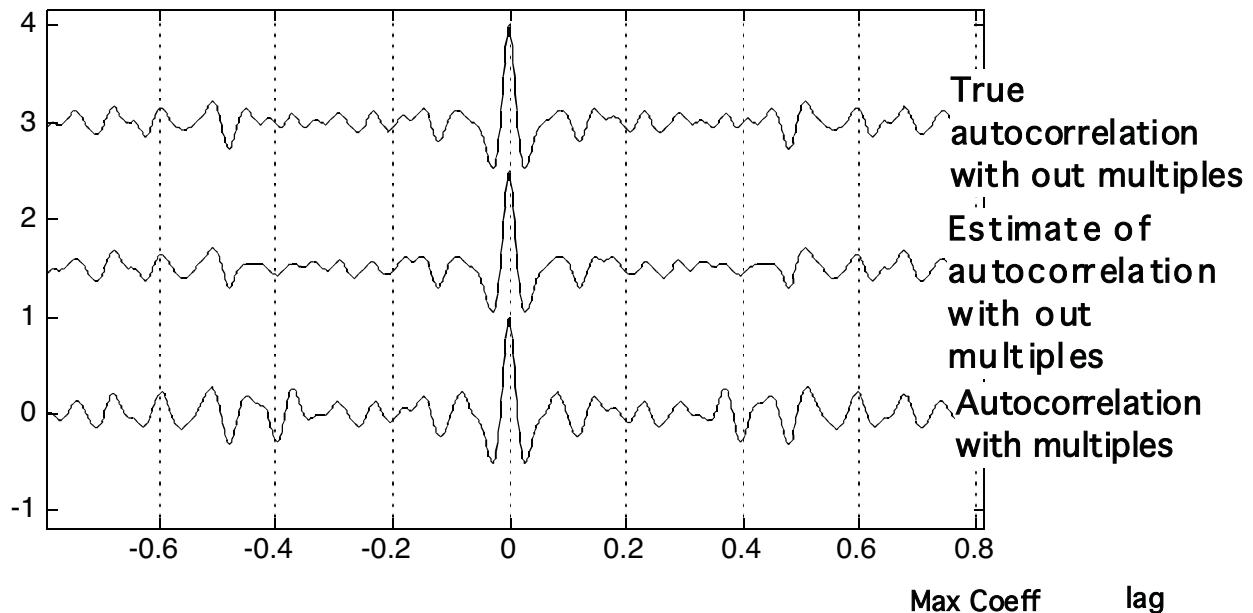
- Design a gapped, minimum phase prediction filter (with stab factor) from the autocorrelation of trdsign.
- Convolve the prediction operator with trin to form the predictable part.
- Subtract the predictable part of trin from trin to form the output trace.

Gapped Predictive Deconvolution

So, running deconpr on the noise free synthetic with multiples using the gap and operator length mentioned and a default stab factor gives:



If we examine the autocorrelations, we see that the periodicity in the autocorrelations at lag of .4 seconds has been suppressed.

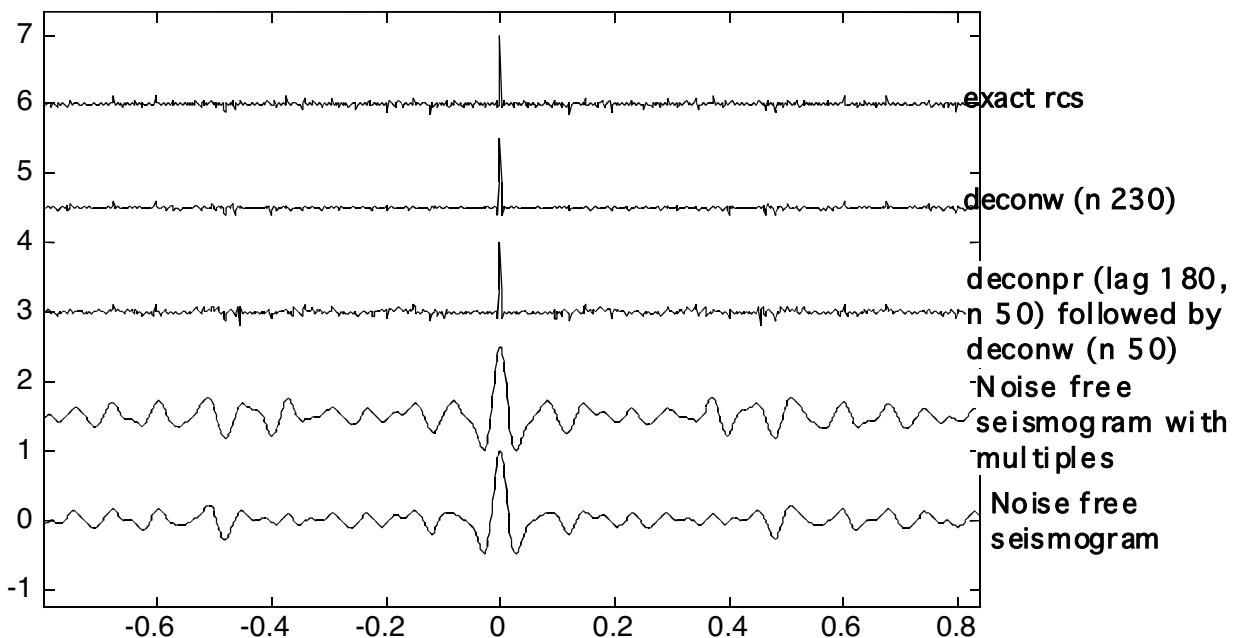
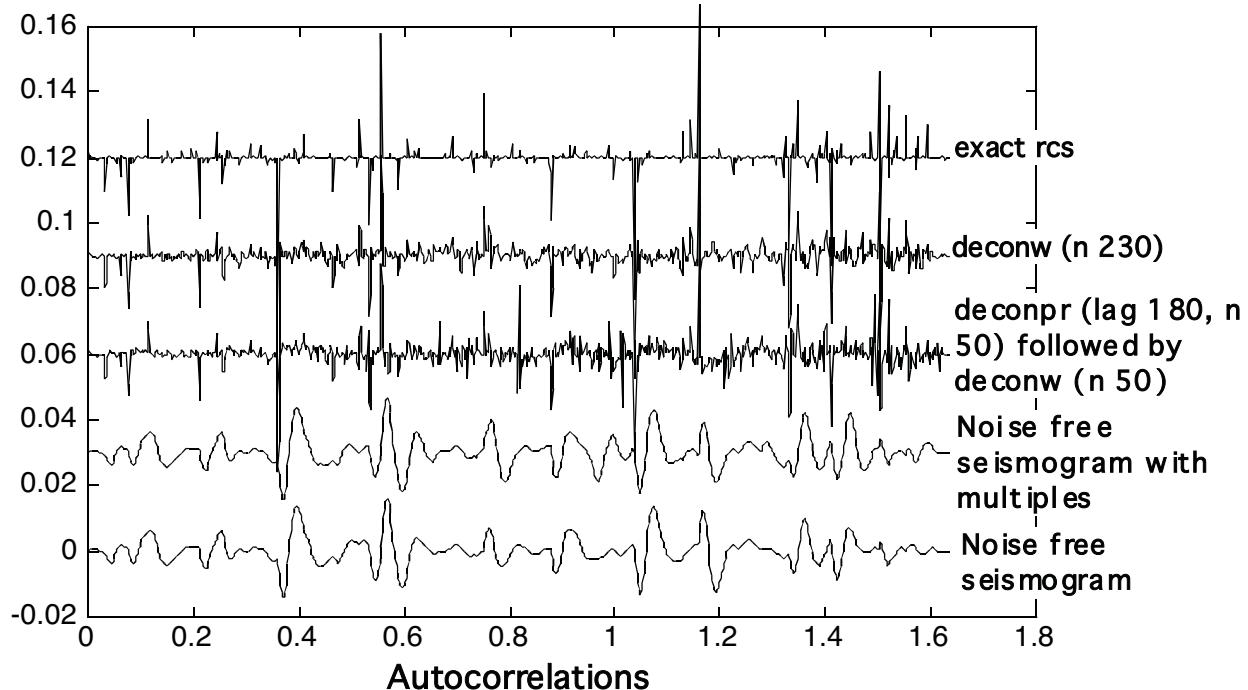


Results from using maxcorr to compare with the noise free, multiple free seismogram

	Max Coeff	lag
With multiples	0.9358	-0.1000
After deconpr	0.9736	-0.1000

Gapped Predictive Deconvolution

Here we compare the results from following our previous deconpr by a deconw ($n = 50$) with a single deconw with $n = 180 + 50 = 230$.



Results from using maxcorr to compare the two decons on this page with the exact rcs

	Max Coeff	lag
deconpr + deconw	0.3451	-0.3000
long deconw	0.3892	-0.2000

Burg (Maximum Entropy) Deconvolution

Burg deconvolution is closely related to Wiener spiking deconvolution in that it is accomplished using prediction error filters of unit lag. The technique was designed by J.P. Burg (Claerbout, J.F., 1976, Fundamentals of Geophysical Data Processing, McGraw-Hill) in response to doubts about the Wiener technique of windowing the autocorrelation. He reasoned that windowing the autocorrelation caused the normal equations to design a prediction filter as though the data had the property that its autocorrelation vanished after n lags. It certainly seems reasonable to expect a better spectral estimation from an algorithm that expects the autocorrelation to continue in some reasonable way. Burg found a technique which designed a prediction error filter directly from the data rather than first forming the autocorrelation and windowing it. His technique will not be developed here but we will quote the following properties:

- The Burg prediction error filter minimizes the sum squared error from forward and backward prediction.
- The so-called Burg spectrum is computed as the inverse of the spectrum of the prediction error filter. (This is not explicitly done in Burg deconvolution.)
- The Burg prediction error filter is minimum phase.
- Though not computed directly, the Burg theory can be shown to be equivalent to a Wiener theory which, instead of truncating the autocorrelation, extrapolates it in a way which maximizes the randomness (entropy) of the implied signal. (Kanasewich, E.R., 1981, Time Sequence Analysis in Geophysics (3rd Edition), University of Alberta Press)

Burg (Maximum Entropy) Deconvolution

Here is the help file from the Matlab function deconvb:

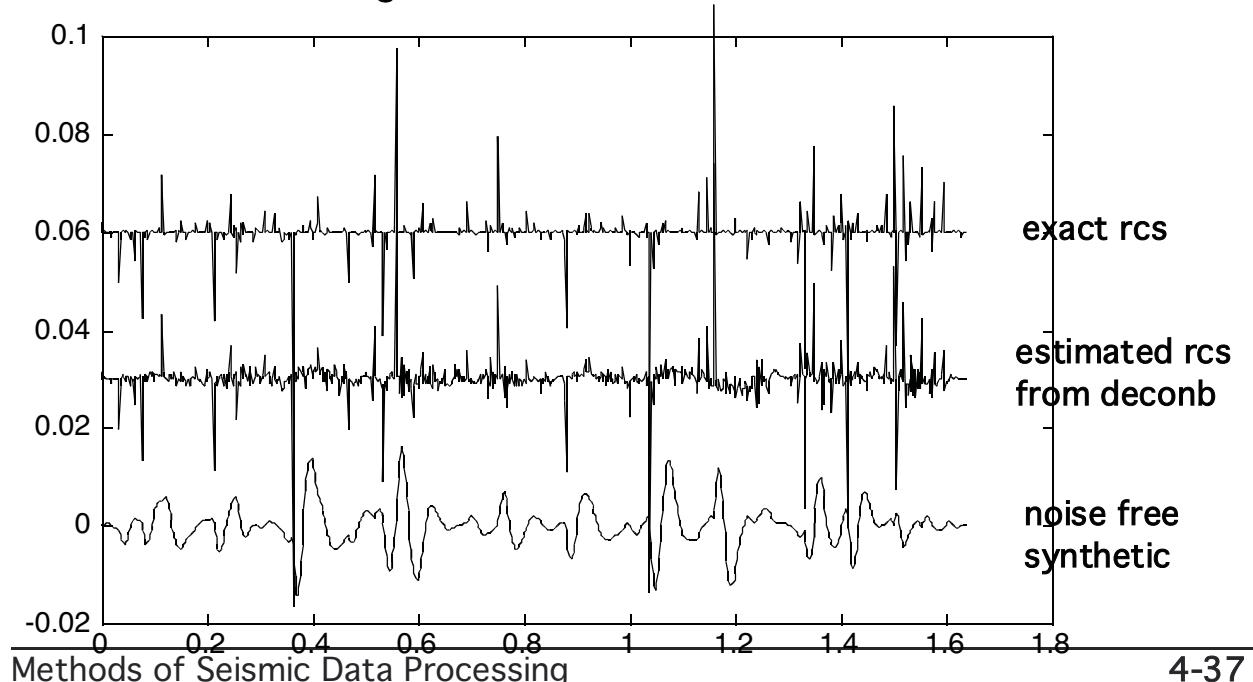
```
%  
% [trout,pefilt]=deconvb(trin,trdsign,l)  
%  
% routine performs a Burg scheme deconvolution of the  
% input trace  
%  
% trin= input trace to be deconvolved  
% trdsign= input trace to be used for operator design  
% l= prediction error filter length (and length of  
% inverse operator  
%  
% trout= output trace which is the deconvolution of trin  
% pefilt= output inverse operator used to deconvolve trin
```

Algorithm :

- Design a unit lag prediction error filter of length l on trdsign.
- Convolve the prediction error filter with trin to form the output trace.

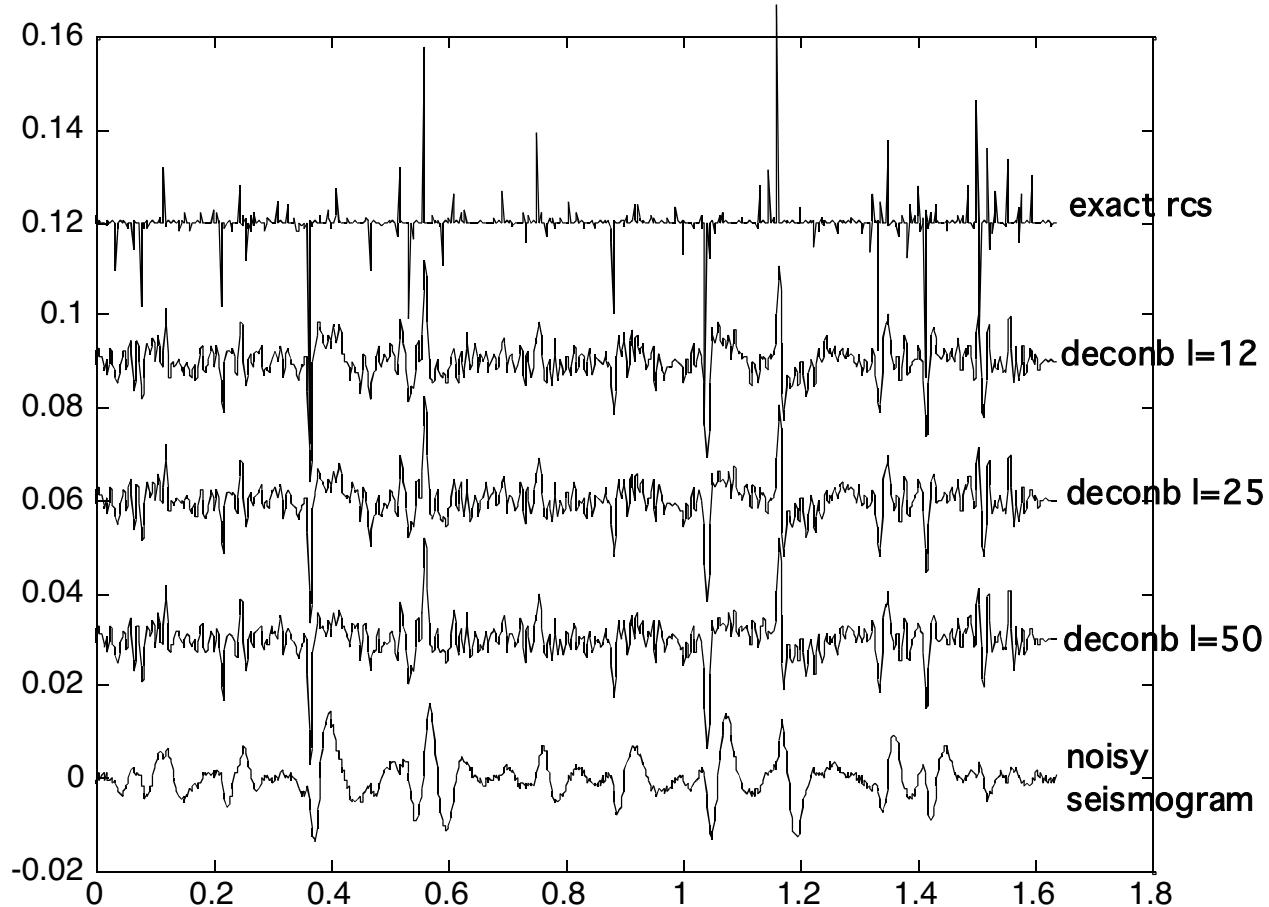
Note that there is no stab factor involved (the algorithm is always stable) and that we must choose the length of a prediction error filter instead of the number of lags on an autocorrelation function. However, as a rough guess, we might consider l to be similar to the number of lags.

Here is the result from deconvolving our noise free synthetic with l=50. It achieves a maximum cross correlation of .4355 at a lag of zero, considerably better than the other algorithms.



Burg (Maximum Entropy) Deconvolution

Below is the result from deconvolving the noisy seismogram with three different prediction filter lengths. All results have been high cut filtered at 70Hz.



Results from using maxcorr
to compare each
deconvolution with the
exact rcs:

	max corr	lag
deconv l=50	0.2227	-1.4000
deconv l=25	0.2168	-1.4000
deconv l=12	0.2093	-1.4000

So we see that, at least on this synthetic, the Burg algorithm does an excellent job, is very stable, and not very sensitive to the choice of the parameter l .

The Minimum Phase Equivalent Wavelet

Any wavelet, no matter what its amplitude or phase spectrum, can be said to have a related wavelet called its *minimum phase equivalent*. If the given wavelet has an amplitude spectrum which is positive definite, then its minimum phase equivalent has the same amplitude spectrum but a phase spectrum computed as the Hilbert transform of the logarithm of the amplitude spectrum. If we let $w(t)$ denote the arbitrary wavelet and F and H be the forward Fourier and Hilbert transforms respectively, then:

$$F[w(t)] = W(f) = A(f)\exp(i\phi(f)).$$

In this expression, $W(f)$ is the complex-valued Fourier spectrum and $A(f)$ and $\phi(f)$ are the real-valued amplitude and phase spectra. Now, if $W(f)$ is positive everywhere, then the minimum phase equivalent wavelet has a Fourier spectrum given by:

$$W_{\min}(f) = A(f)\exp(i\phi_{\min}(f)).$$

where

$$\phi_{\min}(f) = H[\ln(A(f))].$$

In the more general case, when $A(f)$ might have a zero somewhere or when large portions of its domain are dominated by noise, it is customary to compute the minimum phase equivalent by:

$$W_{\min}(f) = A_\mu(f)\exp(i\phi_{\mu\text{-}\min}(f)).$$

$$\phi_{\mu\text{-}\min}(f) = H[\ln(A_\mu(f))].$$

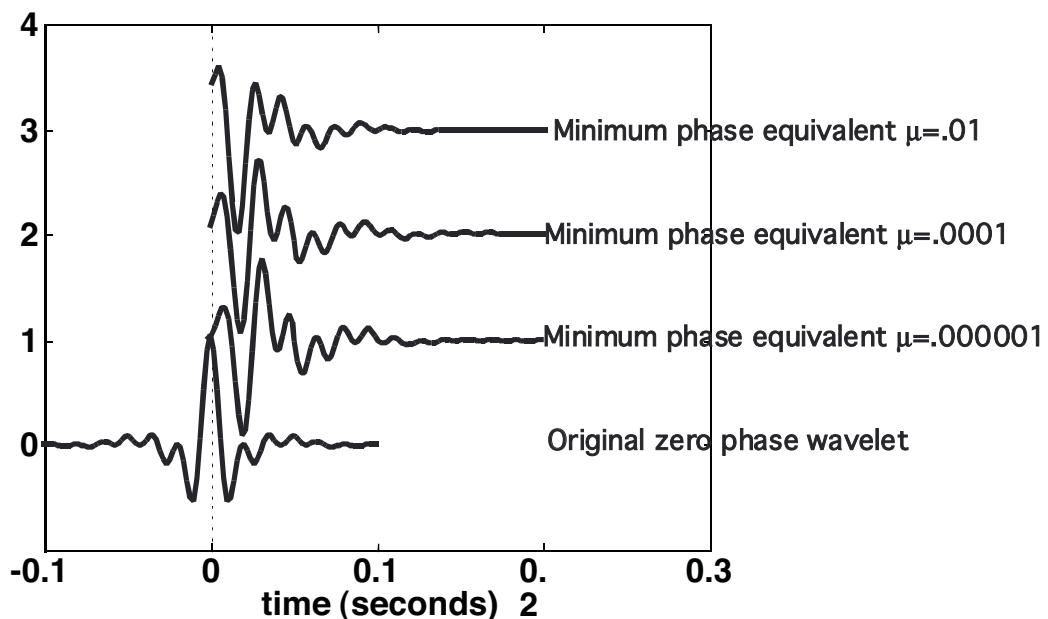
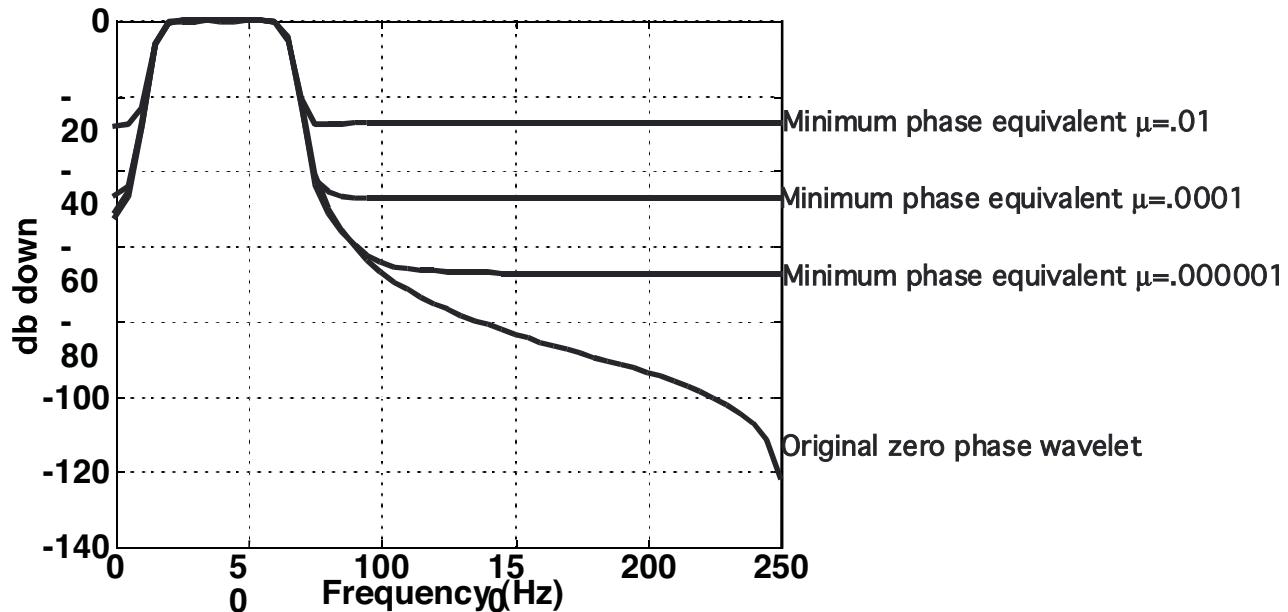
$$A_\mu(f) = A(f) + \mu \max(A(f))$$

In the final expression, μ is a small number, typically between 10^{-1} and 10^{-6} , whose exact value depends on the signal-to-noise ratio and the spectral shape of the signal spectrum.

The Minimum Phase Equivalent Wavelet

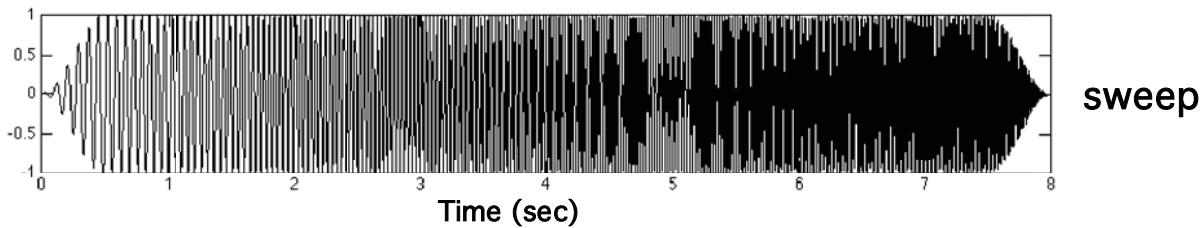
In its most general form, the minimum phase equivalent wavelet is not unique because of its dependence on the "white noise factor" μ .

The simulation below shows a [10,20,60,70] zero-phase Ormsby wavelet and three of its minimum phase equivalents. Note that all have distinctly different phase as a result of their differing μ values.

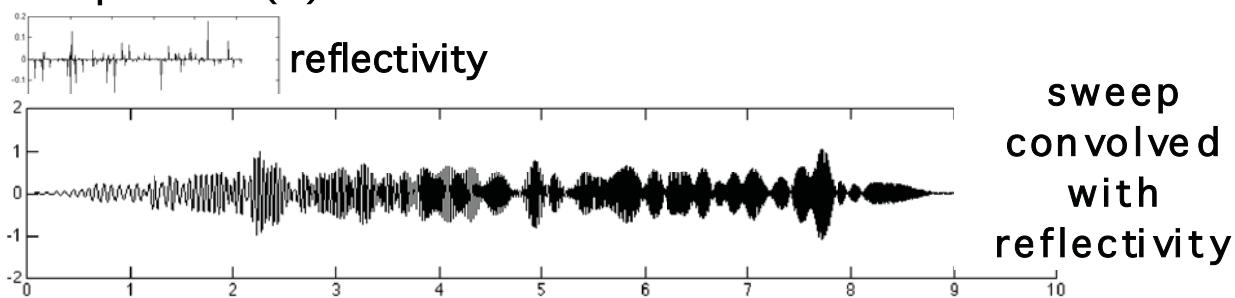


Vibroseis Deconvolution

Exploration with Vibroseis® sources is fundamentally different from the use of explosive sources and needs special consideration in our theoretical development. Instead of an unknown impulsive source waveform, vibroseis attempts to create a known extended source known as a sweep. A sweep is typically a signal which moves continuously through a specified frequency band generating only one frequency instantaneously. Typically sweeps are linear (the same time is swept at each frequency) but non-linear sweeps, which emphasize the high frequencies, are also common. Here is a 10-70 Hz, 8 second, linear sweep:



The convolutional model still fits this source equally well as the impulsive source. That is, given a reflectivity $r(t)$, we can simulate the earth's response by convolving the sweep with $r(t)$:



Obviously, this is a different sort of record than the impulsive source and is much more difficult to interpret because the source waveform is so extended. We need a method of collapsing the source to a compact pulse. That turns out to be the cross correlation method.

Vibroseis Deconvolution

According to the convolutional model, the vibroseis record is:

$$s(t) = \sigma(t) \bullet n_0(t) \bullet a(t) \bullet m(t) \bullet r(t) + noise(t)$$

Most of these terms were defined already in our discussion of the convolutional model. We repeat the definitions here:

$s(t)$ the uncorrelated vibroseis record

$\sigma(t)$ the vibroseis sweep

$n_0(t)$ near surface effects and vibrator distortion

$a(t)$ a convolutional approximation to Q effects

$m(t)$ the subset of all multiples which are convolutional

$r(t)$ the desired reflectivity

$noise(t)$ zero mean, white noise

Now we cross correlate with the sweep and rewrite the model as:

$$s_x(t) = \sigma(t) \otimes s(t) = w_v(t) \bullet w_e(t) \bullet r(t) + \sigma(t) \otimes noise(t)$$

where

$S_x(t)$ is the correlated vibroseis record

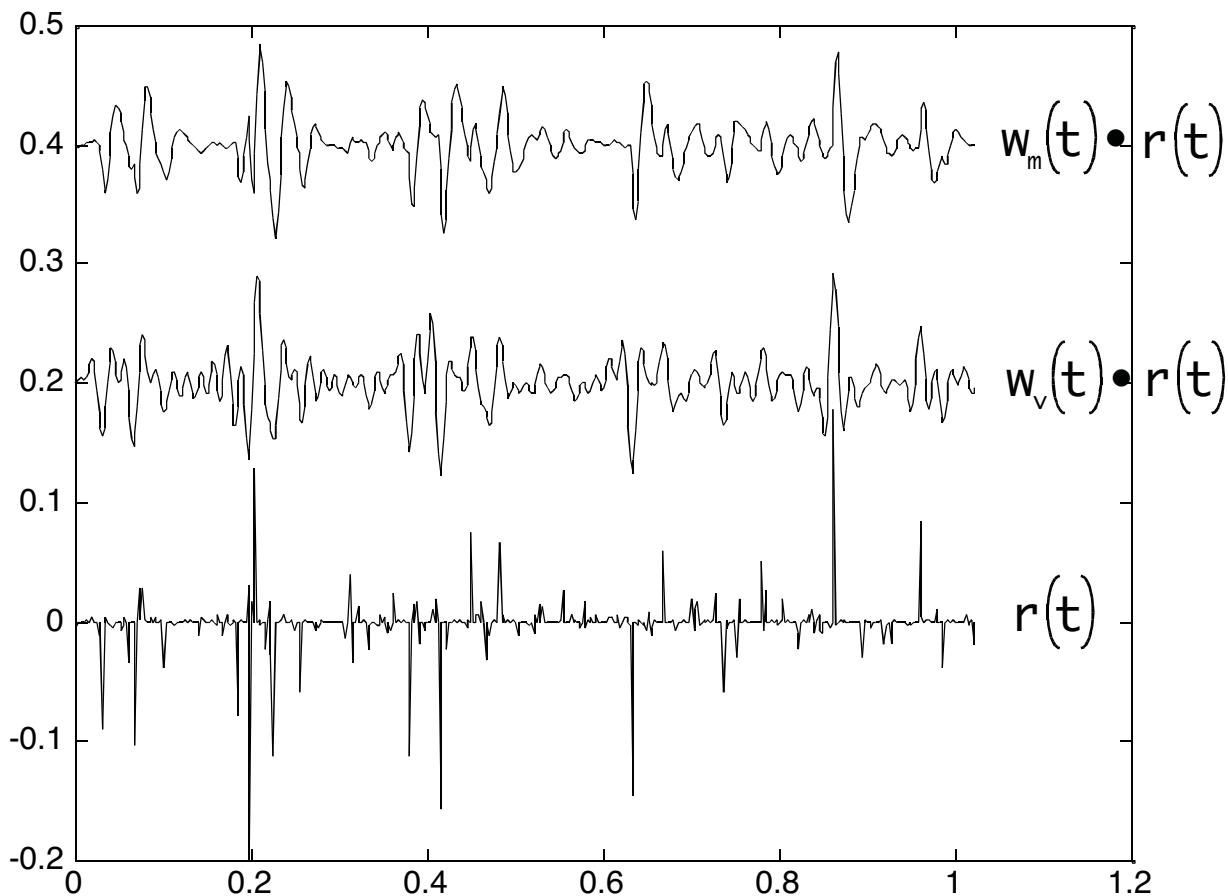
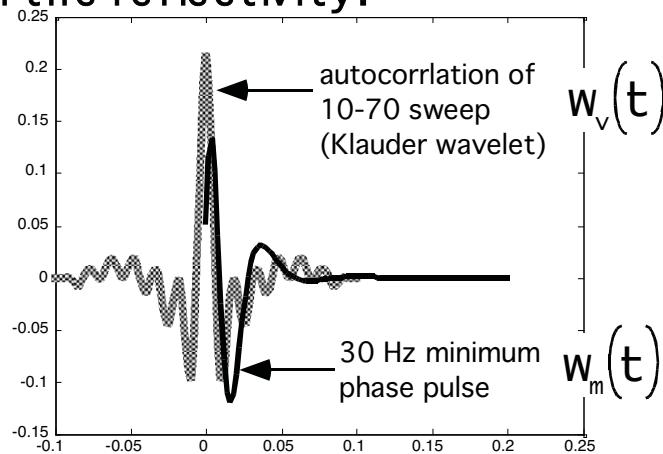
$w_v(t) = (\sigma \otimes \sigma)(t)$ is the autocorrelation of the sweep
(Klauder wavelet)

$w_e(t) = (n_0 \bullet a \bullet m)(t)$ is the effective “earth filter”

Roughly speaking, this says that we can use the convolutional model for correlated vibroseis data if we regard the source waveform as the autocorrelation of the sweep. This is often called the Klauder wavelet.

Vibroseis Deconvolution

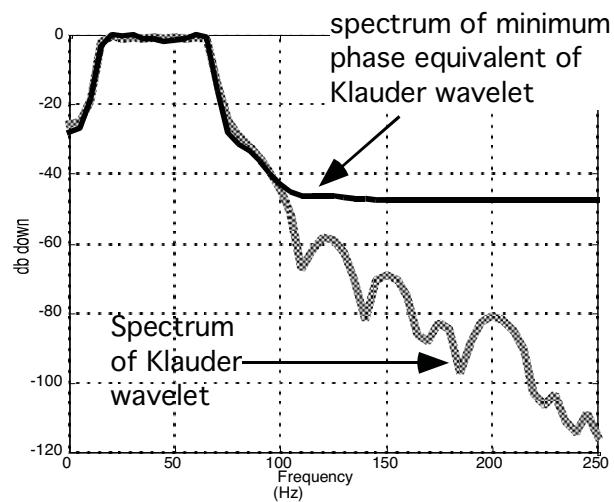
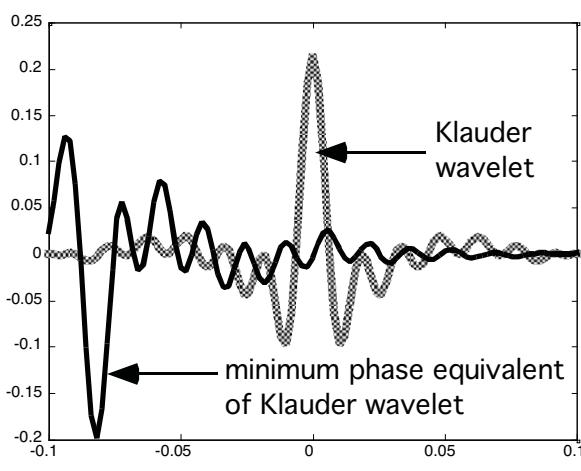
Thus, in the simplest case, we expect a correlated vibroseis record to be the sweep autocorrelation convolved with the reflectivity.



Vibroseis Deconvolution

Deconvolution of the vibroseis synthetic presents a special problem since we cannot assume the wavelet is minimum phase. Most approaches to this problem involve attempting to modify the correlated vibroseis record so that its embedded wavelet is more nearly minimum phase. An immediate problem arises because a minimum phase waveform cannot be bandlimited yet the vibroseis wavelet is explicitly bandlimited. This means that ALL methods which attempt to precondition the embedded vibroseis waveform must employ a white noise or "stab" factor to extend the spectrum. It is usually good practice to ensure that this factor is the same as that used later in the decon algorithm.

Given this, and assuming that the embedded wavelet is the Klauder wavelet, it is a straight forward exercise in signal processing to design a conversion operator which converts the Klauder wavelet to its minimum phase equivalent:

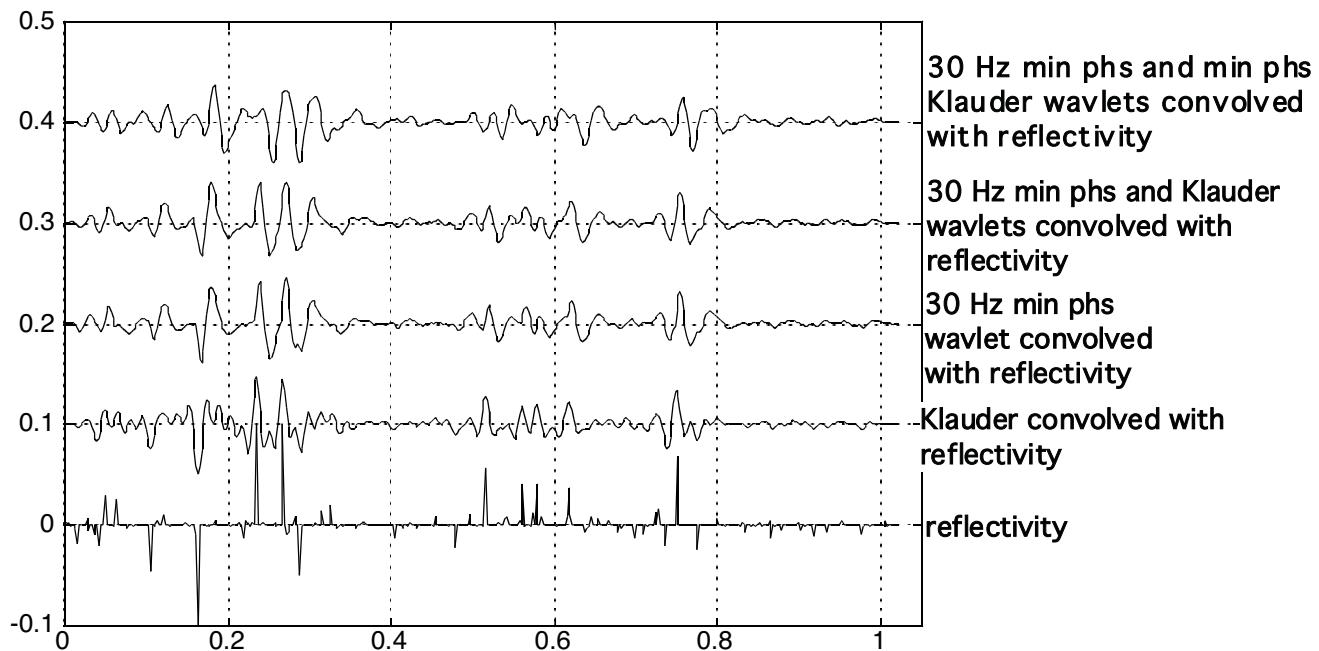


Vibroseis Deconvolution

At this point it is appropriate to ask why the vibroseis record should be deconvolved at all. After all, the signal is generated with a white spectrum over the swept band and is nominally zero elsewhere. Thus the zero phase vibroseis synthetic which consists of Klauder wavelet convolved with reflectivity is already optimal. The answer, of course, lies in the other earth filtering effects such as the near surface effect, multiples, and absorption (Q). Thus, a minimal vibroseis model for deconvolution theory is:

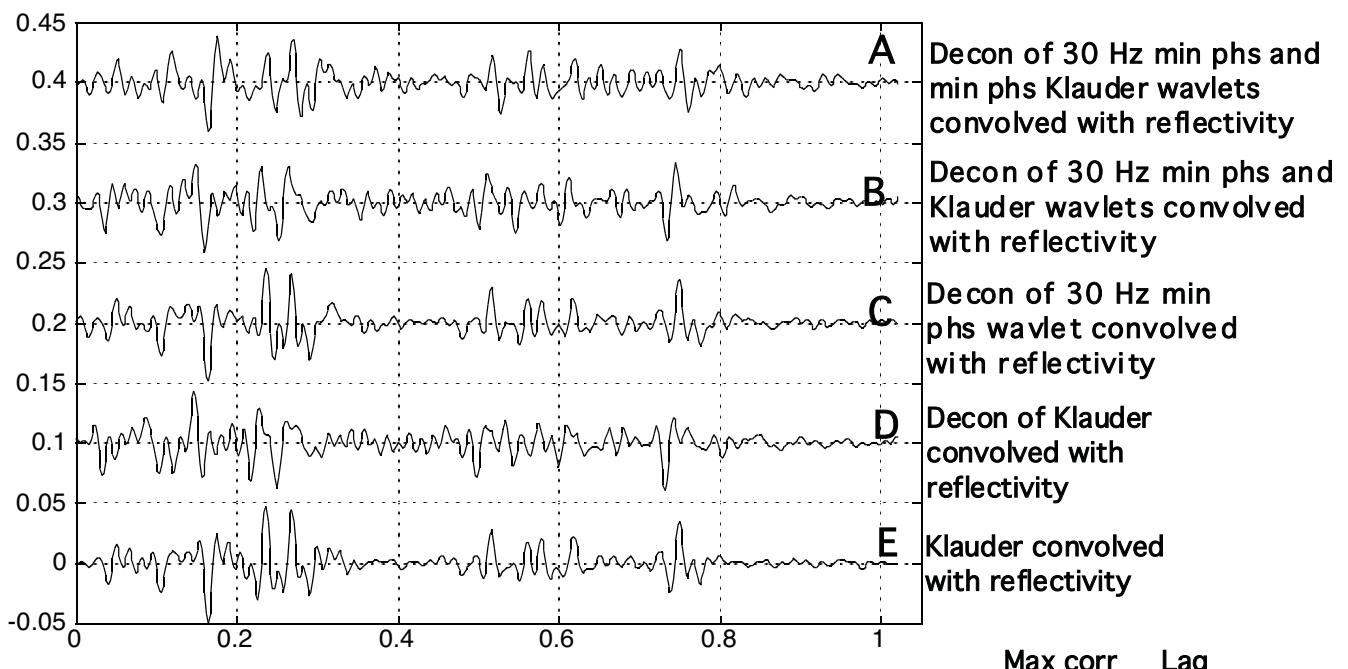
$$s_x(t) = w_v(t) \bullet n_s(t) \bullet r(t)$$

Unlike unlike the impulsive case, the goal of vibroseis deconvolution is to recover $r(t)$ only over the swept band, even in the noise free case. With this in mind, we use the 30 Hz minimum phase wavelet to represent the near surface effects and the Klauder wavelet for the source and create these synthetic seismograms:

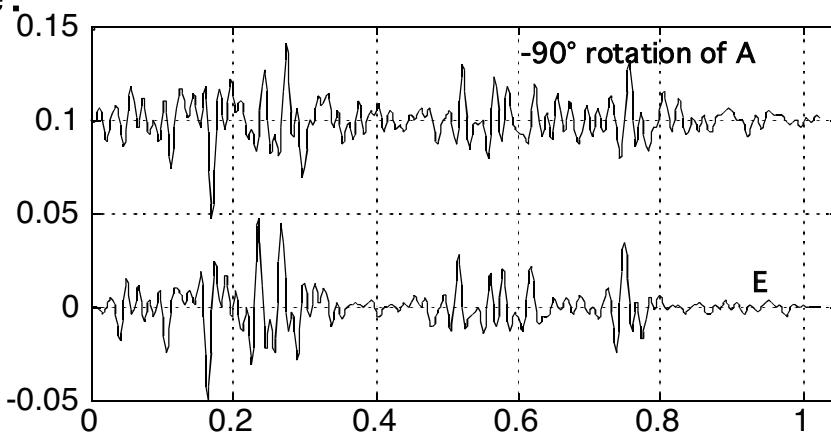


Vibroseis Deconvolution

The results of Weiner decon on these synthetics is below. Note that we compare to the reflectivity convolved with the Klauder wavelet and not to the reflectivity itself. All of these deconvolutions have been filtered back to the swept band.



We can see that the minimum phasing of the vibroseis record produces a better decon but it appears the result has a 90 degree phase rotation. In fact this is the case:



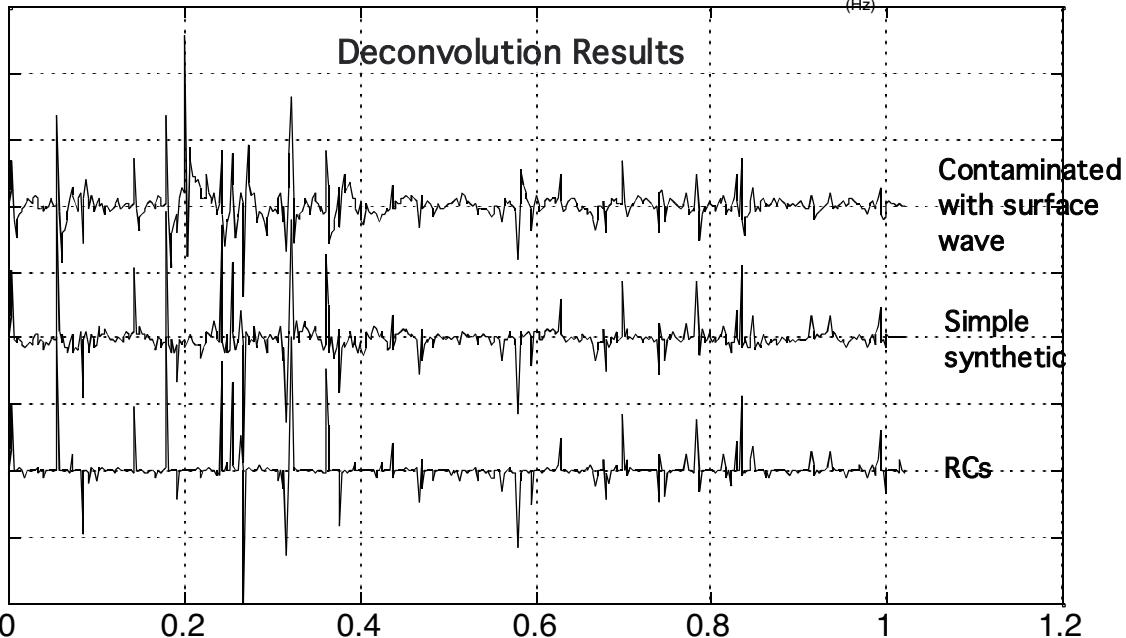
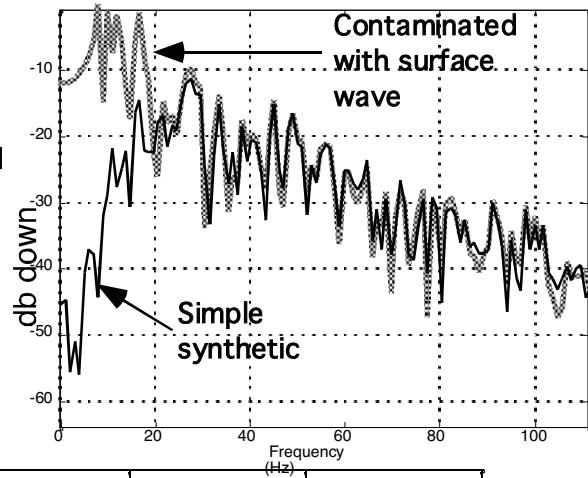
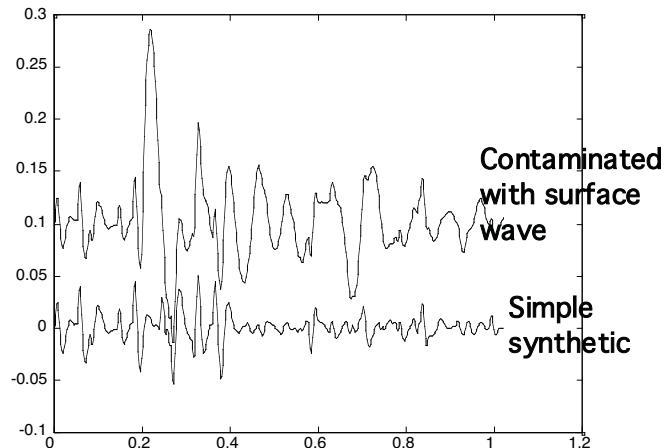
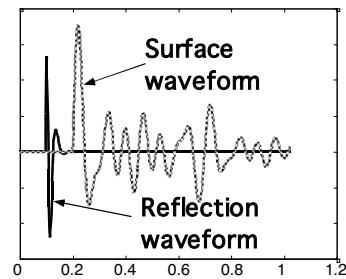
	Max corr	Lag
A	0.3795	0.8000
B	0.3712	-2.3000
C	0.4978	-0.1000
D	-0.4138	-9.1000

Deconvolution Pitfalls

The assumptions behind deconvolution theory help us to understand its basis and, sometimes, to anticipate problems before they arise. Here we will examine some common deconvolution "pitfalls".

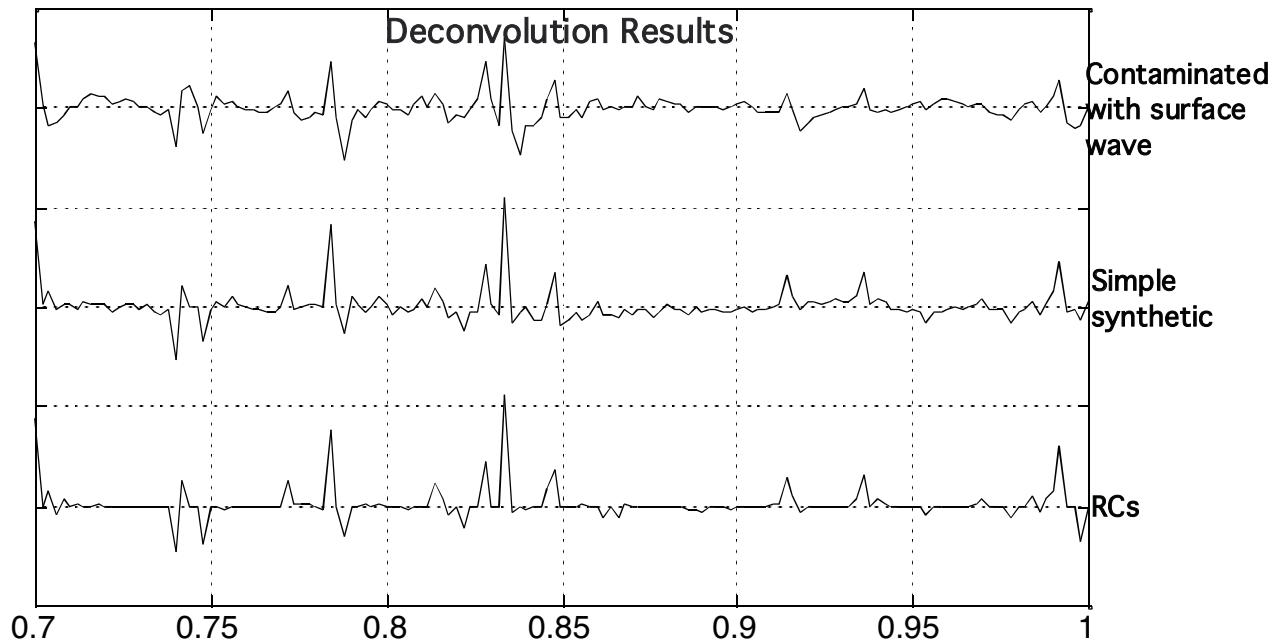
Mixed-wavetypes in the design gate.

The most common example here is the occurrence of a surface wave or similar coherent noise train in the design gate. The simulated surface wave begins at .2 seconds.



Deconvolution Pitfalls

Here is a closeup of the ends of the traces so that the considerable phase distortion can be appreciated:



So we see that the presence of the surface wave has caused a great deal of phase distortion even quite far from the onset of the wave. Since the phase corrections applied by minimum phase deconvolution are deduced from a smoothed representation of the amplitude spectrum, the presence of the surface wave peak in the amplitude spectrum causes erroneous phases to be computed.

Deconvolution Pitfalls

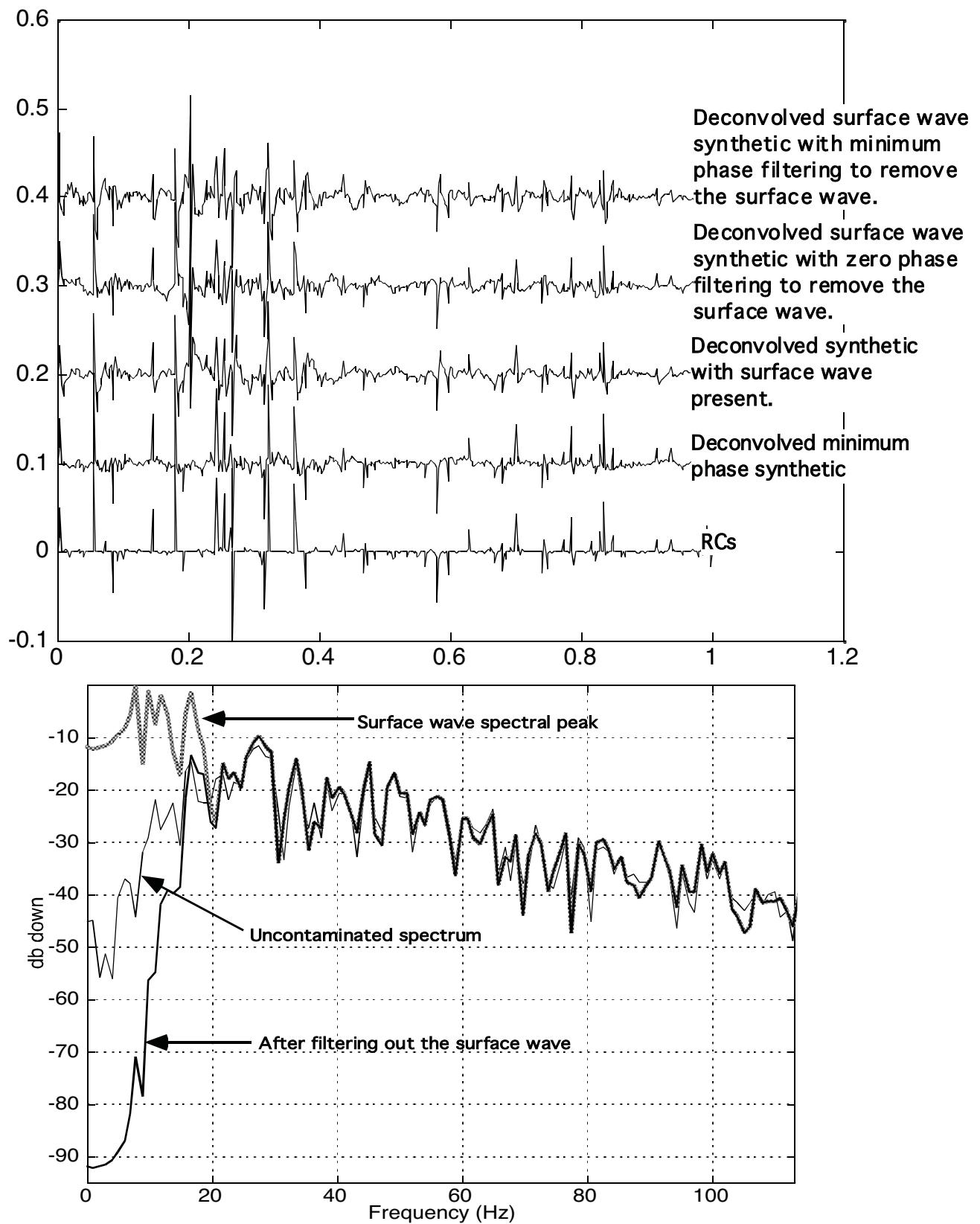
Filtering before deconvolution.

The issue of filtering before deconvolution can be a complex one which takes some surprising twists. It might seem that one could dogmatically insist that all filtering before deconvolution must be minimum phase. However; as we shall see, that is often incorrect. It greatly helps the decision process to consider whether the unfiltered data is in the "minimum phase state" or not.

We will say that seismic data is in the "minimum phase state" if there is a single embedded wavelet and that wavelet is minimum phase. If data is in the minimum phase state, then any filtering should be minimum phase to try to preserve that state. If not, zero phase filtering might actually be preferred if it can be argued to move the data towards the minimum phase state.

The surface wave synthetic which we presented earlier is not in the minimum phase state because it contains two embedded wavelets: the minimum phase reflection wavelet and the non-minimum phase surface waveform. Therefore, minimum phase filtering before deconvolution might not be appropriate. In fact, a zero phase filter designed to knock down the surface wave peak should move the data towards the minimum phase state. The example on the next page shows that this is the case.

Deconvolution Pitfalls

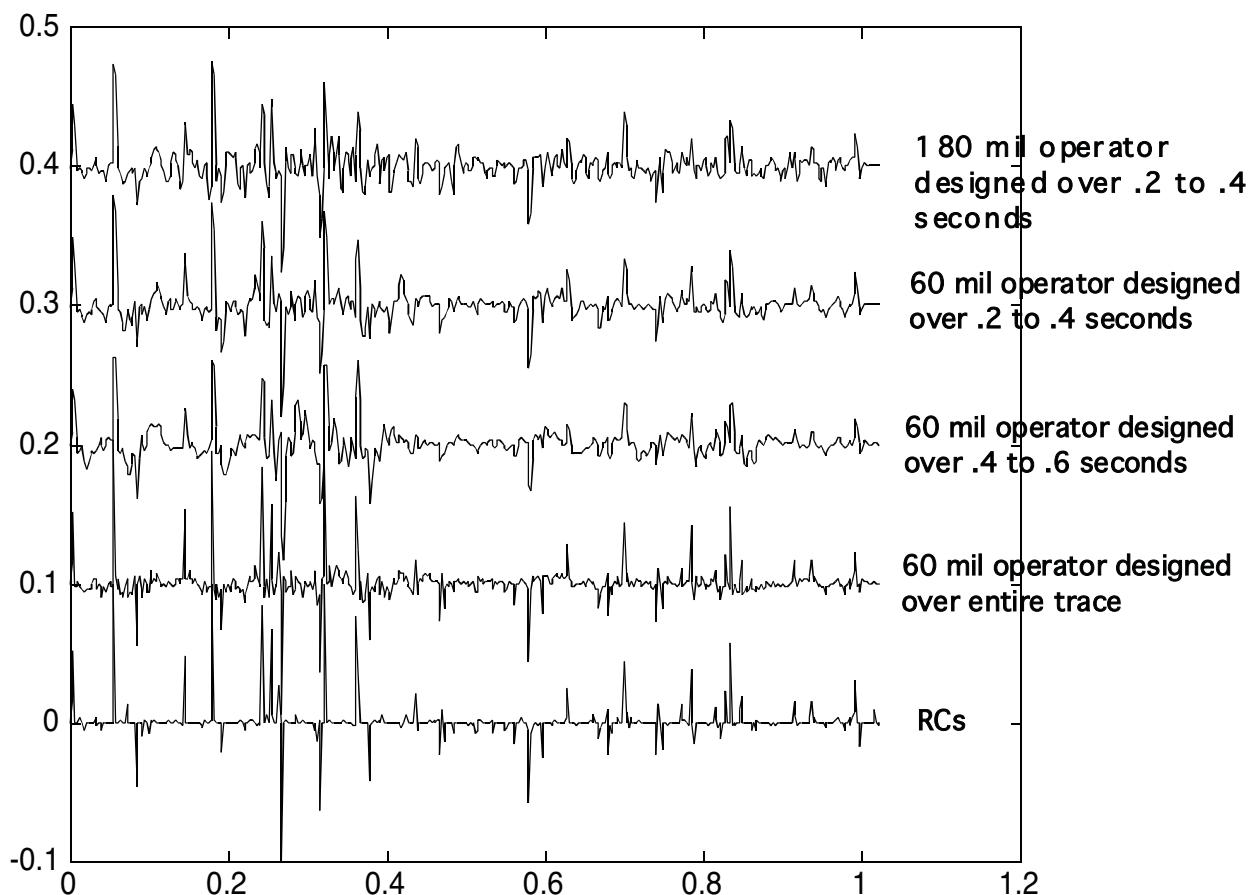


Deconvolution Pitfalls

Design gate considerations.

Typically the deconvolution operator is designed over a subset of the trace chosen for its high signal to noise ratio. Considerations:

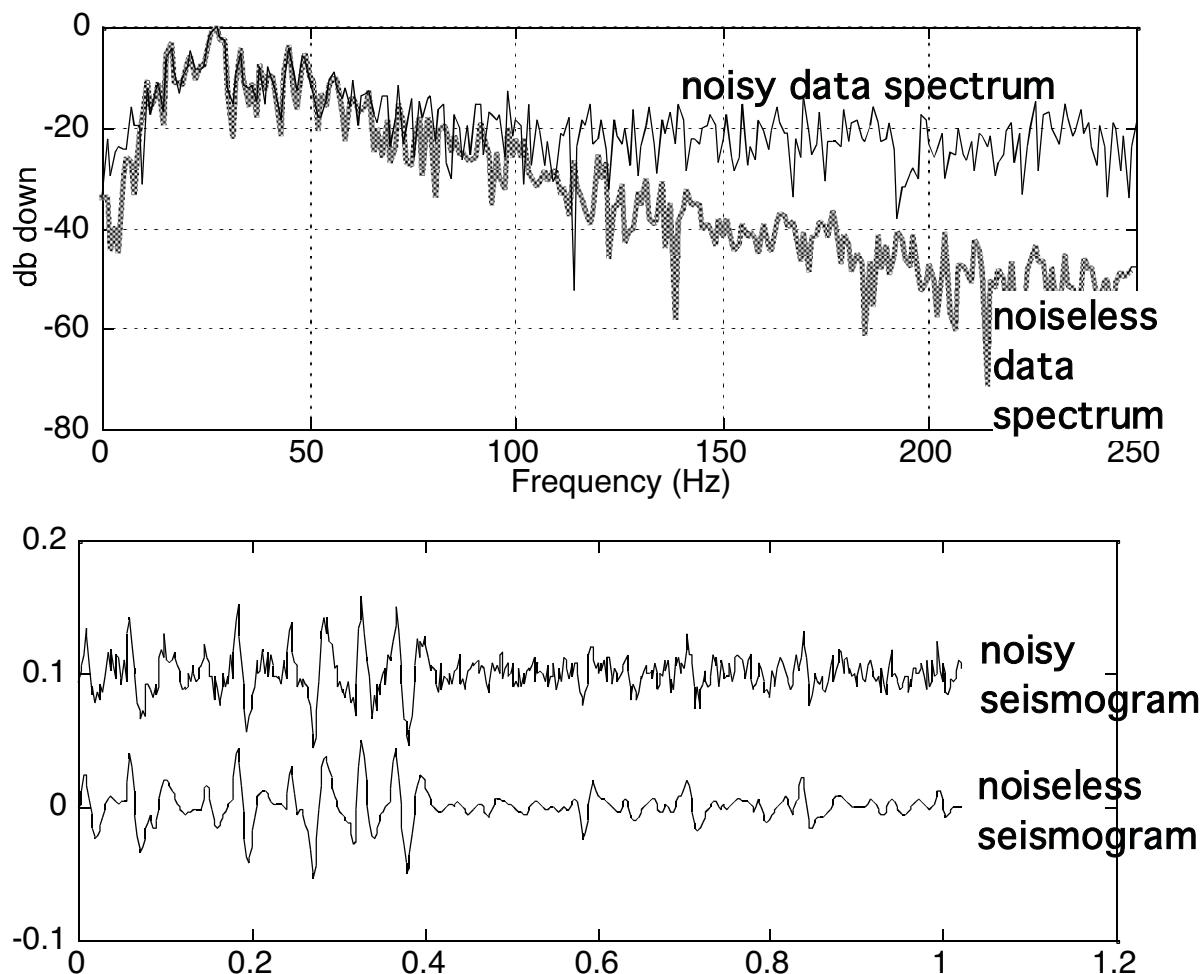
- include the zone of interest
- include large dominant reflectors
- exclude surface waves and below basement
- don't design on noise
- highly non stationary data should avoid very long gates
- operator length should be no more than 1/3 to 1/2 of the gate length



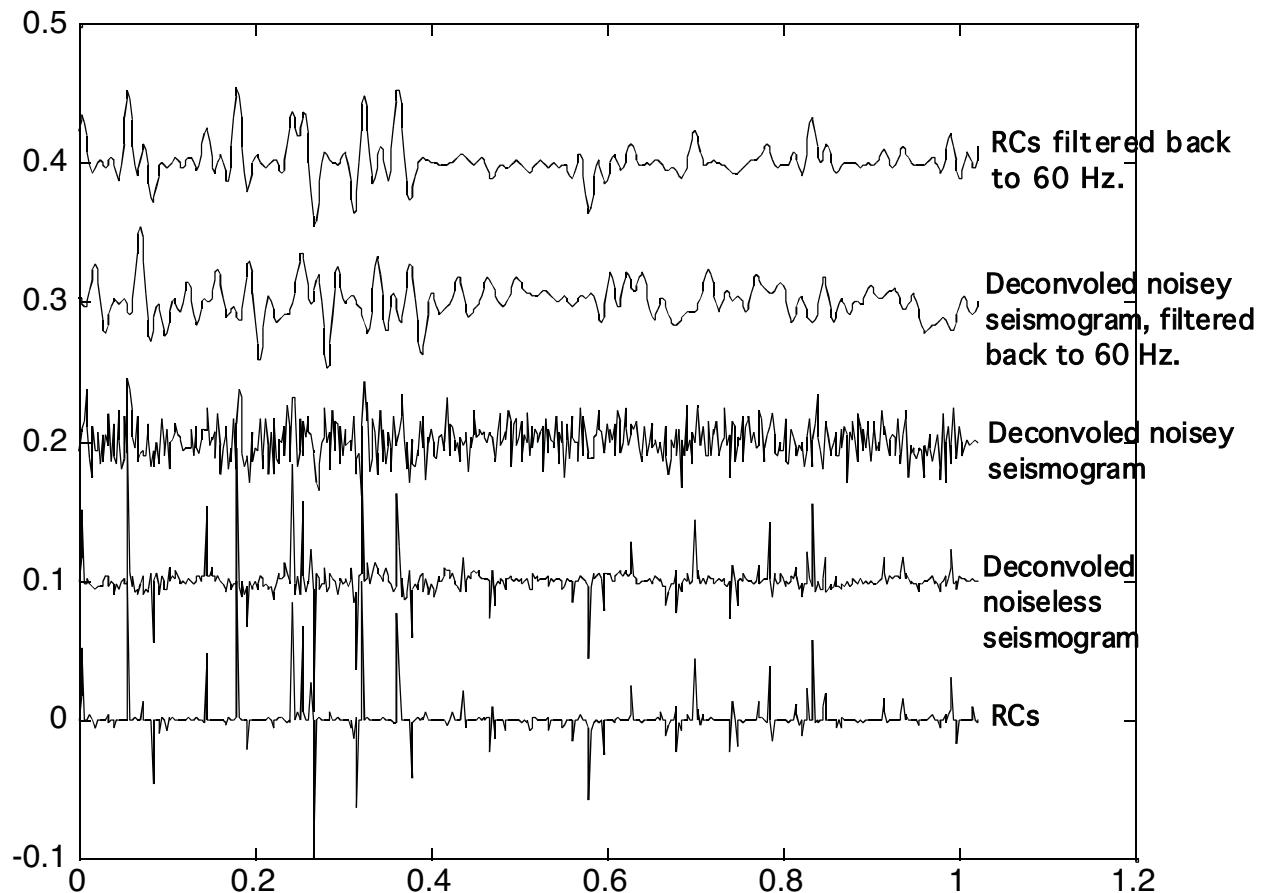
Deconvolution Pitfalls

Filtering after deconvolution.

Real seismic data always contains a great deal of apparently random noise. We've already mentioned the inadvisability of designing the operator on noisy data. Also, it is almost always necessary to filter data back to some signal band after deconvolution. Most decon algorithms cannot distinguish signal from noise and so whiten both. This can have a disastrous effect on such noise sensitive programs as residual statics.



Deconvolution Pitfalls



Though the results of the deconvolution of noisy data are almost always better when filtered back, in general, the match to the RC's is still much worse than without noise.

Deconvolution Pitfalls

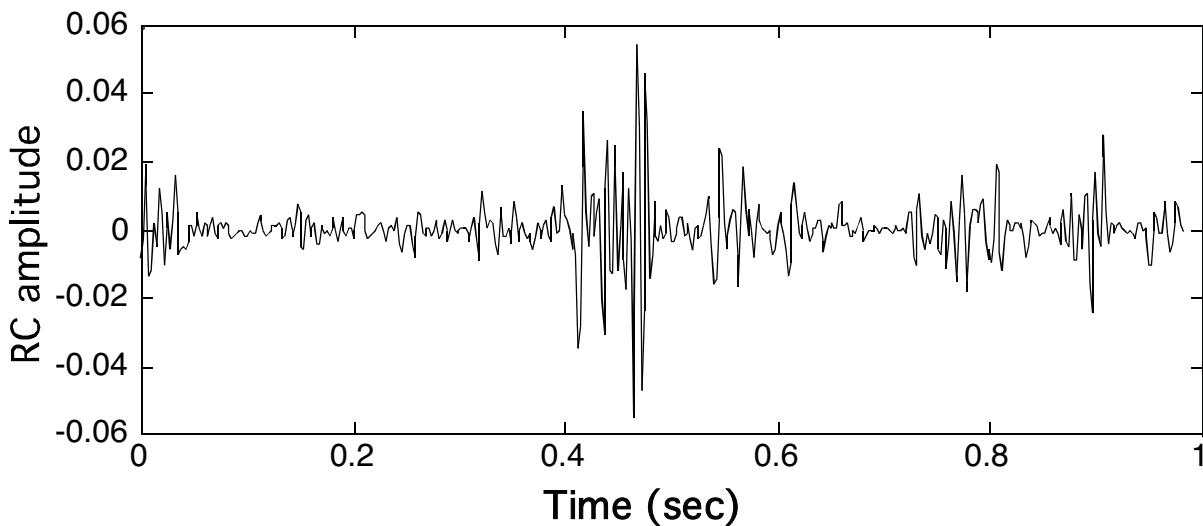
Iterative deconvolution.

It is often assumed that deconvolution is something that needs doing once and is then best forgotten. This attitude usually leads to underwhitened data with residual phase rotations. Since the assumptions of deconvolution are never precisely met, it is often useful to apply several different decons for different purposes. For example, we may use predictive decon to attack a multiple and then spiking deconvolution to sharpen resolution.

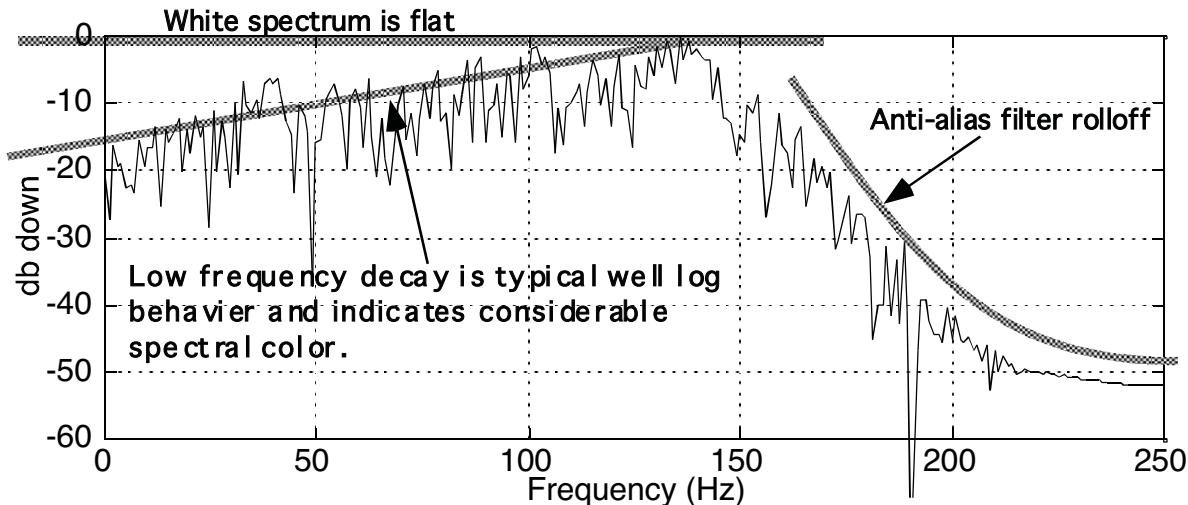
More importantly, deconvolution algorithms cannot distinguish between signal and noise. Thus we must think of them as whitening the spectrum of signal plus noise. If deconvolution is then followed by any process which can reject noise while retaining signal, the resulting data will have a non-white, lower resolution, spectrum. The most common example of this is CMP stacking. Thus it is often necessary to run a post-stack deconvolution or whitening step to ensure maximum resolution. If post-stack minimum phase deconvolution is desired, care should be taken to ensure that zero-phase filtering was not done after the pre-stack deconvolution.

Reflectivity Color

Here is a reflectivity series computed from an Alberta well and plotted versus time.



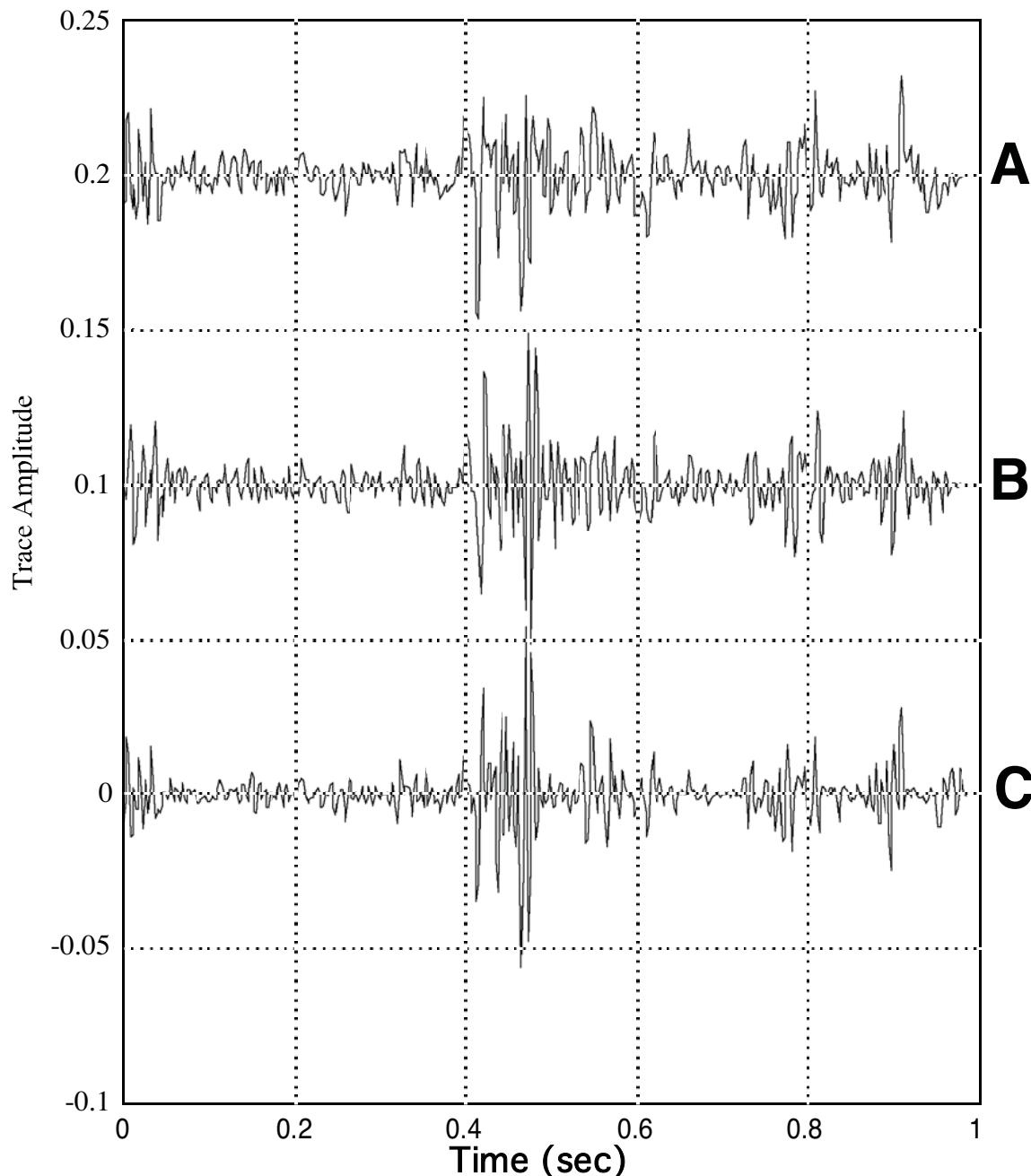
Since standard deconvolution algorithms assume a white reflectivity spectrum, we are motivated to compute the spectrum and see if it is white.



So we see that this spectrum is non-white.
What color is it?

Reflectivity Color

Example of a reflectivity estimate via Weiner deconvolution for a non-white reflectivity. The traditional whitened estimate is shown (A) along with a color corrected estimate (B) and the original well log (C).



Reflectivity Color

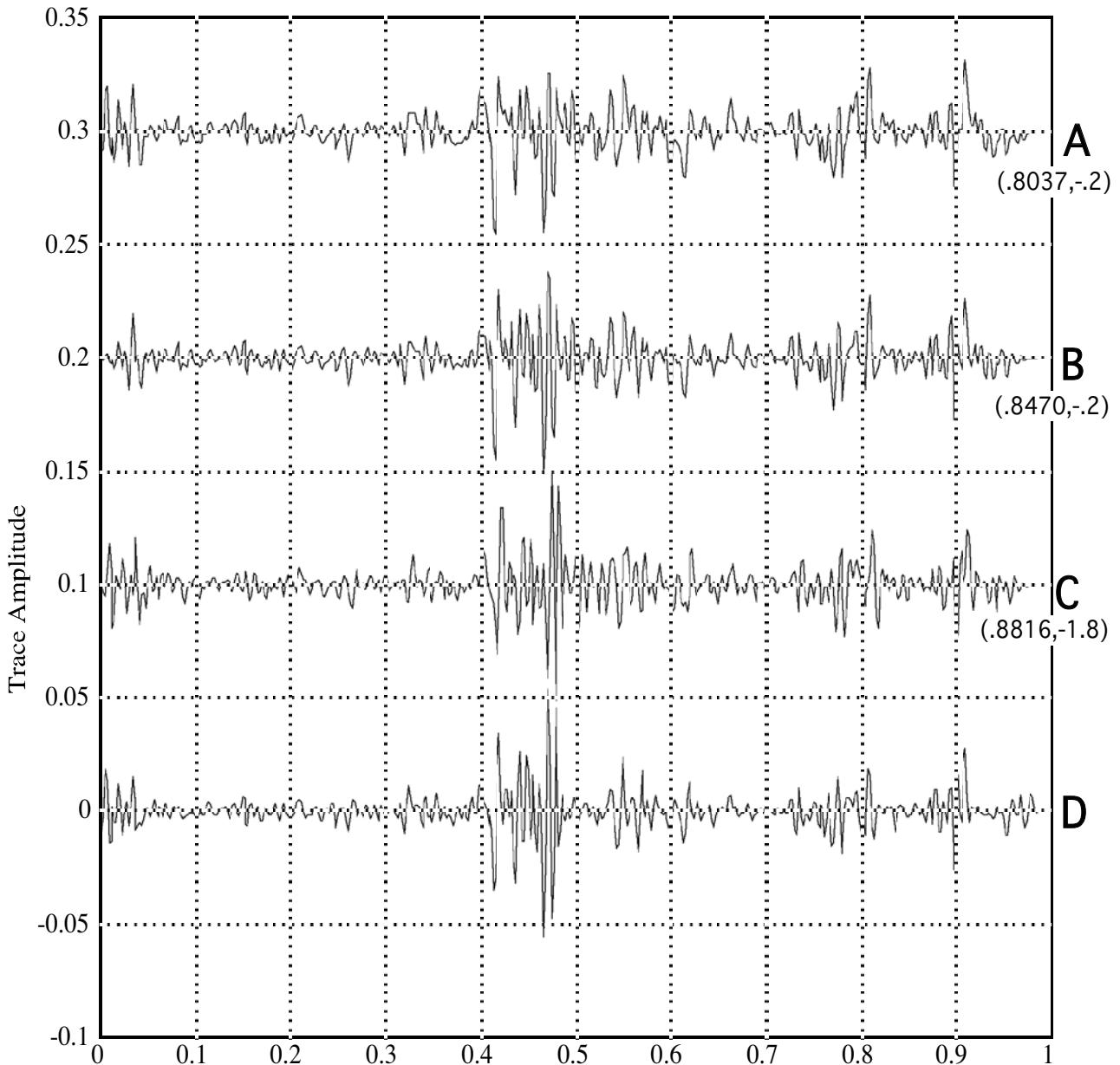
A: Reflectivity estimate from normal Weiner decon

B: A convolved with a zero phase color restoration operator

C: A convolved with a minimum phase color restoration operator

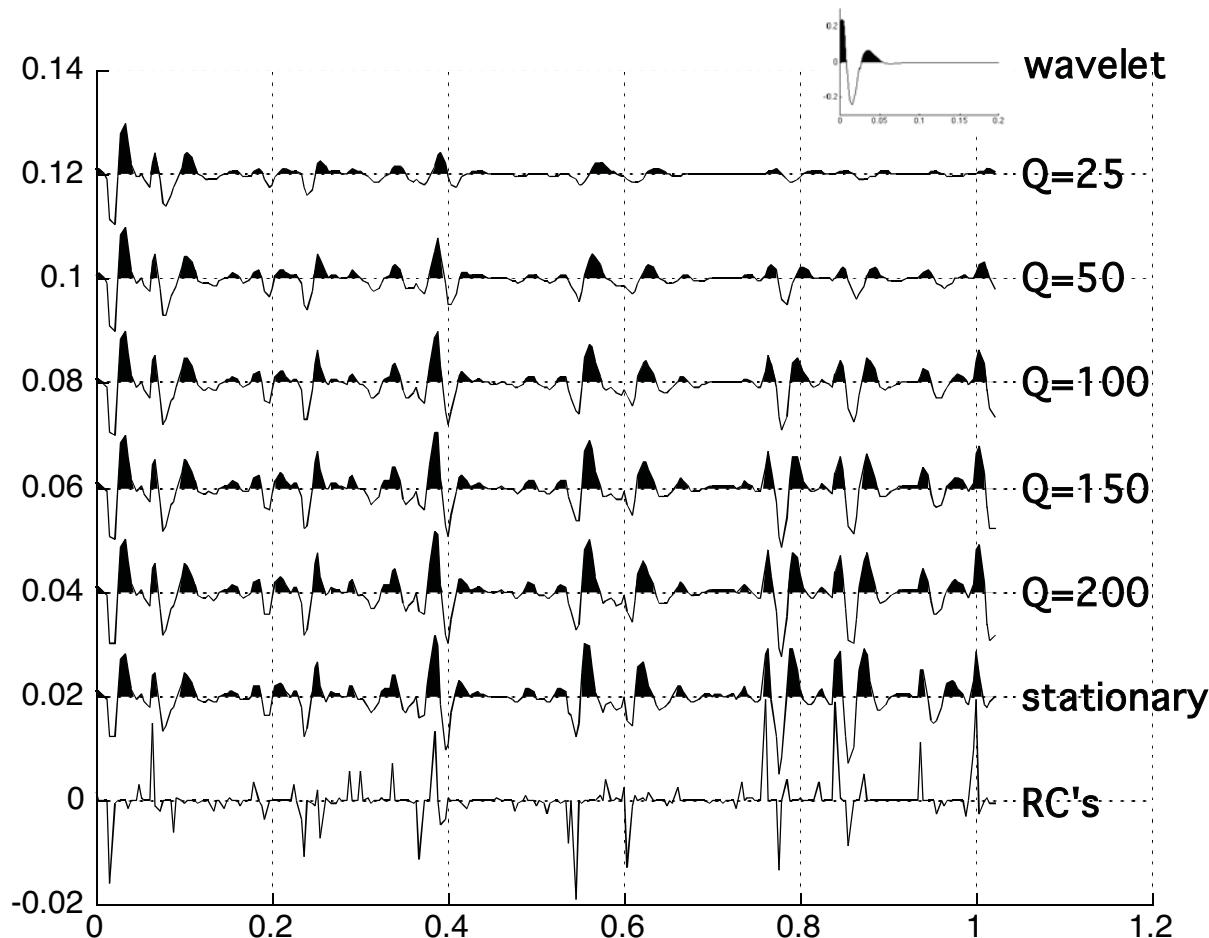
D: Original well log at 2 ms sample rate

Numbers give (Maximum correlation coeff, lag at max {samples}) for the correlation between an estimate and the answer given in D



Q Example

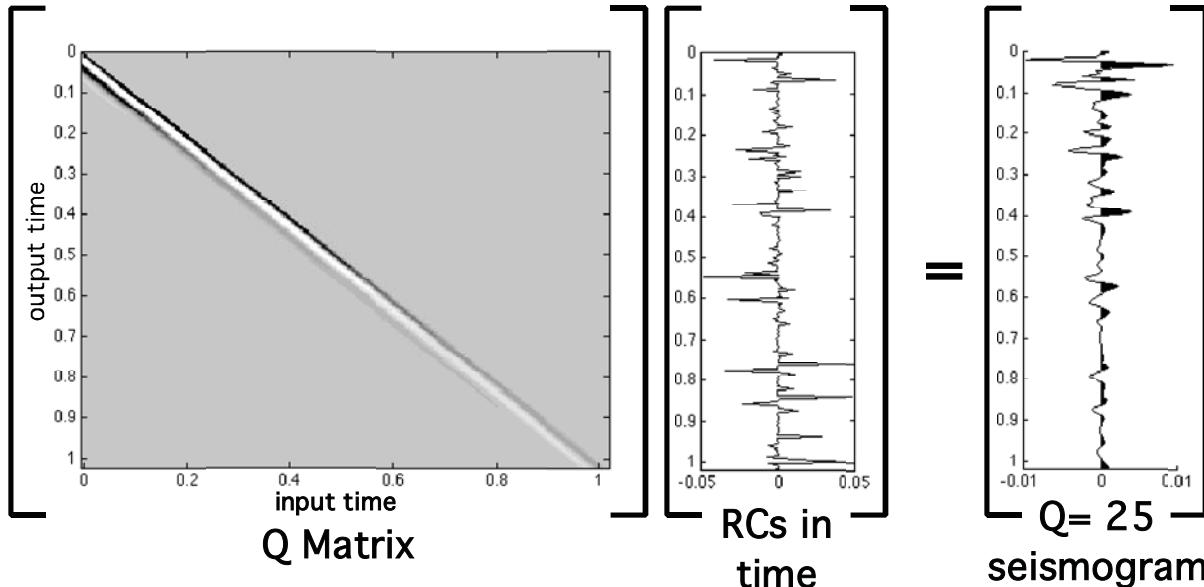
Here is an example comparing a stationary, minimum phase synthetic with a series of constant Q synthetics. Each constant Q synthetic has the same 30 Hz, minimum phase wavelet convolved with it as the stationary synthetic does.



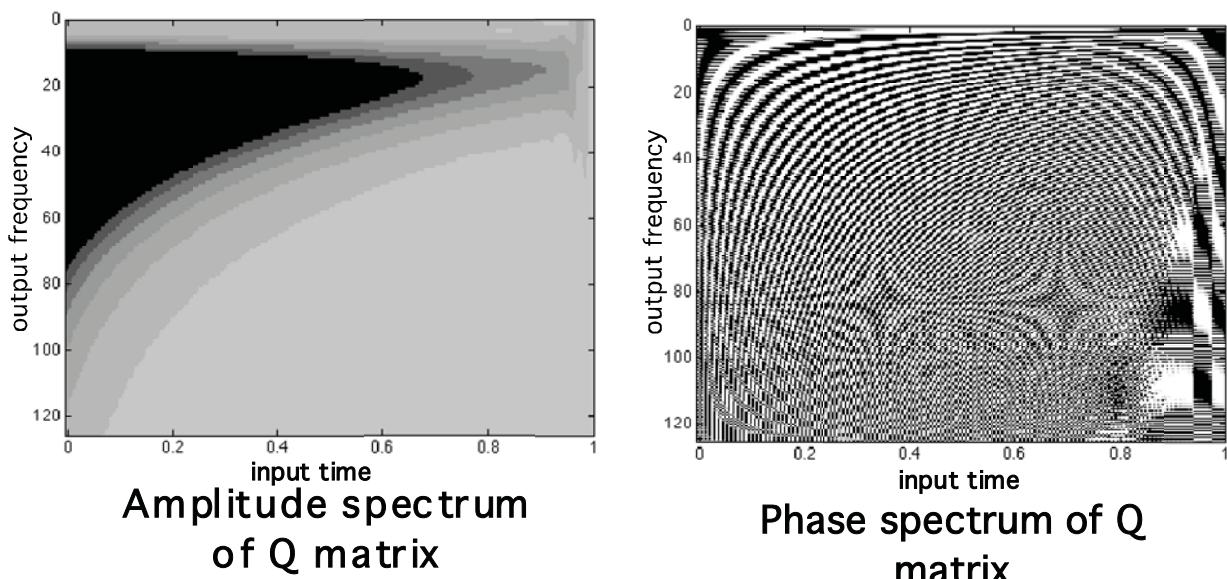
The effect of Q attenuation can be seen to have at least three characteristics: a progressive loss of frequency content with increasing time, a progressive loss of overall amplitude, and a progressive time delay. The construction of one of these synthetics is detailed on the next page.

Q Example

Each of the Q synthetic was created by first constructing the "Q matrix" which applies a Q response to a time series via a generalized convolution. Here the process is depicted graphically for the Q=25 case.

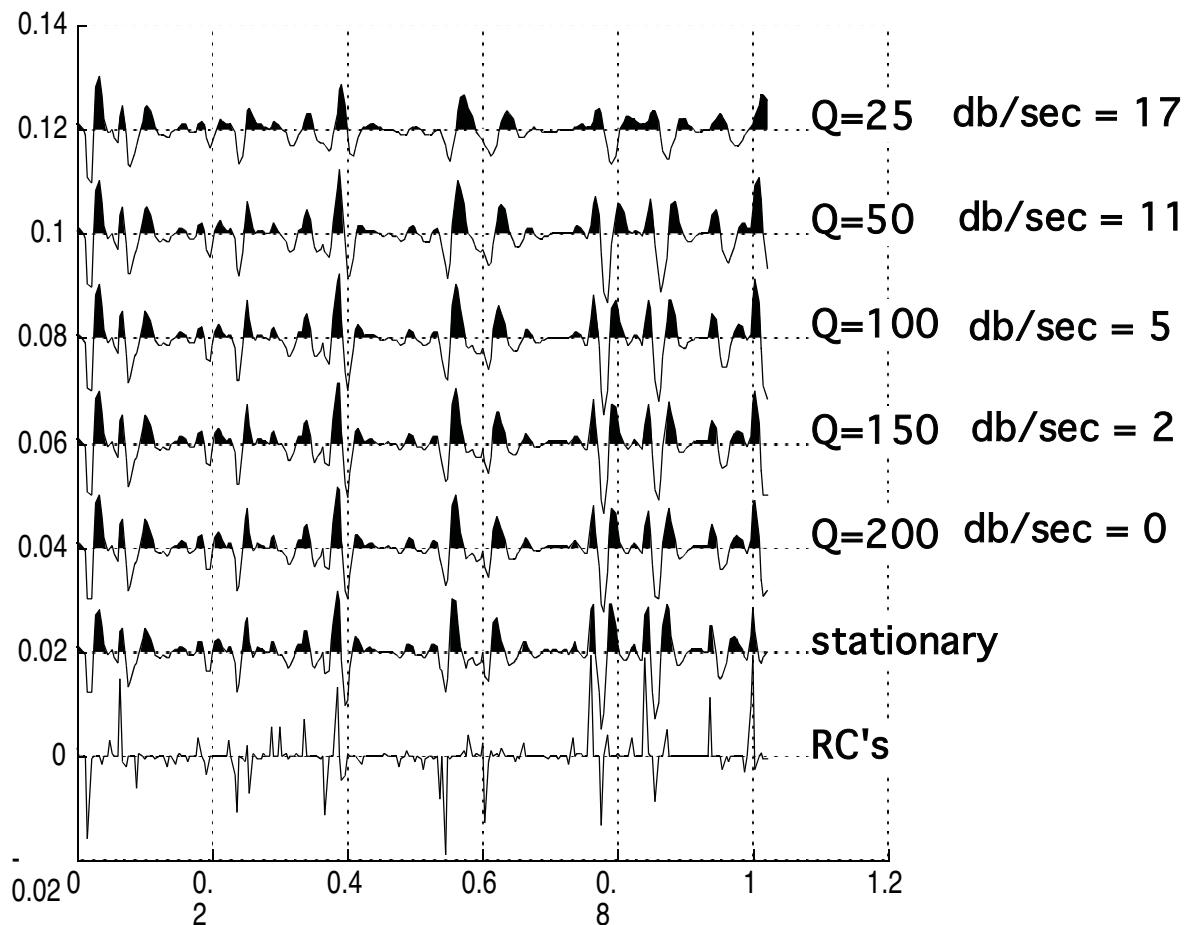


Each column of the Q matrix contains the $Q=25$ impulse response for the input time of the column convolved with the 30 Hz minimum phase source waveform. If we Fourier transform each column, we can see directly the Q amplitude and phase response:



Q Example

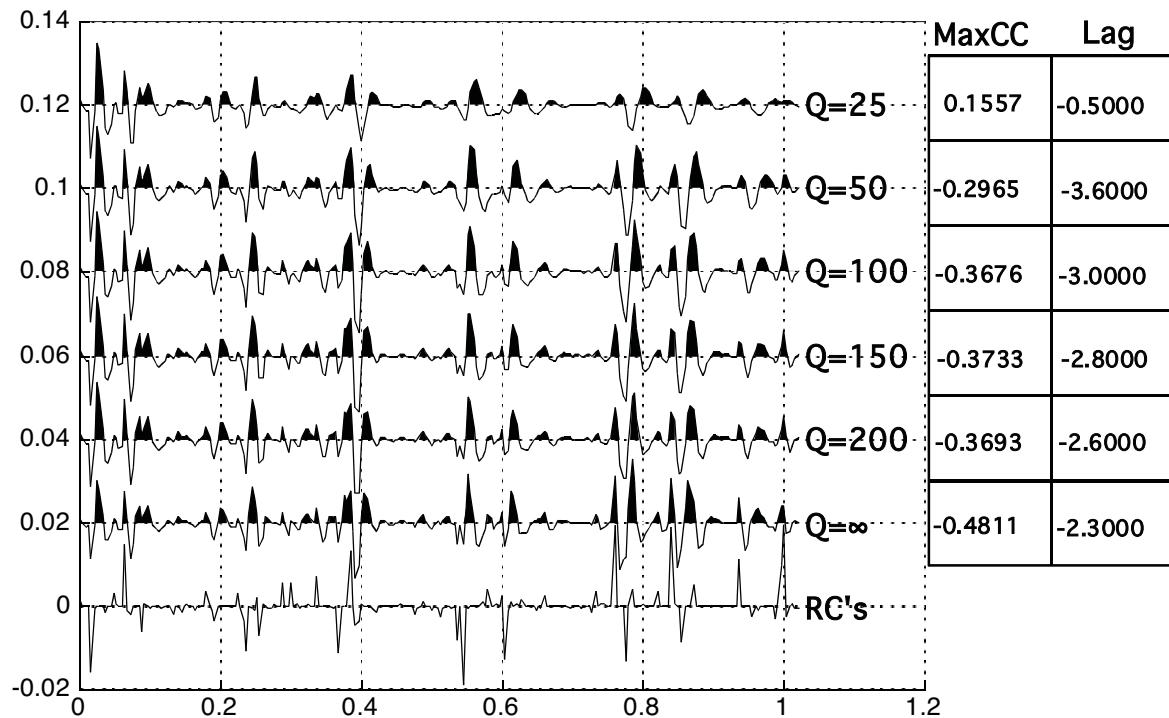
The first step in deconvolving the Q synthetics is to determine exponential gain corrections and apply them. The result is:



These gain factors were determined empirically as is standard practice. It appears that $Q=25$ may be a bit under gained.

Q Example

Next we run Weiner deconvolution with the same parameters for each trace (30 lags and .0001 white noise).



It's clear from this example that the deconvolution result degrades steadily with decreasing Q. Keep in mind that this is a "best case" scenario: no noise, no multiples, minimum phase source, and white reflectivity. Also, even the Q=25 case is not an unreasonable attenuation level because the maximum time in the synthetic is only 1 second. Since t/Q determines the actual attenuation, $1/25$ is the same as $2/50$ or $3/75$.

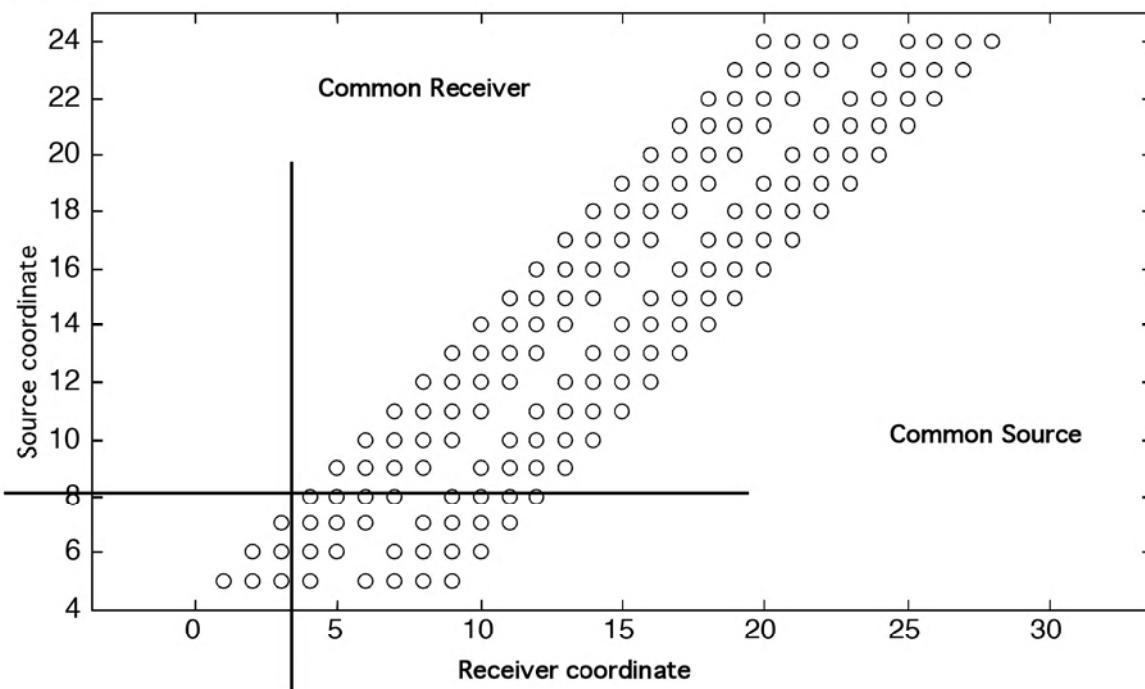
Methods of Seismic Data Processing

**Lecture Notes
Geophysics 557**

**Chapter 5
Surface Consistent Methods**

Seismic Line Coordinates

The acquisition geometry of a seismic survey can be completely described by specifying the coordinates of each shot, $s = [x_s, y_s, z_s]$, and the receivers which are active for each shot, $r(s) = [x_r, y_r, z_r]$. To simplify the discussion, let us assume a 2-D line with all sources and receivers on the $z=0$ plane. Then, we can drop the vector notation and label the j th shot, s_j , and the k th receiver, r_k . If recording arrays are present then these coordinates refer to the array centers. A source-receiver diagram can then be made which shows the location of each source the corresponding live receivers:



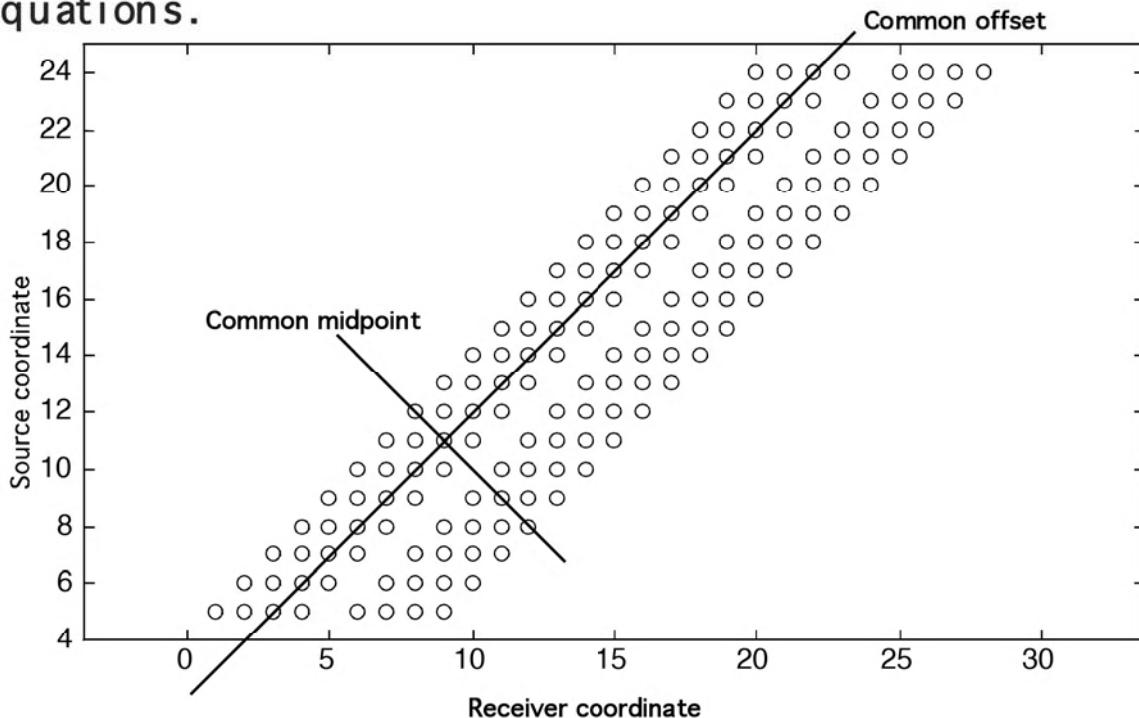
Here we see a source receiver diagram for a live receiver spread consisting of four phones on either side of the source. The source moves by one receiver interval, and the live phones are "rolled along" to record each source record.

Seismic Line Coordinates

Seismic data are necessarily recorded in "common source gathers" but we see that we can equally well consider "common receiver gathers" by collecting one trace from each source record which has the considered receiver live. The ability to sort data from one gather to another and to run processing algorithms on these gathers is fundamental to data processing. The coordinate system (s,r) is often called the acquisition coordinates while another system (x,h) are the "processing coordinates". Source-receiver midpoint, x , and source-receiver half-offset, h , are defined by:

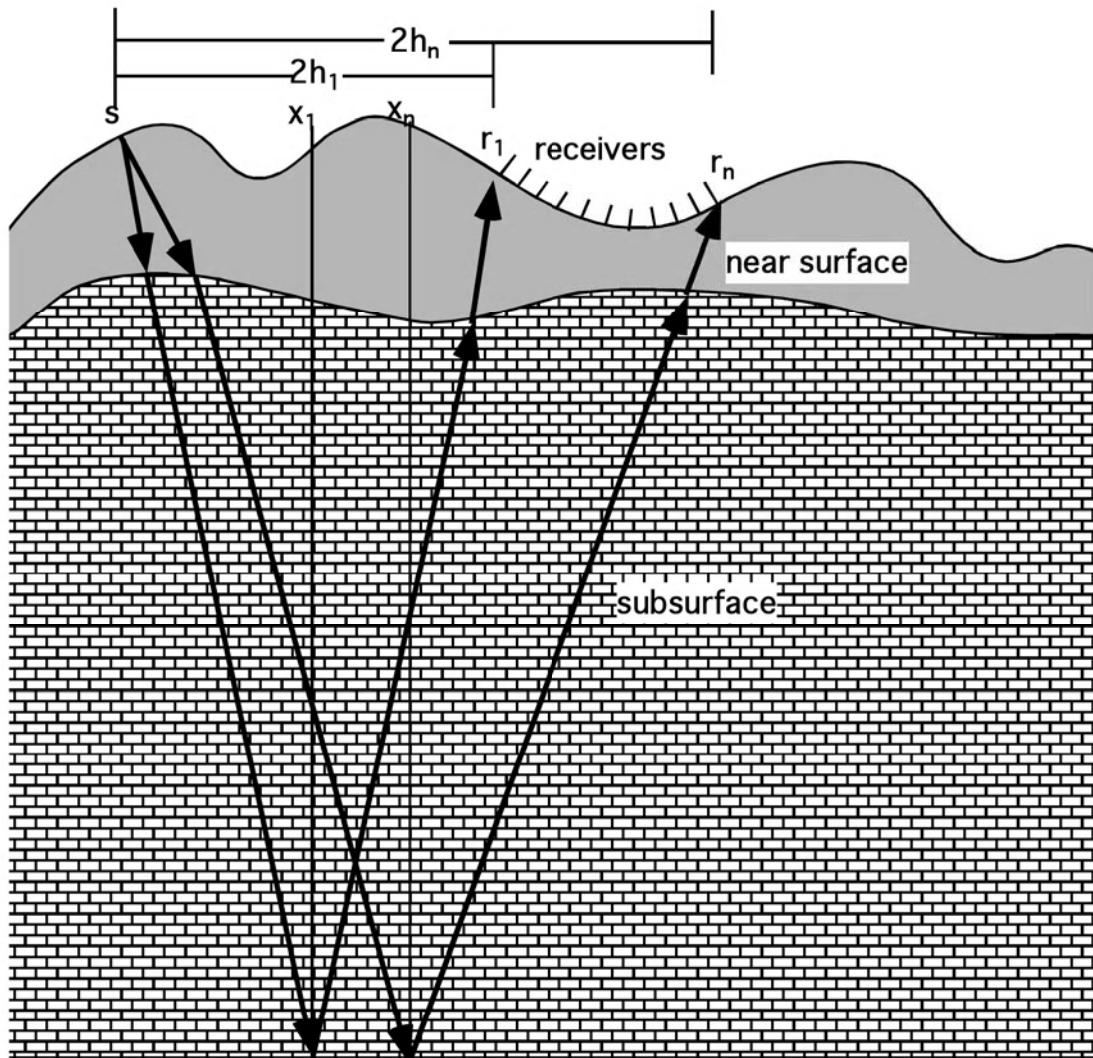
$$x = \frac{1}{2}(s + r) \quad h = \frac{1}{2}(r - s)$$
$$s = x - h \quad r = x + h$$

The half-offset, h , is used in preference to the full offset, H , to provide a more symmetrical set of equations.



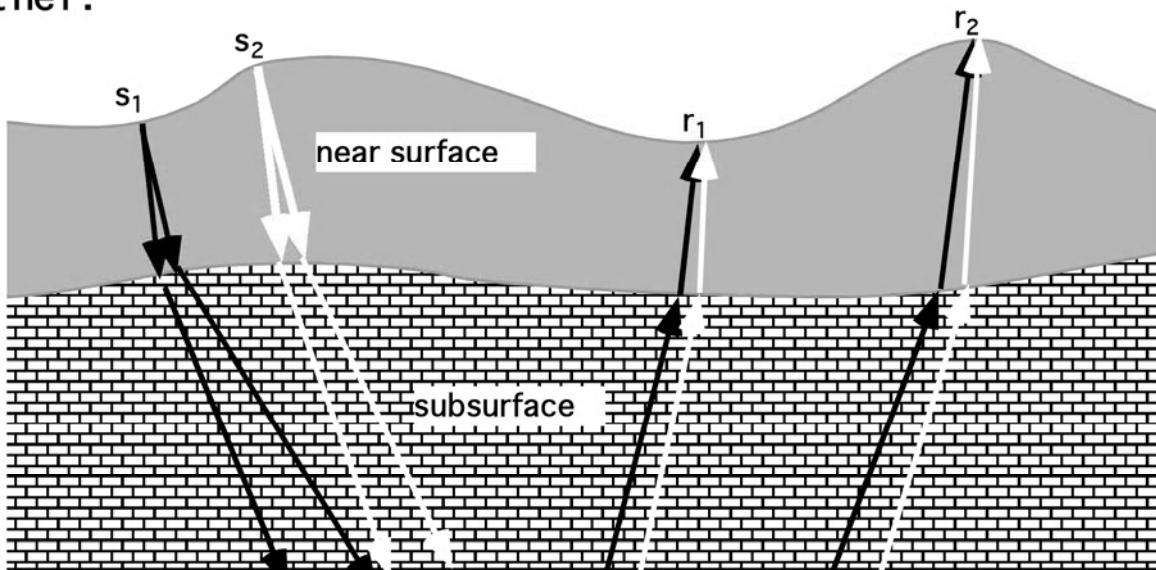
Seismic Line Coordinates

It is generally assumed that the earth consists of a rapidly varying near surface and a slowly varying subsurface. The near surface effects are considered to be only a function of (s, r) while in the subsurface, wave propagation is a function of (x, h) . The division between near surface and subsurface is arbitrary and vague and is not usually necessary to define. Instead we simply search the data for (s, r) consistent effects and attribute those to the near surface and similarly for (x, h) and the subsurface.



A Surface Consistent Convolutional Model

It is quite common (and quite successful) to consider the earth's effect on a seismogram to be divided into those caused by the near surface and those caused by the subsurface. The division between near and subsurface is only vaguely defined as somewhere near the "base of the weathering" or "the second refractor". In practice, near surface effects are judged to be those which are seen to vary only as a function of source and receiver coordinates (s, r). Subsurface effects are generally those which vary as a function of midpoint and offset (x, h). As might be expected, there are many effects which do not fall clearly into one class or the other.



Examples of some near surface effects are: attenuation beneath the source and receiver, source ghosts, source waveforms, source and receiver arrays, statics, surface waves, ...

Examples of some subsurface effects are: Q attenuation, spherical spreading, moveout, most classes of multiples, ...

A Surface Consistent Convolutional Model

Recalling the 1-D convolutional model:

$$\sigma(t) = w_s(t) \bullet ns(t) \bullet m(t) \bullet r(t) + noise(t)$$

$\sigma(t)$ is the seismic trace

$w_s(t)$ is the source waveform (signature)

$ns(t)$ contains all other near surface effects

$m(t)$ contains all convolutional multiples

$r(t)$ is the reflectivity we wish to estimate

$noise(t)$ is white noise

We now extend this model to 3-D in a surface consistent sense by:

$$\sigma(t) = w_s(s, t) \bullet ns(s, r, t) \bullet m(x, h, t) \bullet r(x, h, t) + noise(t)$$

The near surface effects are ideal candidates for stationary decon since they are more likely to be well modeled as stationary convolutions than the subsurface effects. Let us now decompose the near surface term:

$$ns(s, r, t) = a_s(s, t) \bullet m_s(s, t) \bullet a_r(r, t) \bullet m_r(r, t) \bullet rec(r, t)$$

Here a_s and a_r are attenuation responses beneath source and receiver and m_s and m_r are ghost responses and near surface peg-leg multiples for source and receiver. The rec term symbolizes the response of the receivers and recording instrument.

A Surface Consistent Convolutional Model

The attenuation terms are most likely Q effects in the near surface and can be considered as a stationary operations on the subsurface reflectivity. The near surface multiples and ghosts are also well known to be stationary effects.

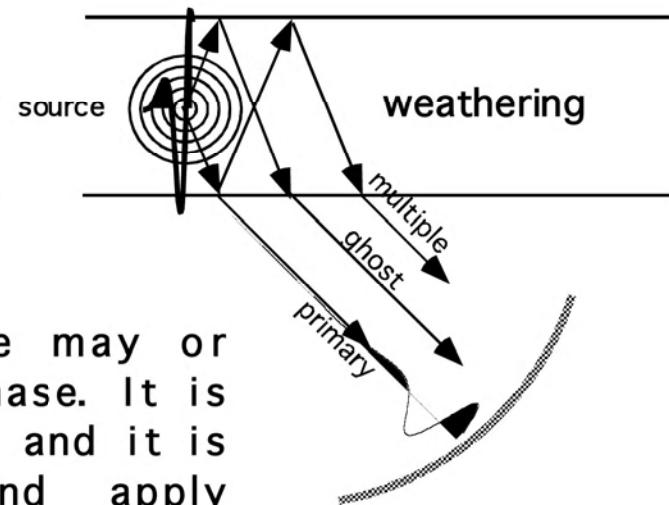
Additionally, both of these terms are arguably minimum phase so they are good candidates for deconvolution.

The instrument response may or may not be minimum phase. It is generally a known effect and it is common to design and apply instrument dephasing operators.

There are other near surface effects that can cause deconvolution problems and that we do not account for such as surface waves, acquisition arrays, etc.

The subsurface effects are grouped as: $m(x, h, t) \cdot r(x, h, t)$

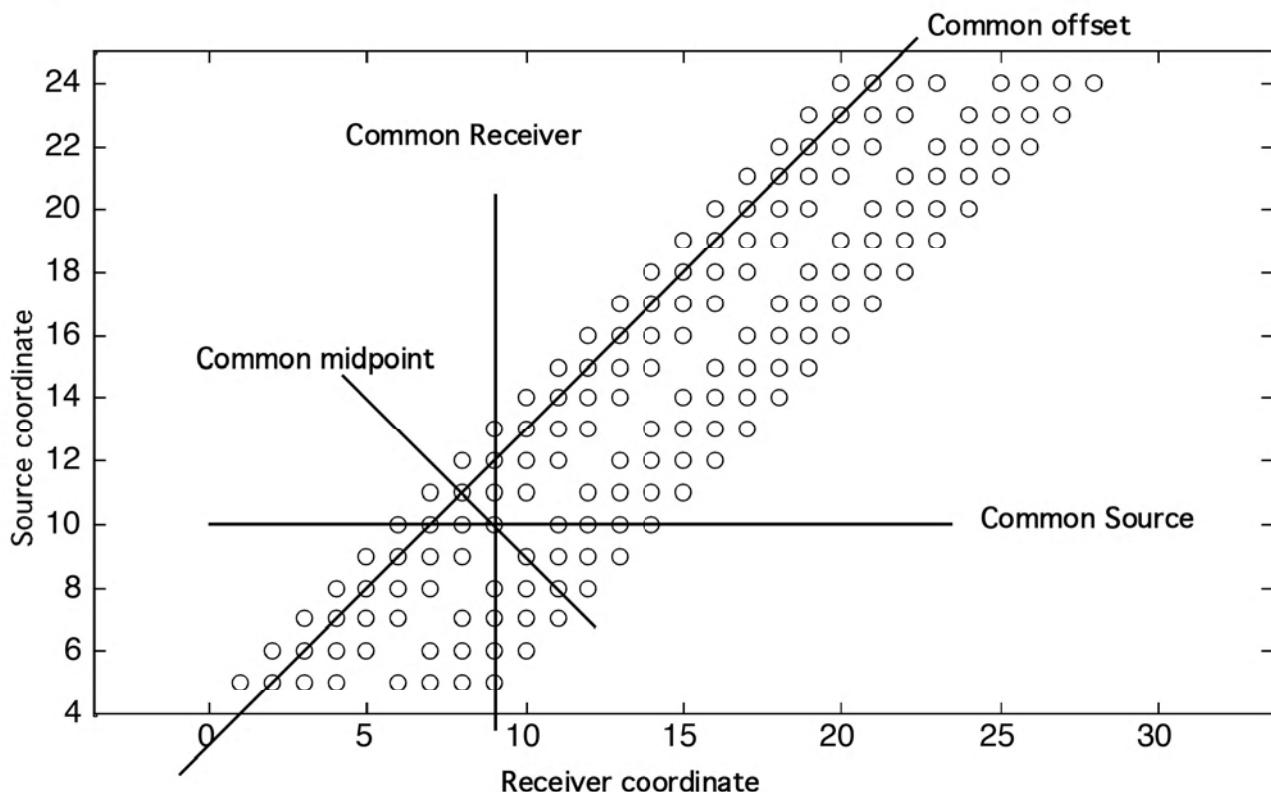
We have already discussed how the assumption of stationarity restricts the class of possible multiples to those whose generating reflector is above the deconvolution design zone. Also, multiples are only perfectly periodic on zero offset traces, so it might appear that we cannot handle any multiples with offset. However, in practice, nmo removal allows us to approximately handle offset traces.



A Surface Consistent Convolutional Model

The assumption of surface consistency is a very powerful one whose advantages over single channel deconvolution include:

- Far fewer operators to design. Consider 200 shots times 1000 receivers is 200,000 traces. If we do source-receiver deconvolution we use all 200,000 traces to design 1200 operators instead of 200,000 operators in the single channel case.
- Surface waves are less of a problem since they can be averaged out or simply left out of the design gate.
- The near surface effects fit the stationary deconvolution model better.
- Non-stationary effects are better modeled as functions of midpoint and offset. To first order, they can be postponed until after stack.



Surface Consistent Methods

Many seismic algorithms are formulated in a manner which is said to be "surface consistent". Among the most successful of are surface consistent amplitude adjustments, deconvolution, residual statics, as well as refraction statics, elevation statics and velocity analysis. The essence of a surface consistent algorithm is that it attempts to decompose some effect, such as amplitude variation, into terms which are functions of a single coordinate such as s , r , x , and h . A good introduction to such method is:

Taner, M.T. and Koehler, F., 1981, Surface Consistent Corrections, Geophysics, vol 46, pp 17-22

The surface consistent model treats the trace $h(t)$, from source s and receiver r , as the convolution of four operators:

$$h(s,r,x,h,t) = a(s,t) \bullet b(r,t) \bullet c(x,t) \bullet d(h,t)$$

where:

- $a(s,t)$ represents source consistent effects such as the source waveform and near surface delays and attenuation beneath the source.
- $b(r,t)$ represents receiver consistent effects such as the recorder response and near surface delays and attenuation beneath the receiver.
- $c(x,t)$ represents midpoint consistent effects such as spatial variation of moveout, reflectivity, and attenuation.
- $d(h,t)$ represents offset consistent effects such as moveout and multiples.

Surface Consistent Methods

If we now recast the basic equation into the frequency domain we have:

$$H(s, r, x, h, \omega) = A(s, \omega)B(r, \omega)C(x, \omega)D(h, \omega)$$

Taking the logarithm of both sides gives a linear equation:

$$\ln(H(\omega)) = \ln(A(\omega)) + \ln(B(\omega)) + \ln(C(\omega)) + \ln(D(\omega))$$

where the notation has been simplified to suppress the dependence of A, B, C, and D on s, r, x, and h respectively.

If we note that the real part of $\ln(H(\omega))$ is the log amplitude spectrum of the trace and the imaginary part is the phase spectrum, we see that we can model either amplitude or phase effects (or both) with this equation.

As an example, suppose we average the log amplitude spectrum over all frequencies to obtain a bulk power measure of the trace. Then we can model it as:

$$H_{\text{mean}} = A_{\text{mean}} + B_{\text{mean}} + C_{\text{mean}} + D_{\text{mean}}$$

We can write one such equation for every seismic trace while the number of unknowns is the number of sources plus the number of receivers plus the number of midpoints plus the number of offsets. For modern high fold data, we will always have many more equations than unknowns and so can solve for a set of surface consistent amplitude scalars by least squares. Any unwanted variation can then be removed by division by the appropriate amplitude scalars.

Surface Consistent Methods

Similarly, we could formulate a surface consistent deconvolution by modelling each frequency of the log amplitude spectrum using the real part of:

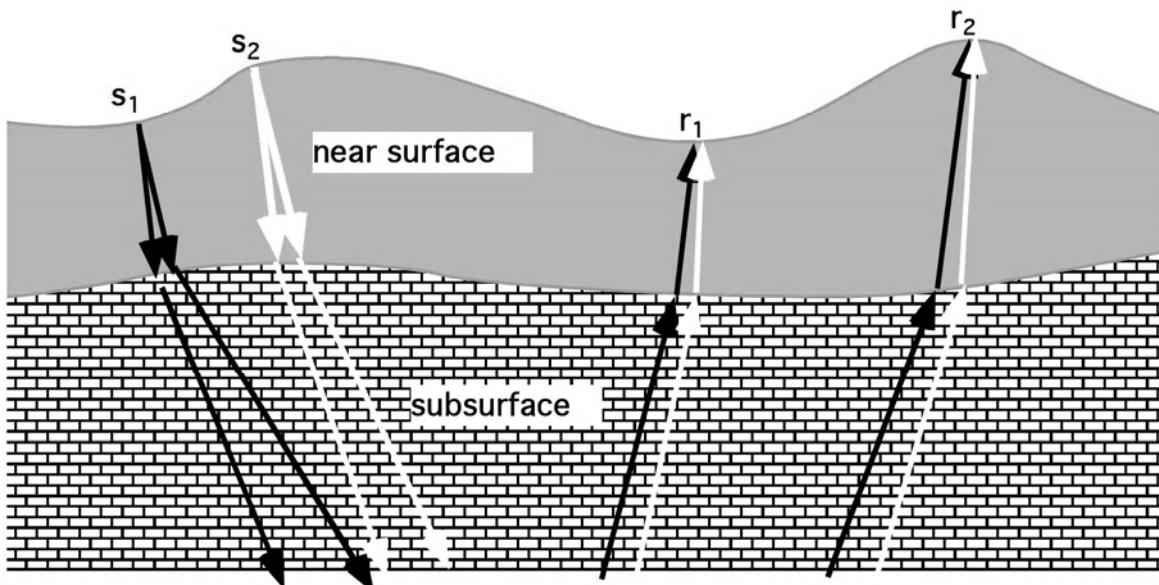
$$\ln(H(\omega)) = \ln(A(\omega)) + \ln(B(\omega)) + \ln(C(\omega)) + \ln(D(\omega))$$

This is considerably more work than the amplitude adjustment problem since a separate least squares problem must be solved for each frequency of interest. Nevertheless, it is a very powerful deconvolution technique which is widely used. Once we have solved for the surface consistent log amplitude spectra, it is a straightforward matter to compute minimum phase deconvolution operators for each source, receiver, midpoint, and offset as desired.

Surface consistent residual statics can also be solved in this fashion but that will be treated in depth later.

Statics and Datums

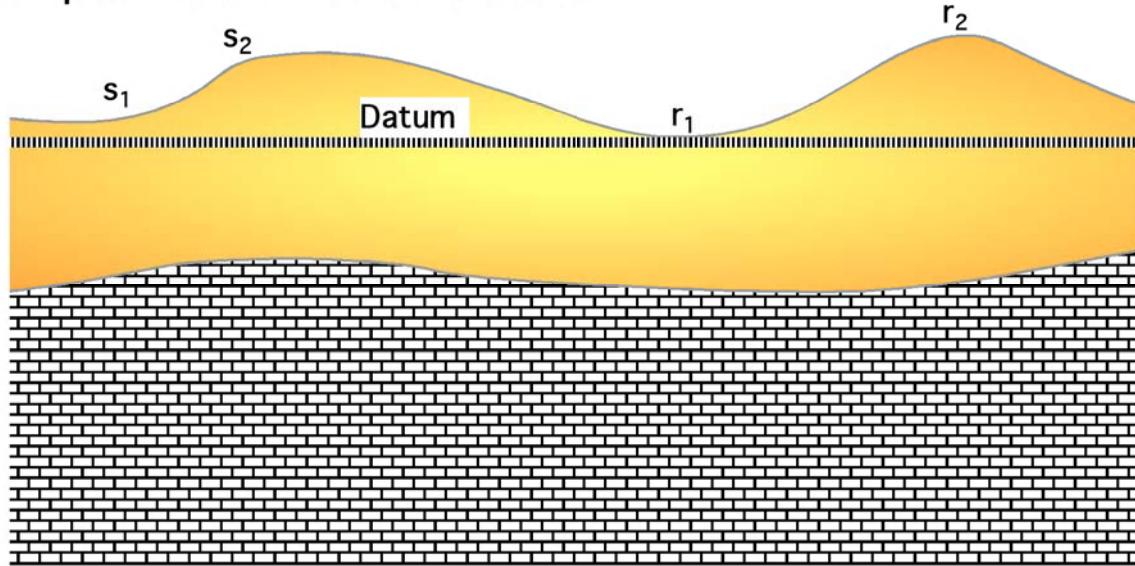
A "static correction" is a constant time shift (i.e. the magnitude of the shift does not vary up and down the trace). It is one of the simplest data processing steps yet it is absolutely essential, and very powerful, in the general event alignment process that precedes the cmp stack. Physically, a relative static shift exists between two traces whenever the near surface beneath the sources or the receivers is different.



This diagram depicts the nearly vertical raypaths that are common in the near surface (due to the strong refraction at its base) which lead directly to static delays. Four traces are shown being recorded: $t(s_1, r_1)$, $t(s_1, r_2)$, $t(s_2, r_1)$, $t(s_2, r_2)$. Comparing traces $t(s_1, r_1)$ and $t(s_1, r_2)$, we see that a static delay exists between the two due to differing conditions beneath the receivers. Similarly, $t(s_1, r_1)$ and $t(s_2, r_1)$ have a relative static delay due to differing conditions beneath the sources.

Statics and Datums

In order to quantify these concepts, we need to introduce the notion of a datum which is essentially the level of "compensation" for all traces:



Let us assume that a calculation has been run which provides us with estimates of the vertical traveltime beneath each source to the base of the near surface (BNS) and from beneath receiver to BNS together with the elevation of the BNS. Examples of such algorithms include refraction statics, elevations statics, and upshot statics (residual statics are a slightly different topic). Then we define the following:

δt_s ... source static or vertical traveltime from any source to BNS

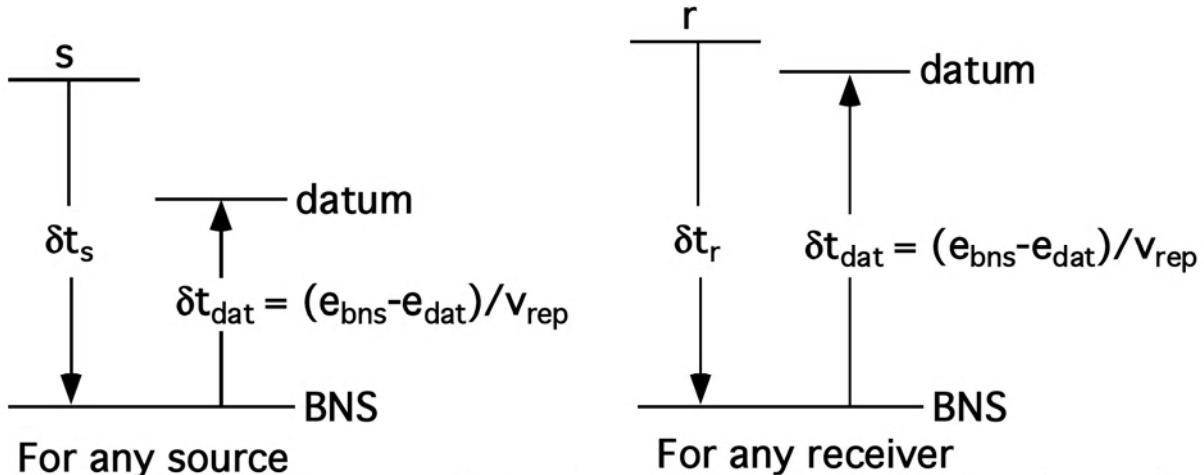
δt_r ... receiver static or vertical traveltime from any receiver to BNS

e_{bns} ... elevation of the base of the near surface at any position on the line

e_{dat} ... elevation of the datum at any position on the line

v_{rep} ... replacement velocity at any position on the line

Statics and Datums



For any trace, the total static to datum is defined as the source/receiver static plus the time from BNS to datum. Note that if the datum is above the BNS, then the datum static will be opposite in sign to the source/receiver static.

$$\delta t_{s \rightarrow dat} = \underline{\delta t_s} + \frac{e_{bns}(s) - e_{dat}(x)}{v_{rep}(x)} \quad \text{shift the source to datum}$$

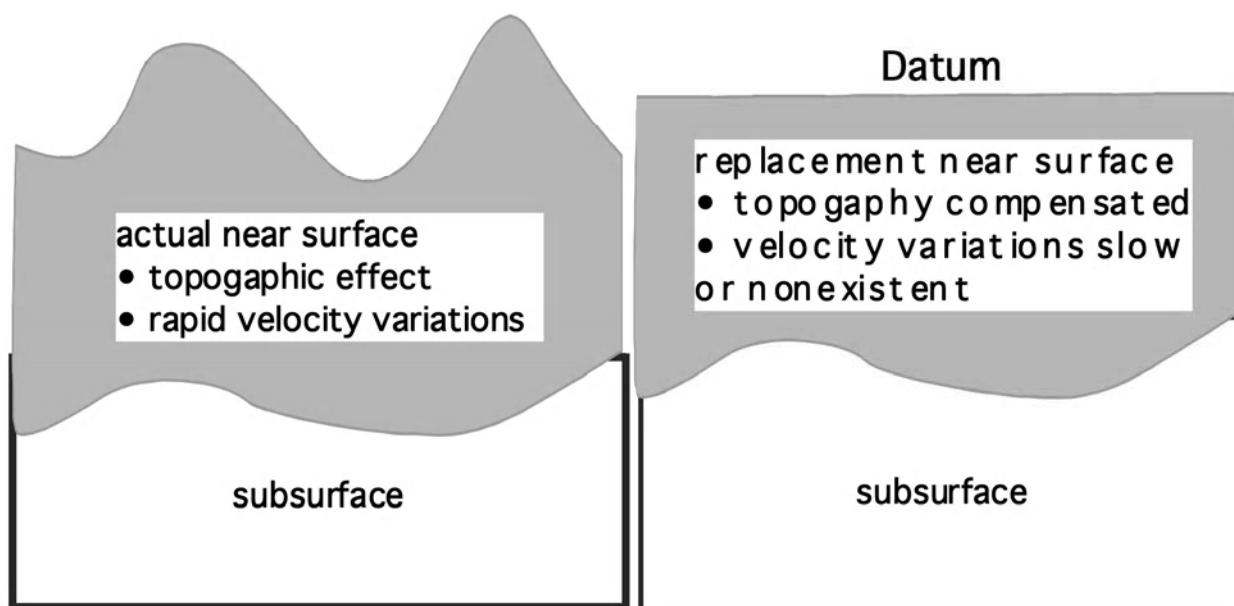
$$\delta t_{r \rightarrow dat} = \underline{\delta t_r} + \frac{e_{bns}(r) - e_{dat}(x)}{v_{rep}(x)} \quad \text{shift the receiver to datum}$$

$$\begin{aligned} \delta t_{trace \rightarrow dat} &= \delta t_{s \rightarrow dat} + \delta t_{r \rightarrow dat} \\ &= \underline{\delta t_s} + \underline{\delta t_r} + \frac{e_{bns}(s) + e_{bns}(r) - 2e_{dat}(x)}{v_{rep}(x)} \end{aligned}$$

In these expressions the datum elevation and the replacement velocity are evaluated at the source/receiver midpoint for the trace while the other terms are evaluated at the appropriate source/receiver coordinates.

Statics and Datums

Conceptually, the statics process can be regarded as stripping off the effects of a variable, topographic near surface and replacing them with a smooth (or perhaps constant) replacement near surface:



The statics model presented here removes most near surface effects though it does not compensate for any structure on the BNS. This can be easily done with a slight increase in algorithmic complexity. Otherwise, to the extent that propagation through the near surface can be modeled with vertical raypaths, this is an accurate replacement process. In the general case, many forms of seismic energy do not travel through the near surface on vertical paths and so are mishandled by this process.

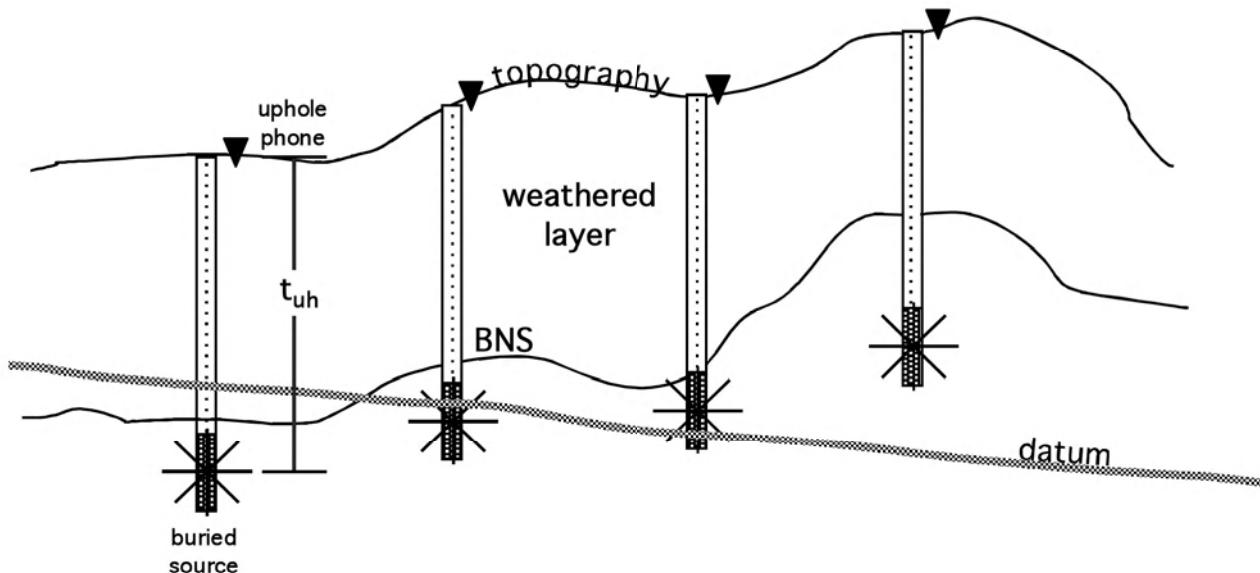
Statics and Datums

The static adjustment process is an approximate form of the general process of "downward continuation" or "wave equation datuming" that is only correct for vertical raypaths. Methods that can correctly handle slanting raypaths are usually formulated with wave equation techniques though they do not respond well to the rapid lateral variations in the near surface. Additionally the statics process does not need a detailed near surface model, only the traveltime delays, while wave equation methods need the more detailed model. Thus statics methods remain an essential part of a data processing flow but we should strive to keep the overall (bulk) adjustment small. That is, we do not want a large bulk time shift to result from the statics solution. Rather, we would prefer to see statics which are roughly zero mean over some length scale. Most commonly, an inappropriate choice of datum or replacement velocity can cause an unwanted bulk shift. Conceptually, the datum should be a smoothed version of the topography while the replacement velocity should be some average through the near surface. In practice, it is common to remove the mean from a statics solution and save it to be applied as the final step of processing. Also, it is common to select a processing (or nmo) datum which meets these criteria and then shift to final (in interpretation datum) at the end of processing.

Large bulk statics will generally degrade the processes of normal moveout removal and migration. Avoid such shifts in processing.

Statics with Uphole Times

It is often the case that buried impulsive sources are used and "uphole" times are measured with a receiver placed near the shothole. In principle, this can be regarded as a determination of the weathering velocity beneath each shotpoint. It is customary to assume that the shot depth is below the weathered layer whose base is denoted BNS.



Making uphole measurements regularly along the line allows a simple method of static correction which is usually preferred over simple "elevation corrections".

Assuming each source is below BNS, we can compute the shot to datum static as: $\delta t_s \approx 0$

$$\delta t_{s \rightarrow \text{dat}} \quad \delta t_s = \frac{e_s(s) - d_s(s) - e_{\text{dat}}(x)}{v_{\text{rep}}(x)}$$

$e_s(s)$ = topographic elevation at shot $d_s(s)$ = depth of shot

$e_{\text{dat}}(x)$ = elevation of datum at s-r midpoint

$v_{\text{rep}}(x)$ = replacement velocity at s-r midpoint

Statics with Uphole Times

Thus the shot to datum static is generally a function of (s, x) but if the datum and replacement velocity are constant, then it is purely a function of s .

The receiver to datum static is similar except that we include an uphole time. If the measured uphole times and shot depths are regarded as sampled versions of continuous functions, then values can be interpolated for every receiver. The receiver static is then:

$$\delta t_{r \rightarrow DAT} = t_{uh}(r) + \frac{e_s(r) - d_s(r) - e_{dat}(x)}{v_{rep}(x)}$$

The total static applied to any trace is thus:

$$\delta t_{trace \rightarrow DAT} = \delta t_{s \rightarrow DAT} + \delta t_{r \rightarrow DAT} = \frac{e_s(s) + e_s(r) - d_s(s) - d_s(r) - 2e_{dat}(x)}{v_{rep}(x)} + t_{uh}(r)$$

Since this is the total time from trace to datum, it must be subtracted from measured arrival times to apply the static correction.

In principle, this is a very good method but in practice several things usually go wrong including:

- Shots occur in the weathered layer
- Up-hole times are picked inaccurately
- Up-hole phones are blown out of the ground
- Shots are too sparse for accurate interpolation

Surface Consistent Residual Statics

The first, and arguably most important, application of surface consistent linear analysis was in the computation of residual statics. These developments occurred in the mid to late 1970's and the result was a quantum leap in the quality of stacked seismic images. (Wiggins R.A., Larner K.L., and Wisecup R.D., 1976, Residual statics analysis as a general linear inverse problem: Geophysics, 41, 922-938)

The fundamental notion is to use cross correlations to measure relative time shifts between traces and to model these time shifts as a linear inverse problem. If T_{ij} is the total traveltime of the trace from the i th shot into the j th receiver, then Wiggins et al. wrote:

$$T_{ij} = S_i + R_j + G_k + M_k X_{ij}^2 \quad (1)$$

$$T_{ij} = T_{\text{pilot}} + \delta T_{ij}$$

T_{ij} is the total traveltime of trace ij

S_i is the source static

R_j is the receiver static

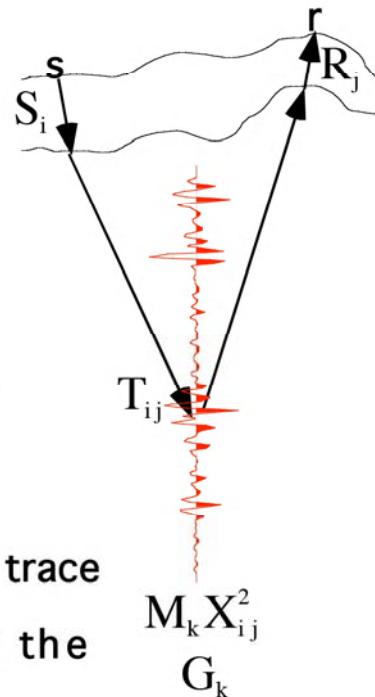
$M_k X_{ij}^2$ is an approximate residual NMO term
(see page 7.2)

G_k is a structure term

T_{pilot} is the traveltime of a reference "pilot" trace

δT_{ij} is computed from cross correlating the
 ij th trace with the pilot

i is the shot position index, j is the receiver position index, and $k=i+j-1$ is the cmp position index.



Surface Consistent Residual Statics

To make things concrete, Wiggins et al. present the formalism for a simple line with 4 live receivers per shot and a rollup of 1 station.

●	☆	☆	☆	☆
●	☆	☆	☆	☆
●	☆	☆	☆	☆
●	☆	☆	☆	☆
<hr/>				
•	•	•	S ₄	S ₃
•	•	•	R ₇	R ₆
•	•	•	G ₁₀	G ₇
•	•	•	M ₁₀	M ₇
•	•	•	S ₂	R ₅
•	•	•	R ₄	G ₅
•	•	•	R ₃	G ₃
•	•	•	R ₂	G ₁
•	•	•	R ₁	M ₃
•	•	•	M ₁	M ₁

For a particular survey geometry, the equations (1) generate one equation per trace (i.e. source-receiver combination). The unknowns are the source, receiver, structure, and rnmo terms. These can be written as a matrix expression:

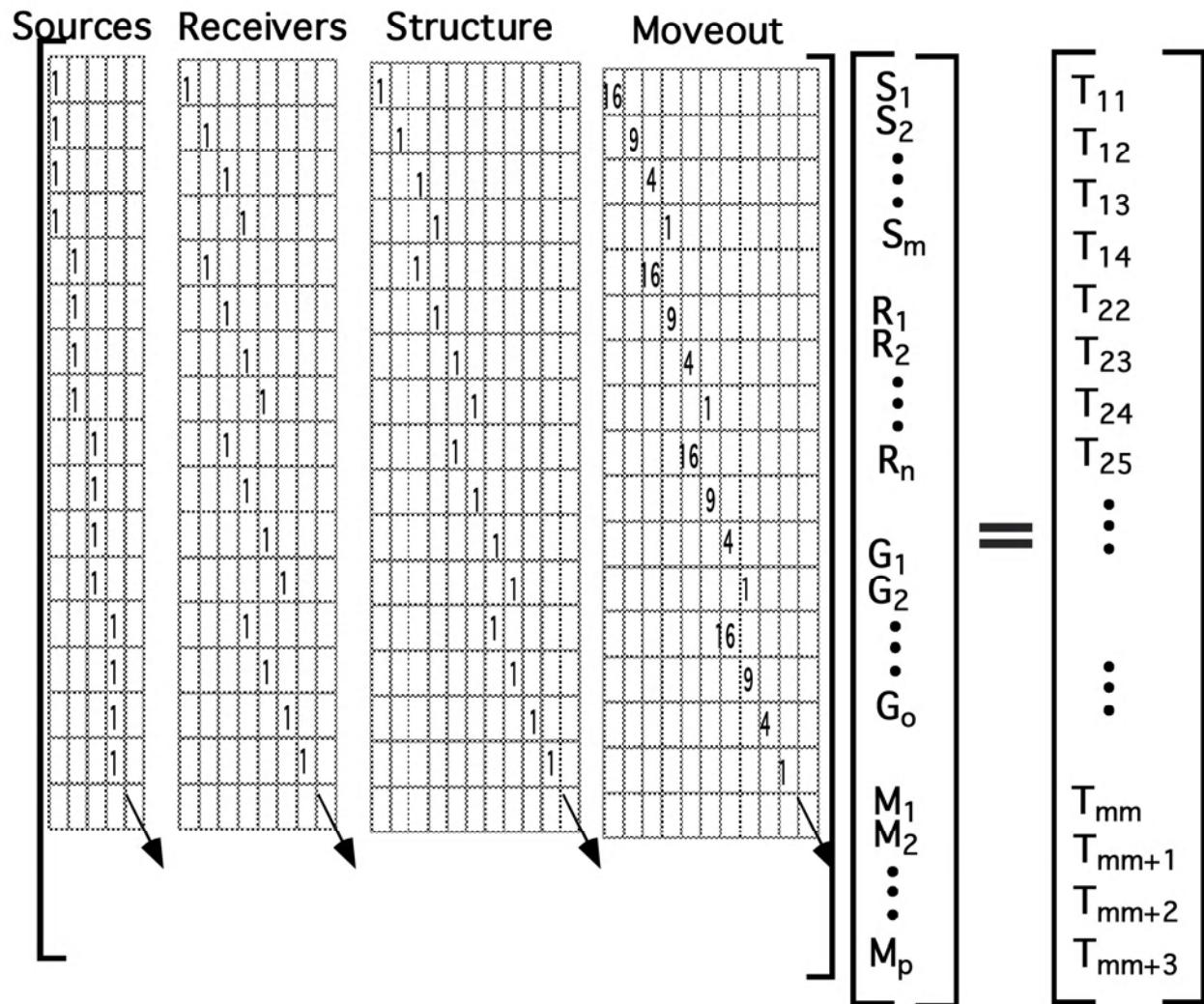
$$\mathbf{A}\bar{\mathbf{p}} = \bar{\mathbf{t}} \quad (2)$$

Generally, there are many more equations than unknowns in (2) and the solution can be formally obtained via least squares:

$$\bar{\mathbf{p}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \bar{\mathbf{t}} \quad (3)$$

Surface Consistent Residual Statics

The structure of the residual statics linear system:



Surface Consistent Residual Statics

A fundamental result of the Wiggins et al analysis is that the ability of the linear system to resolve statics trends is a function of the spatial wavelength of the trends. A statics trend which is significantly larger than a spread length is not reliably resolved. However, unless constraints are built in, the linear system will contain such trends in its solution and these will likely be erroneous. For this reason, residual statics solutions are often forced to zero mean and the long wavelength trends are obtained by other means (or not at all).

The linear system is usually very large but sparse and is generally solved by iterative techniques. Often (but not always) this implies an order to the solution (e.g. S, then R, then G, then M) and the results are nonunique and order dependent. The Gauss-Seidel numerical formalism is most often used. (See Yilmaz p 203 for a discussion.)

Surface Consistent Residual Statics

Other considerations:

- NMO removal. The relationship between statics and velocity analysis is a classic "chicken and egg" scenario. Velocity analysis requires statics corrections first and residual statics function best with minimal residual nmo (rnmo). As a result, data is often improved with iterations of velocity analysis and residual statics. Note that the linear model assumes rnmo is constant with time.
- Simple structure. It is quite common for residual statics algorithms to incorporate constraints which force the structure solution to be a simple one. If this is the case then the correlation window should be chosen appropriately.
- Correlation window. The right hand side of the linear equation set is obtained from trace to trace cross correlations. The single most important decision is to choose the correlation window. Typically, this will be a large (.5 to 1 second) window which may vary spatially along the line. Keep the following in mind when choosing the window:
 - Structure and rnmo should not be strongly time variant over the window. Avoid shallow windows or those with several distinct structural trends. Windows following a package of basement reflectors are often a good choice.
 - Avoid extremely narrow windows or low frequency "ringy" data. These can lead to cross correlation cycle skipping (i.e. picking the wrong peak on the correlation function).

Surface Consistent Residual Statics

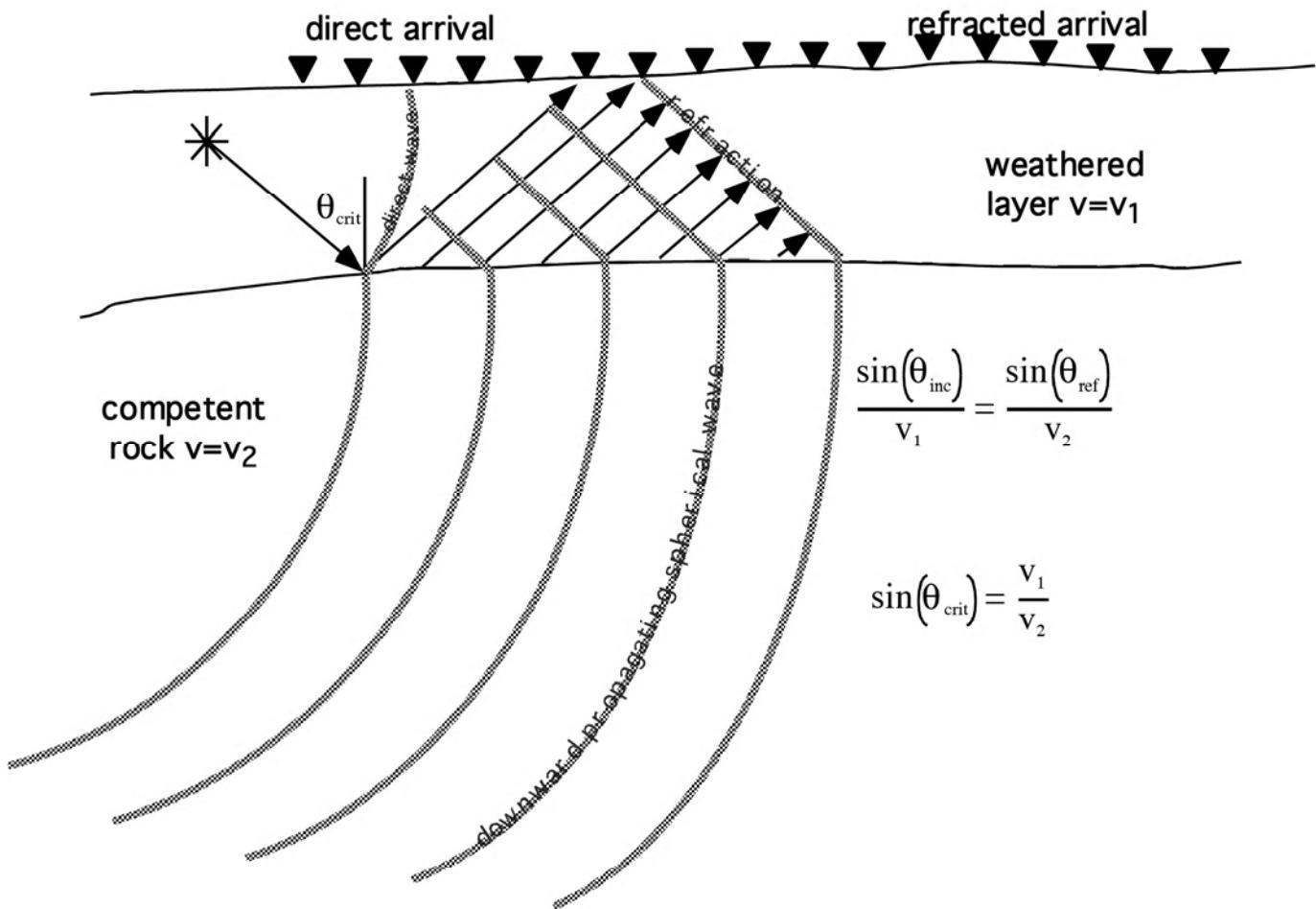
Still other considerations:

- Correlation length. This refers to the maximum number of lags which will be computed on the trace to trace cross correlations. It must be large enough to span the maximum expected time shift due to the sum of all four terms in equation (1). A common mistake is to choose a correlation window based on the expected size of statics without considering structure and rnmo. This will certainly lead to a poor solution. On the other hand, excessively large windows are expensive and are more likely to produce cycle skipping.
- Pilot traces. Rather than correlate each trace against all others (an expensive proposition), many implementations employ *pilot traces* to which traces are correlated. Pilot traces are essentially the best solution yet obtained. They may be formed within the algorithm by stacking traces in cmp gathers and possibly compositing adjacent gathers. Alternately, *external pilot traces* are sometimes provided by stacking, and signal enhancing, the best current solution. Improving the pilot traces in this manner can strongly improve (and bias) a statics solution.
- Data preparation. It is common practice to prepare data for residual statics using conservative bandpass filters, f-k filters, agc's, and other signal enhancing processes and then to apply those statics to data without these processes applied.
- Statics application. Since residual statics are computed from moveout corrected data, they are most properly applied after nmor. However, sometimes they are applied before nmor to improve nmo analysis or for prestack migration.

Refraction Statics

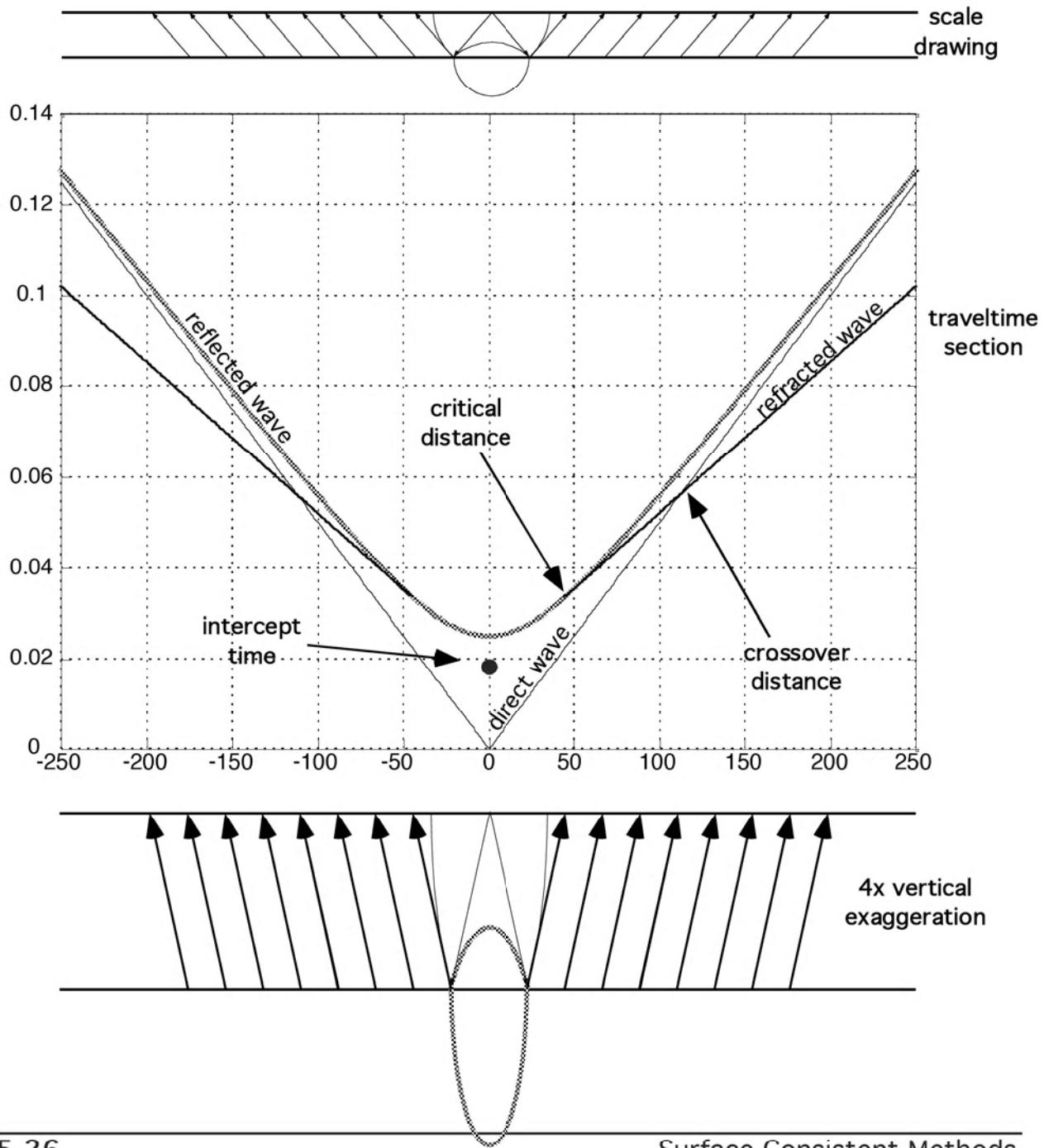
Seismic exploration with refractions has long been a successful technique for crustal studies where very long offset surveys (20 to 100km or more) are used to capture refraction events. These methods contrast with reflection seismology which uses dominantly vertically traveling waves and records their reflections at angles less than the critical angle.

In recent years there has been renewed interest in using the first breaks of seismic reflections data to analyze the very near surface and compute static corrections.



Refraction Statics

Let us simplify the problem and consider a horizontal recording surface and refractor. Let $v_1=2000$ m/s, $v_2=3000$ m/s, and the refractor depth $z_1=25$ m. Then, for a surface source, 3 waves (direct, reflected, and refracted) are generated.



Refraction Statics

A good reference for the equations describing these traveltimes curves is Slotnick (1959, Lessons in Seismic Computing, SEG). In lesson 18, it is shown that:

$$t_{\text{reflection}} = \sqrt{\frac{x^2}{v_1^2} + t_o^2} \quad t_o = \frac{z_1}{v_1}$$

And in lesson 19, Slotnick derives:

$$t_{\text{refraction}} = \begin{cases} \frac{x}{v_2} + t_i, & \text{if } x > x_c \\ \text{nonexistant otherwise} \end{cases}$$

$$t_i = \frac{2z_1}{v_1} \cos(\theta_{\text{crit}}) \quad x_c = z_1 \tan(\theta_{\text{crit}})$$

And obviously

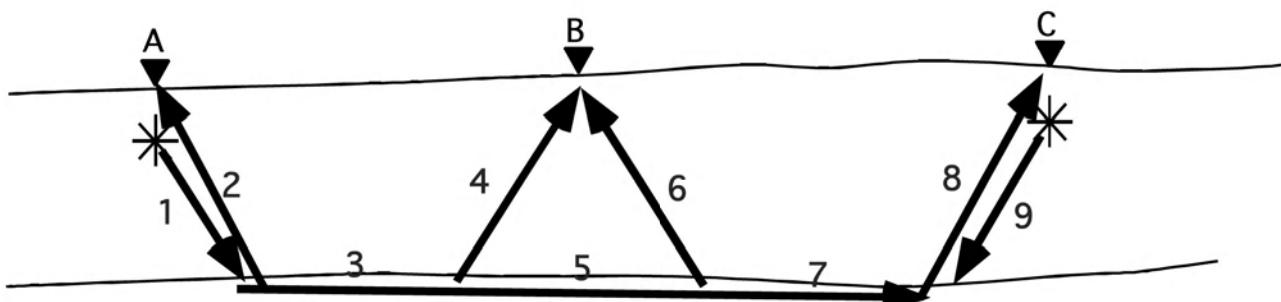
$$t_{\text{direct}} = \frac{x}{v_1}$$

It is interesting to note that the refraction traveltime curve emerges from that of the reflection at the point where their slopes are equal. Also, measurement of the intercept time, t_i , is key to determining the depth to the refractor.

Generally, acquisition spreads for reflection surveying are inadequate to accurately estimate the weathering velocity, v_1 . Therefore it is a critical input parameter in most programs.

Refraction Statics

The method just outlined is difficult to use in practice because the first breaks are not perfectly linear and can be difficult to interpret and because the depth to the weathered layer varies along the line. A variety of methods exist which are more flexible. One of the most simple is the ABC method which is described in Waters (1992, Reflection Seismology, p 197) and is shown below:



Using the numbering scheme in the diagram, the vertical traveltime from B to the refractor (assumed to be the base of the weathered layer) can be estimated in a simple fashion. First, letting t_{AB} denote the time from the surface at A to the surface at B via the refractor and similarly for the other stations, we can write:

$$t_{AB} = t_2 + t_3 + t_4 \quad t_{CB} = t_8 + t_7 + t_6 \quad t_{AC} = t_2 + t_3 + t_5 + t_7 + t_8$$

Then note that:

$$t_{AB} + t_{CB} - t_{AC} = t_4 + t_6 - t_5$$

The quantity, $.5(t_{AB} + t_{CB} - t_{AC})$, is often taken as an estimate of the traveltime through the near surface below B and is called the "delay time".

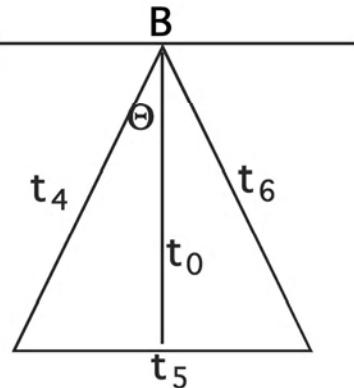
Refraction Statics

Close consideration of the geometry at right leads to:

$$t_0 = z_1/v_1$$

$$t_4 = t_6 = z_1/\cos(\Theta)$$

$$\sin(\Theta) = v_1/v_2$$



It then follows that

$$.5(t_{AB} + t_{CB} - t_{AC} = t_4 + t_6 - t_5) = z_1/(v_1 \cos(\Theta))(1 - (v_1/v_2)^2)$$

In the typical case $v_1 \ll v_2$ and the delay time becomes a good approximation to the vertical traveltime through the weathering, t_o .

If buried sources are to be modeled as in the diagram on the previous page, simple modifications like $t_{AB} = t_1 + t_{uhA} + t_3 + t_4$ will usually suffice.

Methods of Seismic Data Processing

Lecture Notes Geophysics 557

Chapter 6 Velocity Definitions and Simple Raytracing

Velocity in Theory and Practice

It is very common in seismic data processing to make measurements that have the dimensions of velocity (i.e. distance/time). Unfortunately, it is also very common that the people involved do not understand what linkages may exist between their measurements and the actual speed of wave propagation in the earth. This confusion can lead to mistakes in data processing and interpretation.

A great deal of research has been done on this topic and, in order to appreciate it, it is essential to have a basic understanding of various velocity "measures". A velocity measure refers to any measurement, with the dimensions of velocity, which is related in some way to the speed of wave propagation. To understand how it is related is the key to interpreting the significance of the velocity measure.

We have already encountered some velocity measures such as the various apparent velocities. In this chapter, we will examine others such as the average, mean, and rms velocities and show how they are related to actual local wavespeeds. We will also see that velocity measures can often be global averages or averages defined over some interval. Whatever they are, relating them to local wavespeeds is a step towards estimating the physical properties of the subsurface, and that is the ultimate goal of seismic exploration.

Instantaneous Velocity

v_{ins} or just v

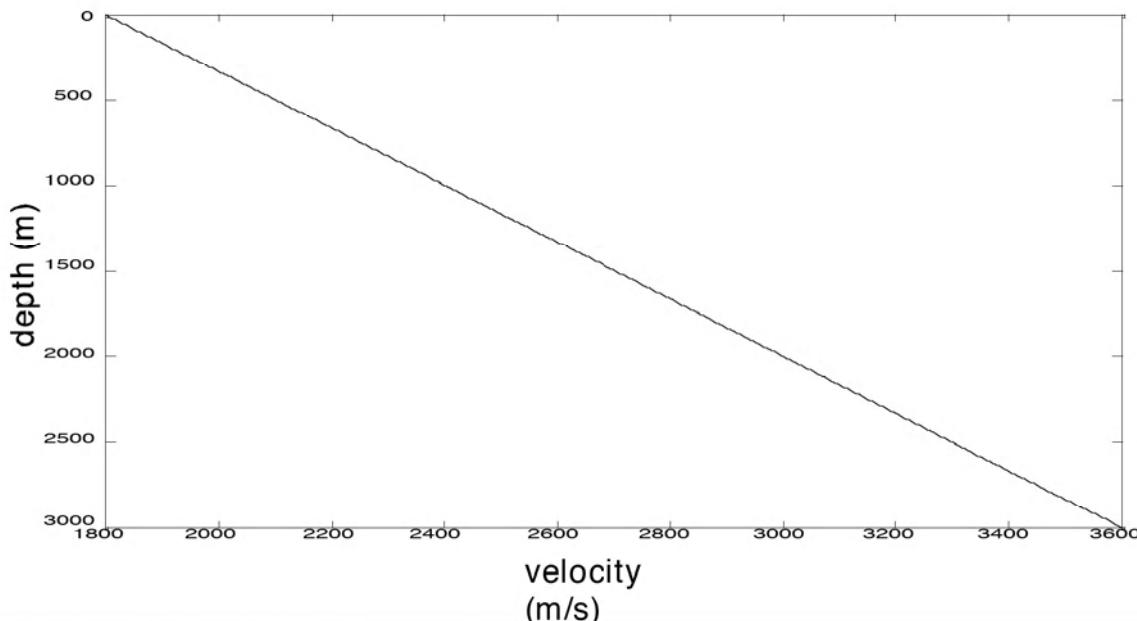
This term generally refers to the speed of propagation of seismic waves in the earth. It can refer to P-waves, S-waves, various types of surface waves, or some combination of these, which should be clear from context. If not explicitly stated, assume P-waves.

Note that, like most seismic "velocities", v_{ins} is not usually a vector quantity and so is not a velocity as physicists would use the term. Rather it is a scalar which can be thought of as the magnitude of a velocity vector.

For seismic experiments, v_{ins} is generally a function of position (x, y, z) in the earth but does not change with time. It can rarely be assumed spatially constant but often significant variations occur only in the vertical direction, as is common in sedimentary basins.

Example: The "universal function":

$$v = v_0 + cz, \text{ where } v_0 = 1800 \text{ m/sec and } c = .6/\text{sec.}$$



Vertical Traveltime

τ or t

When velocity (instantaneous) is a general function of position, $v=v(x,y,z)$, we can still restrict our attention to a single (x,y) location and there consider $v(z)$. Vertical traveltimes, τ , at any such location turns out to be a useful substitute for z , even though Snell's law may prevent energy from actually propagating along a vertical path, because there is always a one-to-one correspondence between τ and z . Vertical traveltimes is calculated by considering a small interval dz over which $v(z)$ is essentially constant and expressing $d\tau$ as:

$$d\tau = \frac{dz}{v(z)}$$

The total one-way traveltime from the surface to depth z , is simple the sum, or integral, of many such small contributions:

$$\tau(z) = \int_0^z \frac{d\tilde{z}}{v(\tilde{z})}$$

Defined in this way, $\tau(z)$ is a mathematically function which always increases with increasing z . Such monotonically increasing functions always have inverses. That is, given $\tau(z)$ we can always find $z(\tau)$ and vice versa.

Vertical Traveltime

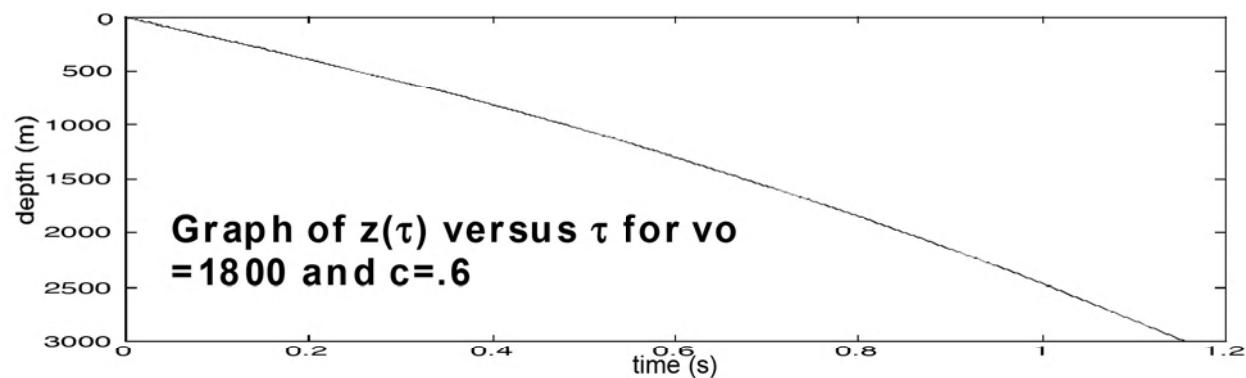
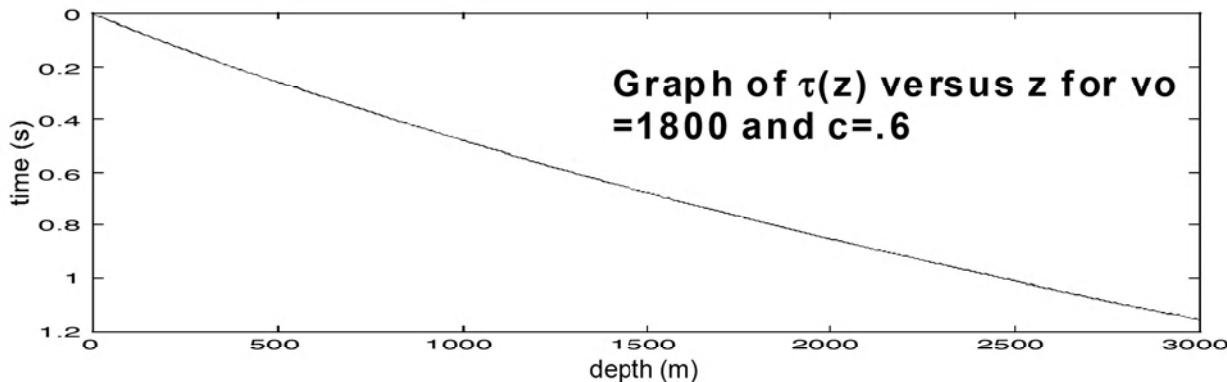
Consider again the example of linear variation of v with z .

$$\tau(z) = \int_0^z \frac{dz}{v(z)} = \frac{1}{c} \int_{v_o}^{v_o + cz} \frac{d\xi}{\xi} = \frac{1}{c} \ln \left(1 + \frac{cz}{v_o} \right)$$

Then, we can find $z(\tau)$ by solving this result for z :

$$z(\tau) = \frac{v_o}{c} \left(e^{c\tau} - 1 \right)$$

These relationships between time and depth are called "time-depth curves" and provide a simple method of vertical time↔depth conversion.



v_{ins} as a function of vertical traveltime

$$v_{ins}(\tau)$$

For any fixed (x,y) location, every depth, z , has a unique vertical traveltime value associated with it. Therefore, v_{ins} , which is physically a function of z may always be expressed as a function of $\tau(z)$.

Example: $v(z) = v_o + cz$

Then we have shown that: $\tau(z) = \frac{1}{c} \ln \left(1 + \frac{cz}{v_o} \right)$

and:

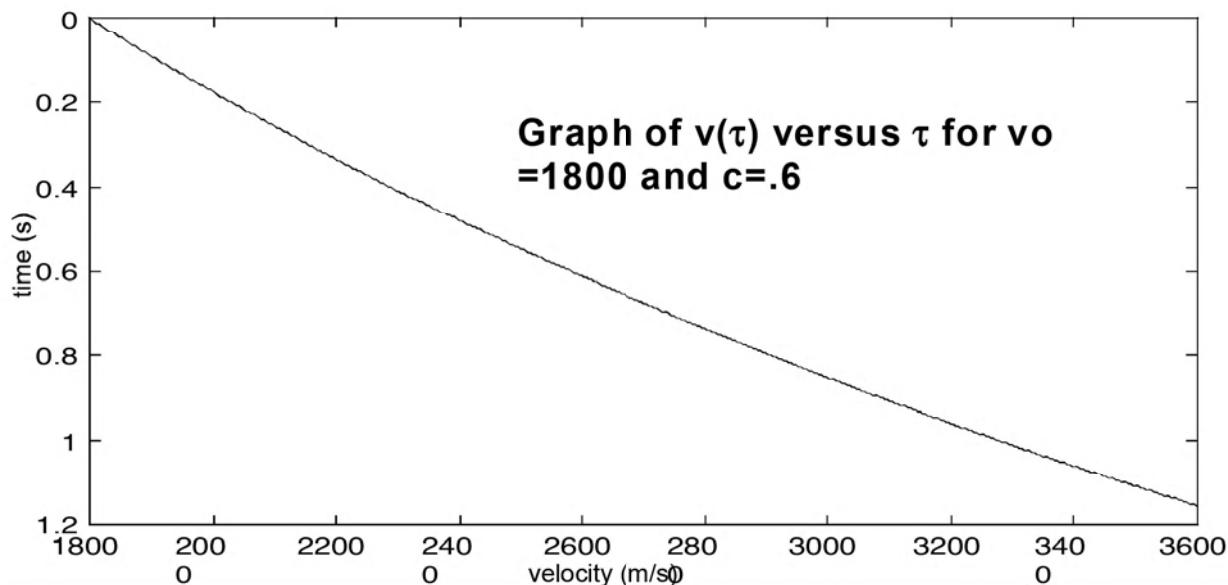
$$z(\tau) = \frac{v_o}{c} (e^{c\tau} - 1)$$

so:

$$v(\tau) = v(z(\tau)) = v_o + c \left[\frac{v_o}{c} (e^{c\tau} - 1) \right]$$

$$v(\tau) = v_o e^{c\tau}$$

A linear variation of velocity with depth is exponential with time.



v_{ins} as a function of vertical travelttime

We often obtain a good approximation to $v_{ins}(\tau)$ from seismic velocity analysis. One use of $v_{ins}(\tau)$ is in calculating z given τ .

$$dz = v_{ins}(\tau) d\tau$$

$$z(\tau) = \int_0^\tau v_{ins}(\tilde{\tau}) d\tilde{\tau}$$

Compare this with the analogous expression for computing vertical travelttime given $v_{ins}(z)$:

$$\tau(z) = \int_0^z \frac{dz}{v_{ins}(\tilde{z})}$$

Average Velocity

$$v_{ave}$$

The typical industry definition of the average velocity, v_{ave} , is:

For any depth, z , the average velocity is given by the depth, z , divided by the vertical traveltimes, $\tau(z)$. Thus:

$$v_{ave}(z) = \frac{z}{\tau(z)} \quad \text{and} \quad v_{ave}(\tau) = \frac{z(\tau)}{\tau}$$

Thus, expressed as a function of z , v_{ave} becomes:

$$v_{ave}(z) = \frac{z}{\int_0^z \frac{dz}{v(\tilde{z})}}$$

while, as a function of time, a much more pleasant expression results

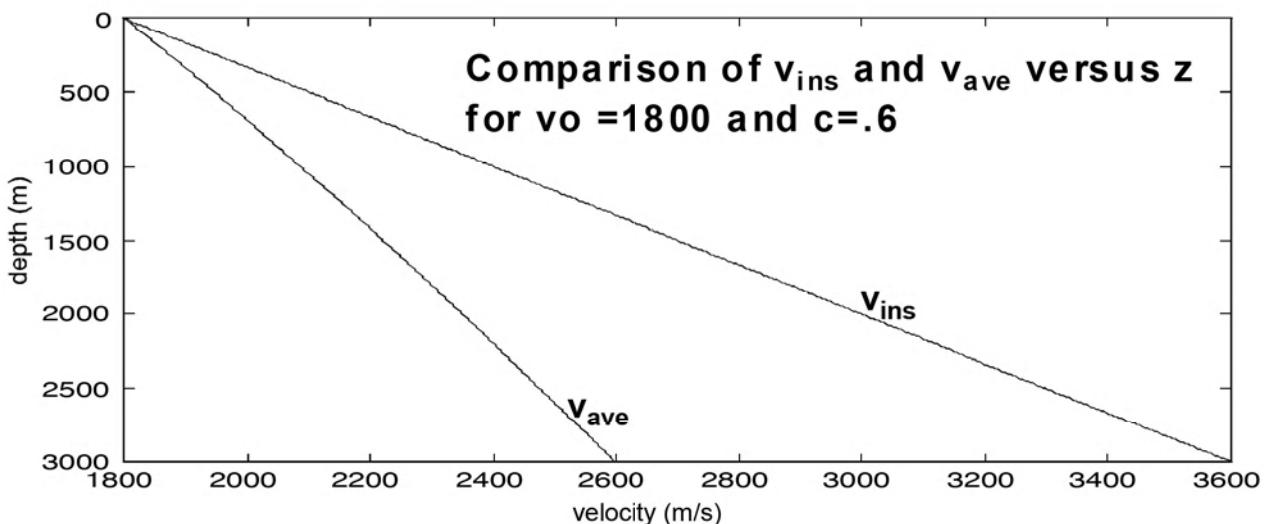
$$v_{ave}(\tau) = \frac{1}{\tau} \int_0^\tau v(\tilde{\tau}) d\tilde{\tau}$$

Thus we see that v_{ave} represents a mathematical average of the instantaneous velocity over time and not depth.

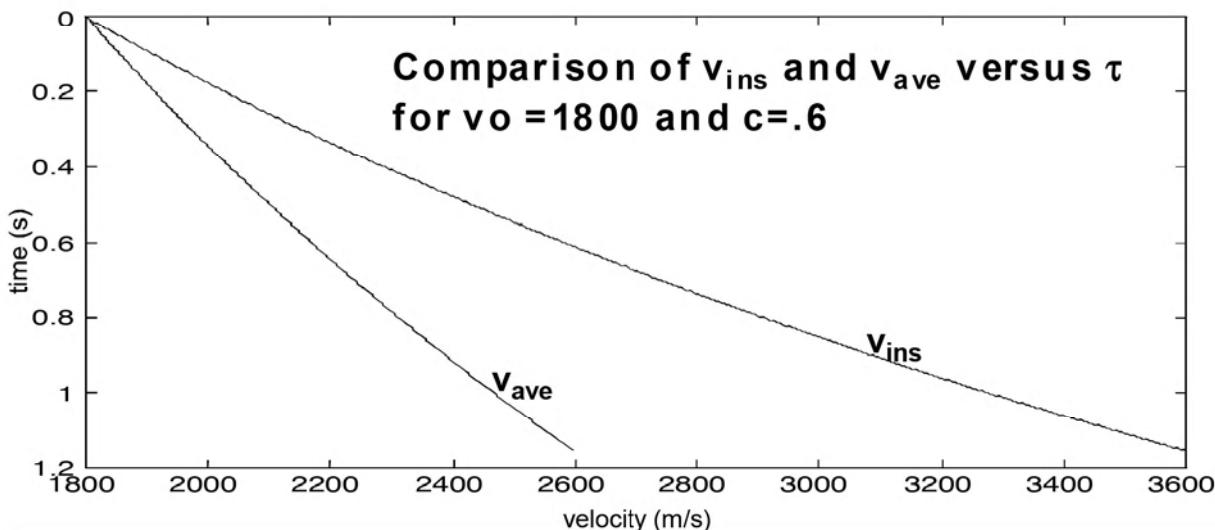
Average Velocity

Continuing with our example of linear variation of velocity with depth:

$$v_{ave}(z) = \frac{z}{\int_0^z \frac{dz}{v(\tilde{z})}} = \frac{cz}{\ln \left(1 + \frac{cz}{v_o} \right)}$$



and $v_{ave}(\tau) = \frac{1}{\tau} \int_0^\tau v(\tilde{\tau}) d\tilde{\tau} = \frac{v_o}{C\tau} (e^{c\tau} - 1)$



Mean Velocity

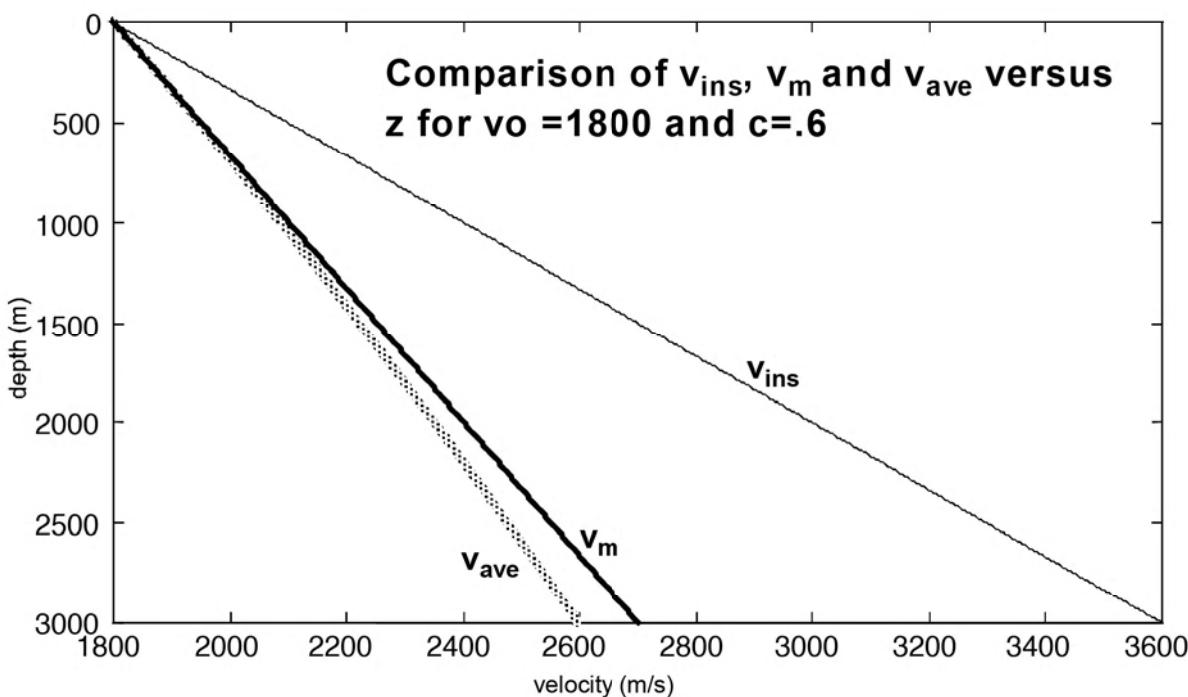
$$v_m$$

We have seen that the "average velocity", v_{ave} , used in seismology is actually a mathematical average taken over vertical traveltimes. Similarly, we can speak of the mathematical average over depth, called the mean velocity, which is defined by:

$$v_m(z) = \frac{1}{z} \int_0^z v(\tilde{z}) d\tilde{z}$$

For the linear variation of v with z , we have:

$$v_m(z) = \frac{1}{z} \int_0^z v(\tilde{z}) d\tilde{z} = \frac{1}{z} \int_0^z (v_o + c\tilde{z}) d\tilde{z} = v_o + \frac{1}{2}cz$$
$$v_m(\tau) = v_m(z(\tau)) = v_o + \frac{1}{2}c \left(\frac{v_o}{c} (e^{c\tau} - 1) \right) = \frac{v_o}{2} (1 + e^{c\tau})$$



RMS Velocity

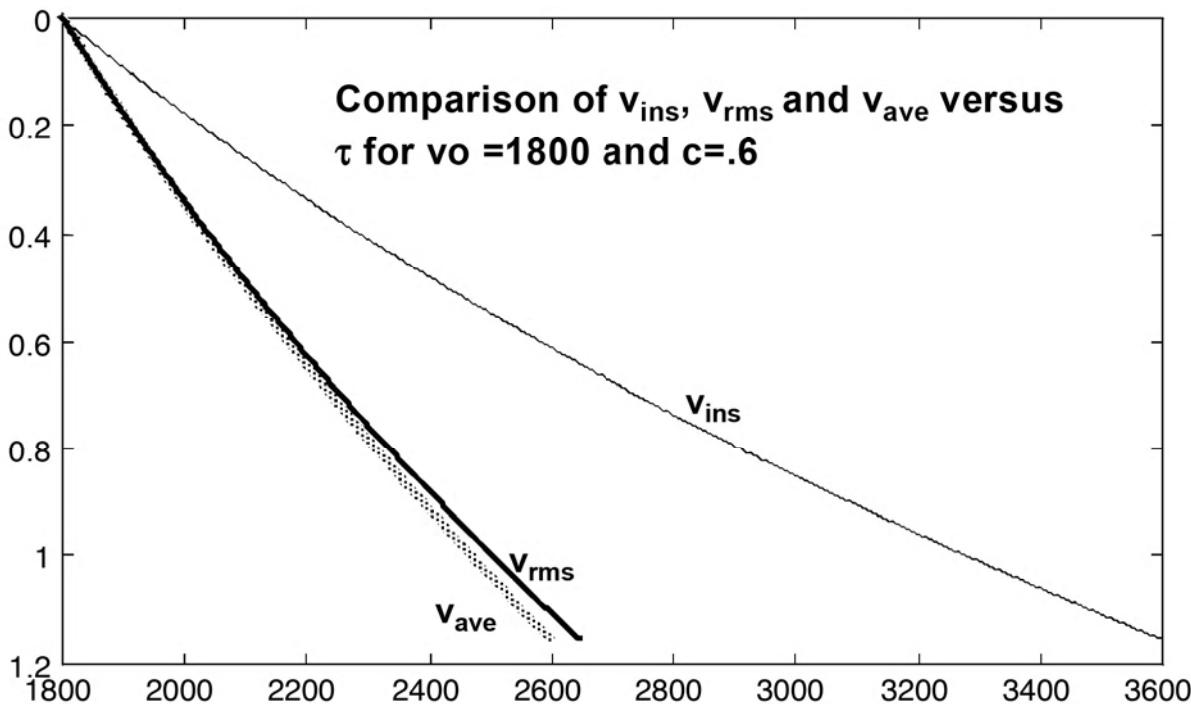
$$v_{rms}$$

Another useful velocity measure is v_{rms} where the "rms" stands for root mean square. Similarly to v_{ave} , v_{rms} is calculated by integrating with respect to vertical traveltime but this time the integrand is v^2 .

$$v_{rms}^2(\tau) = \frac{1}{\tau} \int_0^\tau v^2(\tilde{\tau}) d\tilde{\tau}$$

For the linear with depth example, we have:

$$v(\tau) = v_o e^{c\tau}$$
$$v_{rms}^2(\tau) = \frac{1}{\tau} \int_0^\tau v_o^2 e^{2c\tilde{\tau}} d\tilde{\tau} = \frac{v_o^2}{2c\tau} (e^{2c\tau} - 1)$$



RMS Velocity

A number of relationships exist between the various velocity averages: v_{ave} , v_m , and v_{rms} . Several of importance are:

$$v_{rms}^2 = v_{ave} v_m$$

$$v_{rms} \geq v_{ave}$$

The first is easily proven as follows:

$$\begin{aligned} v_{rms}^2 &= \frac{1}{\tau} \int v^2(\tau) d\tau = \frac{z}{\tau} \left(\frac{1}{z} \int v (vd\tau) \right) \\ &= \frac{z}{\tau} \left(\frac{1}{z} \int v dz \right) = v_{ave} v_m \end{aligned}$$

The second follows from a well known mathematical inequality known as Schwartz's Inequality. This quite general result confirms the intuition that anytime you average the squares of a series of numbers, the square root of that result is always greater than the direct average of the numbers themselves. This is quite obviously true when the numbers are randomly varying in both sign and magnitude but is also true for any other variation no matter how systematic. The equality holds only when the functions (i.e. the numbers) are constant.

Interval Velocity

$$v_{int}$$

Corresponding to any of the velocity averages, an "interval" velocity can be defined which is essentially the particular average applied across a small interval rather than from the surface. For example, the average and rms velocities across an interval defined by τ_1 and τ_2 are simply:

$$v_{aveint}(\tau_1, \tau_2) = \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} v(\tilde{\tau}) d\tilde{\tau}$$

$$v_{rmsint}(\tau_1, \tau_2) = \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} v^2(\tilde{\tau}) d\tilde{\tau}$$

Note that these quantities must be considered as functions of both the upper and lower bounds of the integrals. Also, in the limit as the interval shrinks so small that $v(t)$ can be considered constant, both interval velocities approach the instantaneous velocity.

It follows from the definition of average velocity, that the ratio of the depth interval to the time interval gives the average velocity across the interval:

$$v_{aveint}(\tau_1, \tau_2) = \frac{z_2 - z_1}{\tau_2 - \tau_1} = \frac{\Delta z}{\Delta \tau}$$

The average velocity across a large interval, for example from 0 to τ can be easily broken into a sum of interval velocities across subintervals.

Interval Velocity

The average velocity across a large interval, for example from 0 to τ can be easily broken into a sum of interval velocities across subintervals.

$$v_{\text{aveint}}(\tau_0, \tau_2) = \frac{1}{\tau_2 - \tau_0} \int_{\tau_0}^{\tau_2} v(\tilde{\tau}) d\tilde{\tau} = \frac{1}{\tau_2 - \tau_0} \left[\int_{\tau_0}^{\tau_1} v(\tilde{\tau}) d\tilde{\tau} + \int_{\tau_1}^{\tau_2} v(\tilde{\tau}) d\tilde{\tau} \right]$$
$$(\tau_2 - \tau_0) v_{\text{aveint}}(\tau_0, \tau_2) = (\tau_1 - \tau_0) v_{\text{aveint}}(\tau_0, \tau_1) + (\tau_2 - \tau_1) v_{\text{aveint}}(\tau_1, \tau_2)$$

This result is called the "addition formula" for average interval velocities. Obviously, this could be repeated for any number of subintervals with a similar result. In fact, if we let v_k be the average velocity across the k th subinterval and let $\Delta\tau_k$ be the traveltimes across it, then the average velocity across the total interval is:

$$v_{\text{aveint}}(\tau_0, \tau_2) = \frac{1}{\tau_2 - \tau_0} \sum_{k=1}^n v_k \Delta\tau_k$$

where: $\tau_2 - \tau_0 = \sum_{k=1}^n \Delta\tau_k$ is the total traveltimes across the interval

This is essentially a discrete layer equivalent to the integral definition of v_{ave} and is often used in practical situations.

Interval Velocity

The analysis just used to examine the average velocity across an interval would be repeated in similar fashion for either v_m or v_{rms} . The latter case is far more important due to the close link between data derived stacking velocities and v_{rms} . The addition formulae for v_{rms} is:

$$(\tau_2 - \tau_0) v_{rms,int}^2 (\tau_0, \tau_2) = (\tau_1 - \tau_0) v_{rms,int}^2 (\tau_0, \tau_1) + (\tau_2 - \tau_1) v_{rms,int}^2 (\tau_1, \tau_2)$$

Similarly, the generalization for n subintervals is:

$$v_{rms,int}^2 (\tau_0, \tau_2) = \frac{1}{\tau_2 - \tau_0} \sum_{k=1}^n v_k^2 \Delta \tau_k \quad \text{with} \quad \tau_2 - \tau_0 = \sum_{k=1}^n \Delta \tau_k$$

We must, of course, interpret v_k as the rms velocity across the kth subinterval.

In practice, stacking velocities are often assumed to closely approximate rms velocities. Thus, it often occurs that we know the stacking velocities to two closely spaced reflectors and wish to use this information to estimate the interval velocity between them. We can use the addition formula for this purpose if we take τ_0 be the datum from which the stacking velocities were measured, τ_1 to be the time to the first reflector, and τ_2 to be the time to the second. Then, the interval rms velocity between the reflectors is estimated by:

$$v_{rms,int}^2 (\tau_1, \tau_2) = \frac{(\tau_2 - \tau_0) v_{rms,int}^2 (\tau_0, \tau_2) - (\tau_1 - \tau_0) v_{rms,int}^2 (\tau_0, \tau_1)}{(\tau_2 - \tau_1)}$$

Interval Velocity

In this expression, we use the stacking velocity to the second reflector in place of $v_{rmsint}(\tau_0, \tau_2)$ while $v_{rmsint}(\tau_0, \tau_1)$ becomes the stacking velocity to the first reflector. For historical reasons, interval velocities so calculated are often referred to as Dix interval velocities after C.H. Dix who first derived the connection between stacking and rms velocities. There are several important properties of Dix interval velocities:

- They are rms interval velocities
- For very small intervals, they will approximate local instantaneous velocities.
- Proper calculation requires rms velocities to the top and bottom of the interval.
- They are only meaningful to the extent that stacking velocities are well approximated by rms velocities.
- Noisy or invalid v_{rms} estimates can lead to imaginary or otherwise unphysical interval velocities.

With respect to the last point, note that the Dix interval velocity equation yields the square of the interval velocity and hence must evaluate to a positive number. This requires that:

$$v_{rmsint}^2(\tau_0, \tau_2) > v_{rmsint}^2(\tau_0, \tau_1) \frac{\tau_1 - \tau_0}{\tau_2 - \tau_0}$$

or

$$v_{rmsint}(\tau_1 + \Delta\tau) > v_{rmsint}(\tau_1) \sqrt{\frac{\tau_1}{\tau_1 + \Delta\tau}} \quad \text{where} \quad \Delta\tau = \tau_2 - \tau_1$$

Interval Velocity

Thus the rate at which v_{rms} can decrease and still yield a physical result is constrained. Note that there is no constraint on the rate of increase.

Question: Suppose you know the following:

- $\tau_1, \tau_2, z_1, z_2 \dots$ depths and times to two reflectors
- v_{rms1} and $v_{rms2} \dots$ estimates of v_{rms} to both reflectors

Consider:

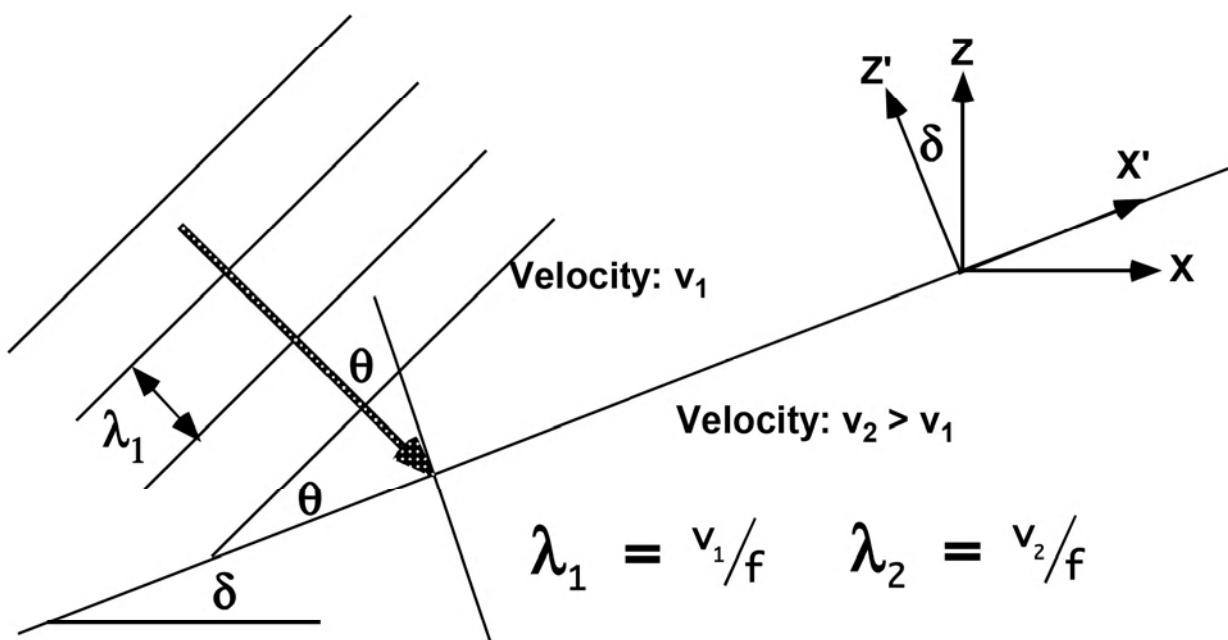
$$V_{int} = \frac{z_2 - z_1}{\tau_2 - \tau_1} \quad \text{and} \quad V_{int \text{ sint}} = \sqrt{\frac{v_{rms2}^2 \tau_2 - v_{rms1}^2 \tau_1}{\tau_2 - \tau_1}}$$

Do you expect these results to be similar?

If not, which will most likely be larger?

Snell's Law

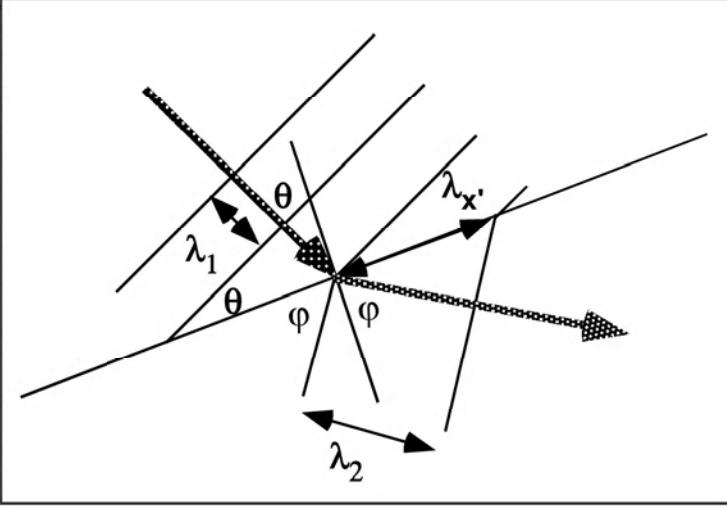
Snell's law is the relation which governs the angles of reflection and refraction of wavefronts (or equivalent raypaths) at velocity interfaces. To understand its origins, consider the diagram below which shows a periodic plane wave approaching a dipping interface:



Here we assume velocity $v_2 > v_1$ which means that the wavelength in the second medium will be greater than that in the first. If we draw a series of wavefronts, of greater wavelength in the second medium, we find that we must tilt them at an angle with respect to the incoming wavefronts in order to match the wavefronts at the boundary. This is shown on the next page.

Snell's Law

Considering the point of contact of the leading wavefront in the first medium, we see that it must also be the contact point of the leading wavefront in the second medium. If it were not, then the second wavefront must either lead the first, which amounts to a effect before the cause, or it must lag, which would be a cause without an immediate effect.



Thus we see that the wavelength component measured along the interface, the x' direction, must be the same in both media. Since the frequency of oscillation does not change, we conclude that the apparent velocity of the wavefront, measured along the interface, is conserved.

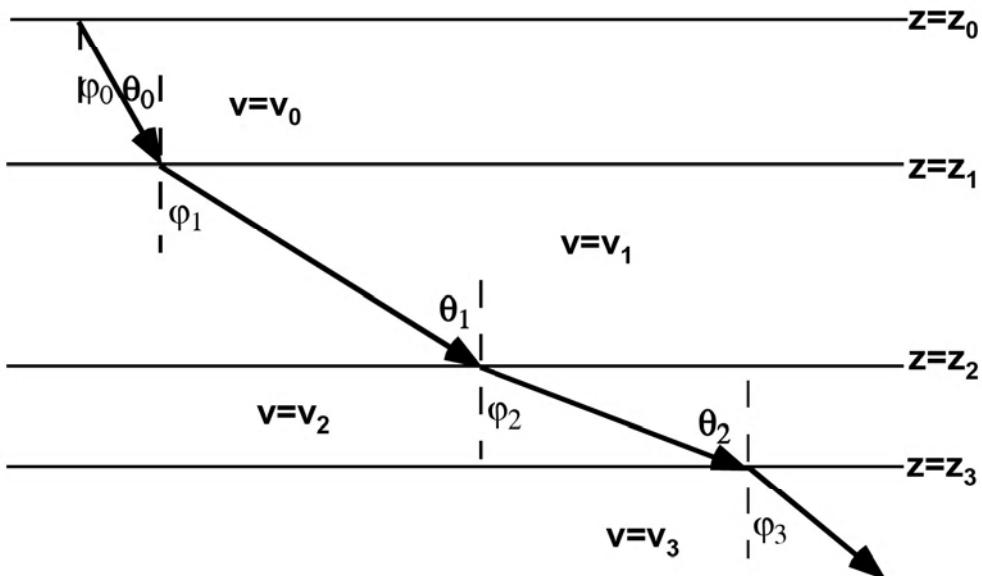
$$v_{1x'} = \lambda_{x'} f = v_{2x'}$$
$$v_{1x'} = \frac{v_1}{\sin(\theta)} = v_{2x'} = \frac{v_2}{\sin(\phi)}$$

This result is Snell's law. The angles involved can be considered as the angles between the wavefronts and the interface or between the raypaths and the normal to the interface.

Snells Law: Wavefront propagation across a velocity interface always conserves the apparent velocity along the interface.

Raytracing in a $V(z)$ Medium

Raytracing in a $v(z)$ medium can be simply described by conservation of the horizontal apparent "slowness" otherwise known as the ray parameter. Consider the application of Snell's law to the horizontally stratified earth, with layers of finite thickness, as shown below:



We have seen that Snell's law requires the conservation of the apparent velocity at each interface. Usually this is stated in terms of the apparent "slownesses" which are simply the velocity inverses. In conformity with the notation in the diagram, we have, at the j th interface:

$$\frac{\sin(\theta_{j-1})}{v_{j-1}} = \frac{\sin(\phi_j)}{v_j} \quad \text{but} \quad \phi_j = \theta_j$$

so

$$\frac{\sin(\theta_{j-1})}{v_{j-1}} = \frac{\sin(\theta_j)}{v_j}$$

Raytracing in a V(z) Medium

It is obvious now that, if we repeat this analysis at any other interface we will reach a similar conclusion:

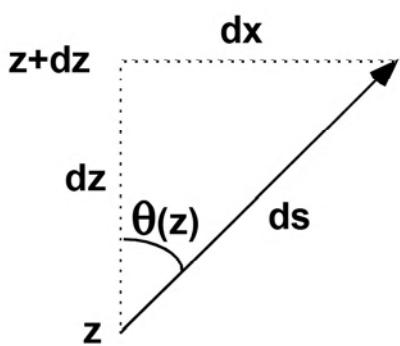
the quantity: $p = \frac{\sin(\theta_j)}{v_j}$ is conserved throughout the entire wave propagation.

p is generally referred to as the "ray parameter" since it is a unique constant for any ray and thus "parameterizes" the possible rays. We see that the conservation of p and its identification as the horizontal apparent slowness is a direct consequence of Snell's law. If we generalize this result to a continuous variation of v with z, we have:

$$p = \frac{\sin(\theta(z))}{v(z)} = \text{a constant for any particular ray}$$

Raytracing in a $V(z)$ Medium

We are now in a position to derive general expressions for the traveltime and distance travelled for any ray in a $v(z)$ medium. Consider a small "differential" element of such a raypath as shown below:



The horizontal distance, dx , is:

$$dx = \tan(\theta(z))dz$$

While the differential travelttime is:

$$dt = \frac{ds}{v(z)} = \frac{dz}{v(z)\cos(\theta(z))}$$

We incorporate Snell's law by simply saying: $\sin(\theta(z)) = pv(z)$

$$\text{Thus: } \cos(\theta(z)) = \sqrt{1 - \sin^2(\theta(z))} = \sqrt{1 - p^2 v^2(z)}$$

$$\text{and } \tan(\theta(z)) = \frac{\sin(\theta(z))}{\cos(\theta(z))} = \frac{pv(z)}{\sqrt{1 - p^2 v^2(z)}}$$

So the differential raypath expressions become:

$$dx = \frac{pv(z)}{\sqrt{1 - p^2 v^2(z)}} dz \quad \text{and} \quad dt = \frac{dz}{v(z)\sqrt{1 - p^2 v^2(z)}}$$

Raytracing in a V(z) Medium

We now obtain expressions for macroscopic raypaths by simply integrating these results:

$$x = \int_{z_1}^{z_2} \frac{pv(z)}{\sqrt{1-p^2v^2(z)}} dz$$

$$t = \int_{z_1}^{z_2} \frac{dz}{v(z)\sqrt{1-p^2v^2(z)}}$$

Given a ray parameter p , and a velocity function $v(z)$, we can compute the distance traveled and the traveltime by evaluating these integrals.

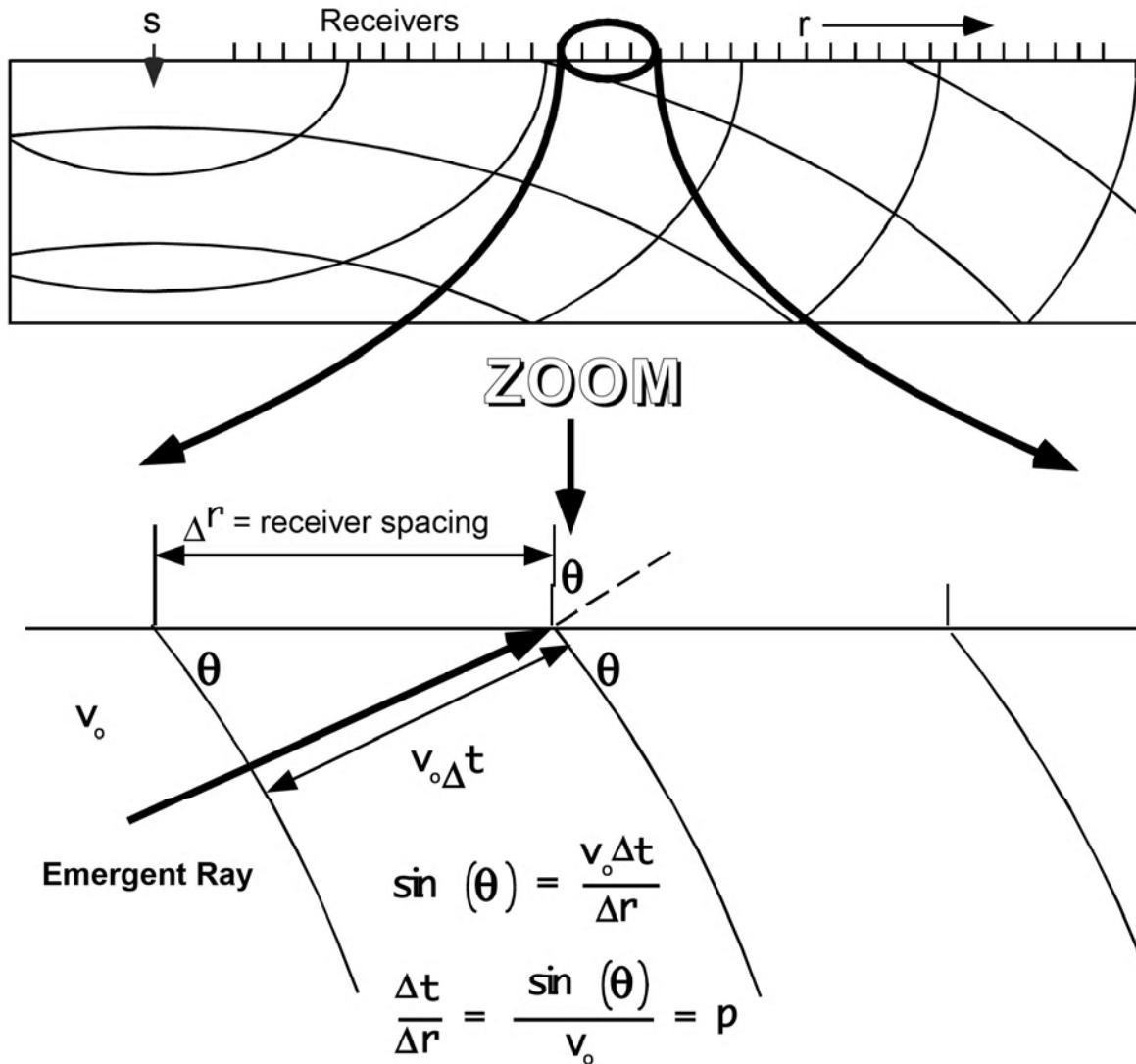
If we wish expressions for discrete layering, these are easily obtained:

$$x = \sum_{k=1}^n \frac{pv_k}{\sqrt{1-p^2v_k^2}} \Delta z_k \quad t = \sum_{k=1}^n \frac{\Delta z_k}{v_k \sqrt{1-p^2v_k^2}}$$

Note that these expressions are coupled through the ray parameter p . Though they are exact, there is no known general solution which eliminates p and gives $t(x)$ or $x(t)$ directly. Recalling that, in a real seismic experiment, many different wavefronts and hence p values, arrive at a given x , we realize that $t(x)$ must be a very complex multi-branch function.

Measurement of the Ray Parameter

It is a simple matter to measure the ray parameter of a wavefront emergent at a receiver array. In essence, the measurement is simply that of the apparent velocity of the wavefront as it sweeps across the receivers.



Since the wavefronts can be considered to be locally planar near any two receivers, we can use ordinary trigonometry to deduce that a measurement of $\Delta t / \Delta r$ gives the ray parameter. $\Delta t / \Delta r$ is also sometimes called the time dip or the horizontal slowness.

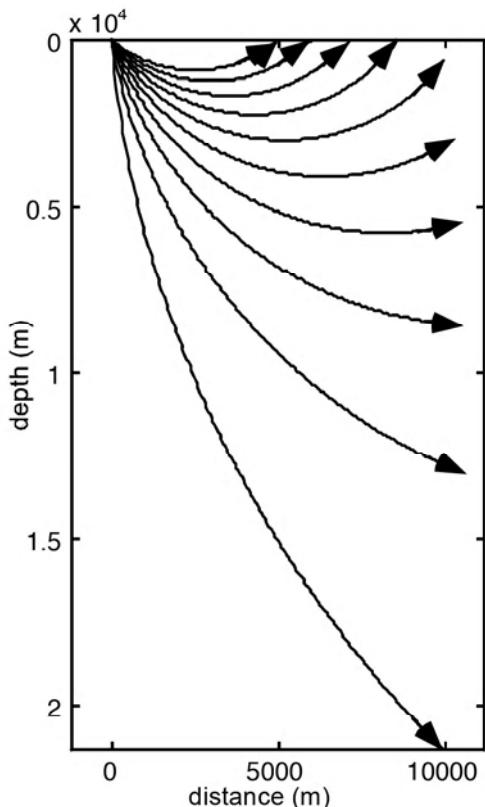
Raypaths when $v = v_o + cz$

It is not difficult to integrate the $v(z)$ raypath expressions for the case of linear velocity with depth (See Slotnick, 1959, Lessons in Seismic Computing, pp 205-211). The results are:

$$x = \frac{1}{pc} \left\{ \sqrt{1-p^2v_o^2} + \sqrt{1-p^2(v_o+cz)^2} \right\}$$

$$t = \frac{1}{c} \ln \left[\left(\frac{v_o+cz}{v_o} \right) \left(\frac{1+\sqrt{1-p^2v_o^2}}{1+\sqrt{1-p^2(v_o+cz)^2}} \right) \right]$$

Slotnick then shows the x equation to be that of a circle of radius $1/(pc)$ and centered at $x=\cos(\theta_o)/(pc)$ and $z=-v_o/c$.

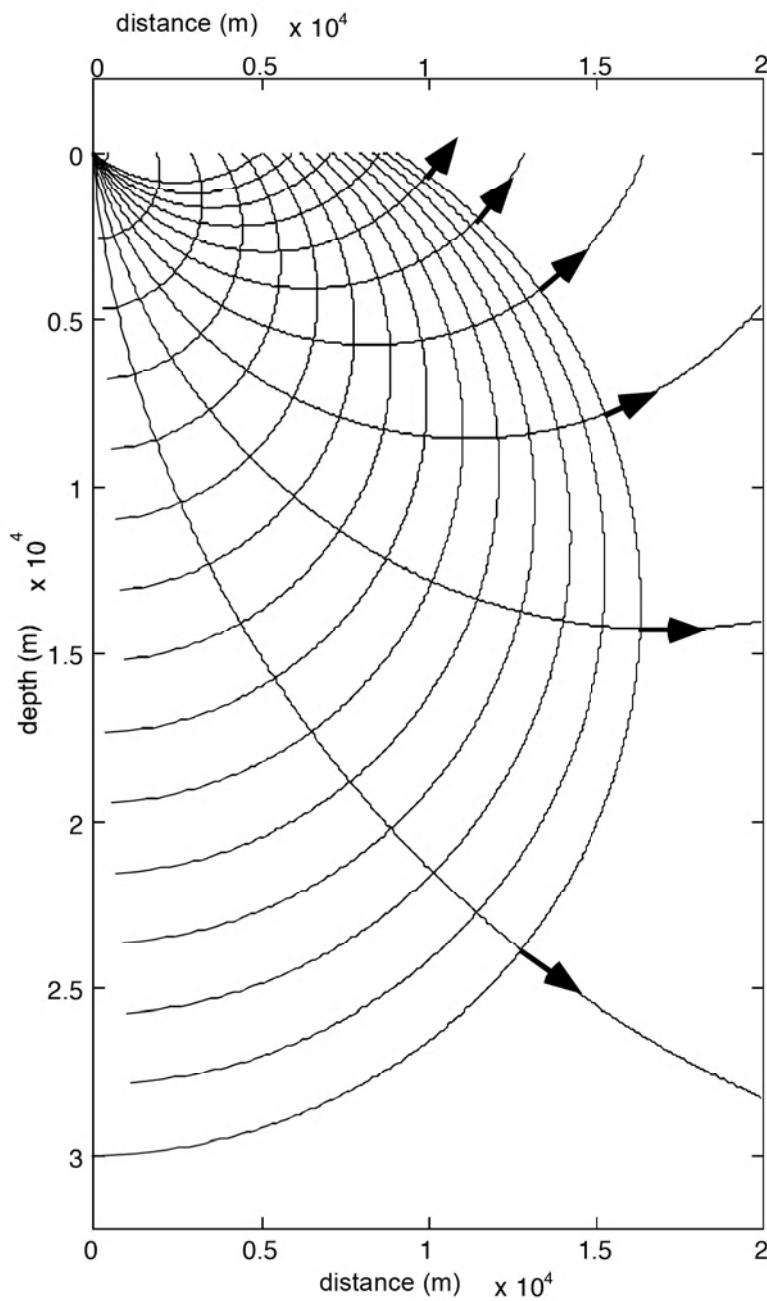


Raypaths for $v = 1800 + 0.6z$ and takeoff angles of $5^\circ, 10^\circ, 15^\circ \dots 50^\circ$. The raypaths are circles whose center is at a negative z and an x which is displaced to the right with increasing angle.

Raypaths when $v = v_o + cz$

Slotnick also shows the wavefronts to be circles whose centers are along the z axis at

$$z = \frac{v_o \cosh (cT - 1)}{c}$$



with radii of

$$r = \frac{v_o \sinh (cT)}{c}$$

where T is the traveltime along the wavefront.

The diagram shows the same raypaths as the previous page (for $v = 1800 + .6z$) and also shows 15 wavefronts for times:

0.2569	(all times in sec on ds)
1.0424	
1.5741	
1.9766	
2.3005	
2.5717	
2.8048	
3.0093	
3.1914	
3.3555	
3.5050	
3.6421	
3.7688	
3.8865	
3.9965	

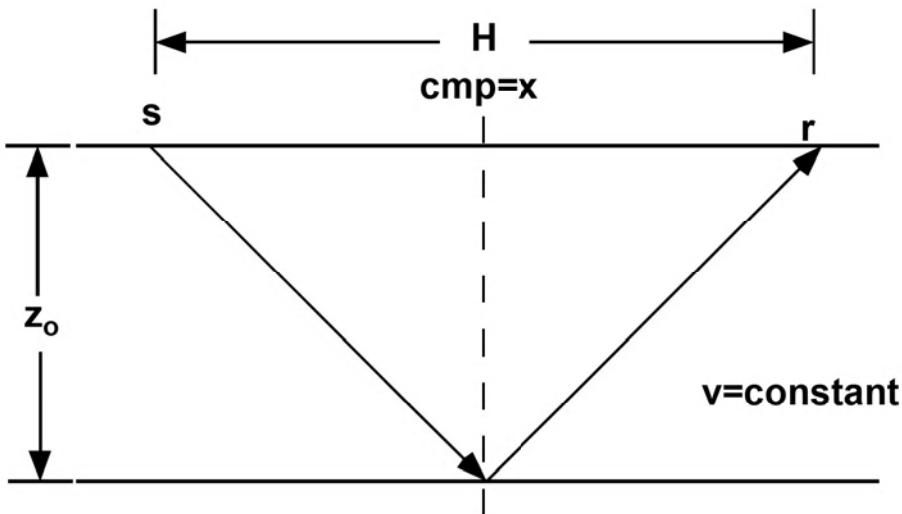
Methods of Seismic Data Processing

Lecture Notes Geophysics 557

Chapter 7 Normal Moveout and Stack

Normal Moveout

Consider the problem of computing the traveltime in the case shown below. By symmetry, we need only consider the downgoing path and double the result.



$$t_H = 2 \frac{\sqrt{z_0^2 + \frac{H^2}{4}}}{v} \longrightarrow t_H^2 = \frac{4z_0^2}{v^2} + \frac{H^2}{v^2}$$

$$\text{so } t_H^2 = t_0^2 + \frac{H^2}{v^2}, \quad t_0 = \frac{2z_0}{v}$$

This is the equation of a hyperbola in (t, H) space. So we see that the simplest possible reflection signature is hyperbolic. The time difference $t_H - t_0$ is called the "normal moveout" or nmo.

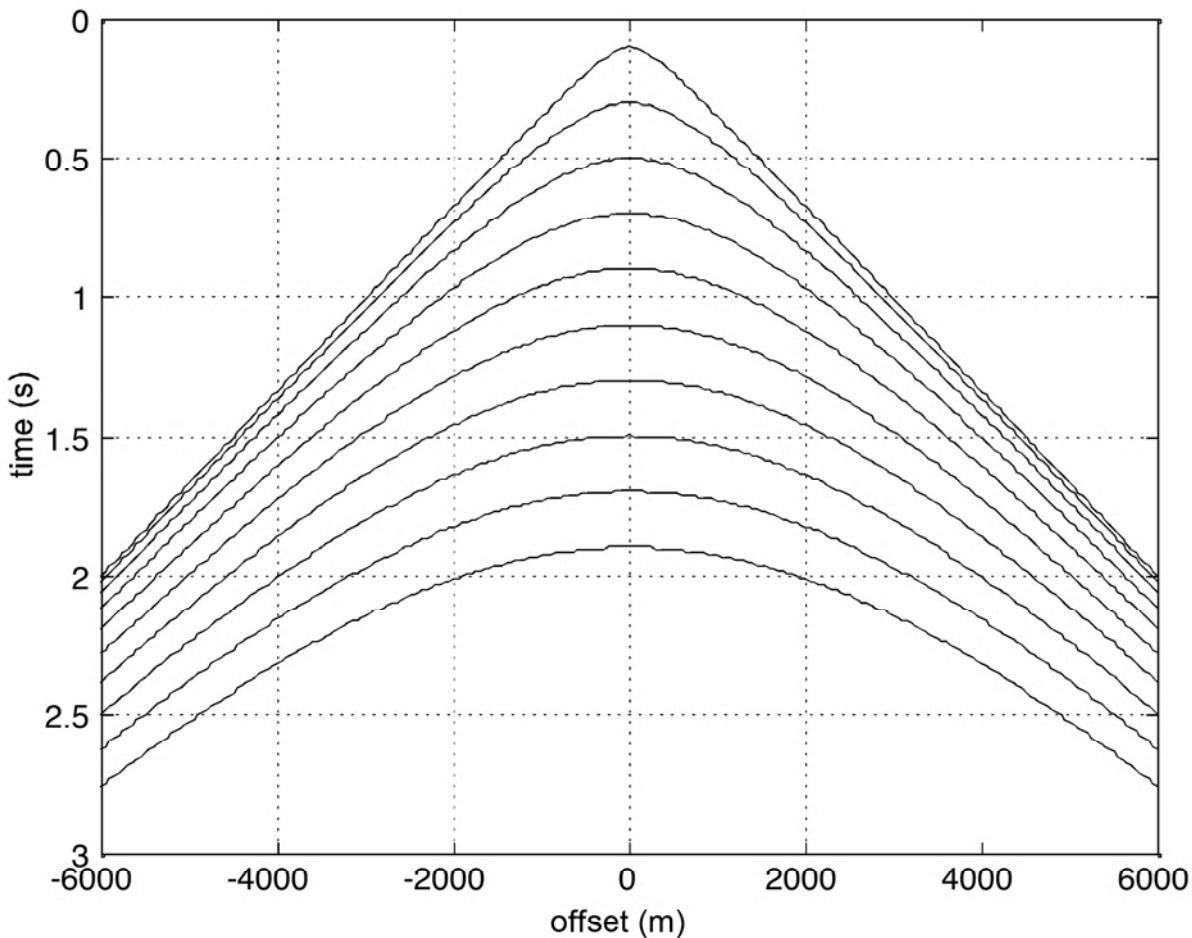
MOVEOUT

$$\Delta t = t_H - t_0 = t_H \left(1 - \sqrt{1 - \frac{H^2}{t_H^2 v^2}} \right) = \text{nmo}$$

$$\Delta t \approx \frac{H^2}{2t_H v^2}, \quad H \ll vt_H = \text{parabolic approximation to nmo}$$

Normal Moveout

Here is a graph of ten nmo curves for $v=3000$ m/s and a number of different t_0 times.

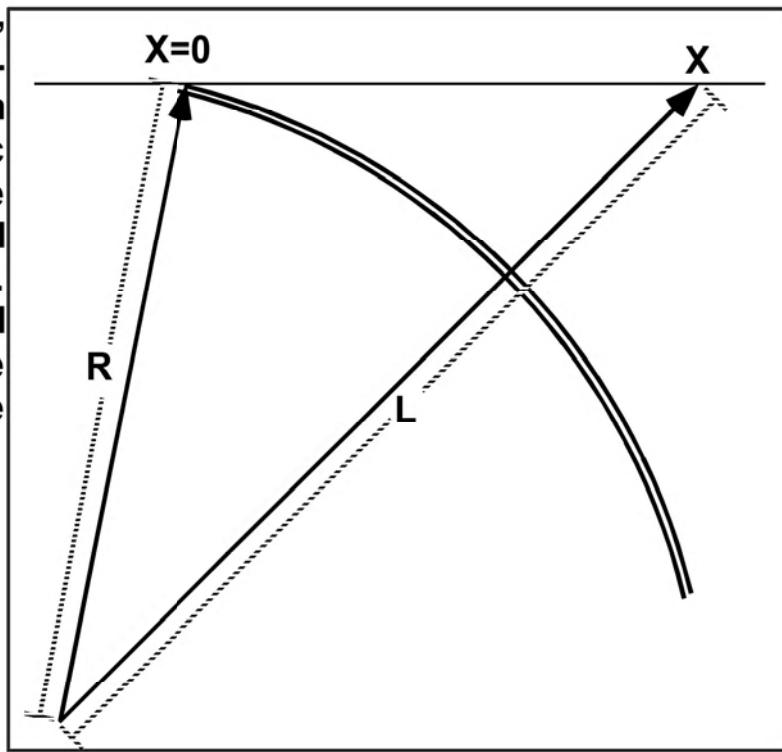


We note that the family of hyperbolae all approach the same asymptotes defined by $t = +/- H/v$.

Normal Moveout

Consider the approach of a spherical wavefront to a recording array as shown below. We will show that its traveltime signature, (t_x, x) is a hyperbola.

The inference is then that hyperbolic traveltime curves are indicating spherical wavefronts. (However, ellipsoidal wavefronts also have hyperbolic traveltime curves.)



If the arrival time at $x=0$ is denoted t_0 , then the time at X can be written as:

$$t_x = t_0 + \frac{(L - R)}{V}$$

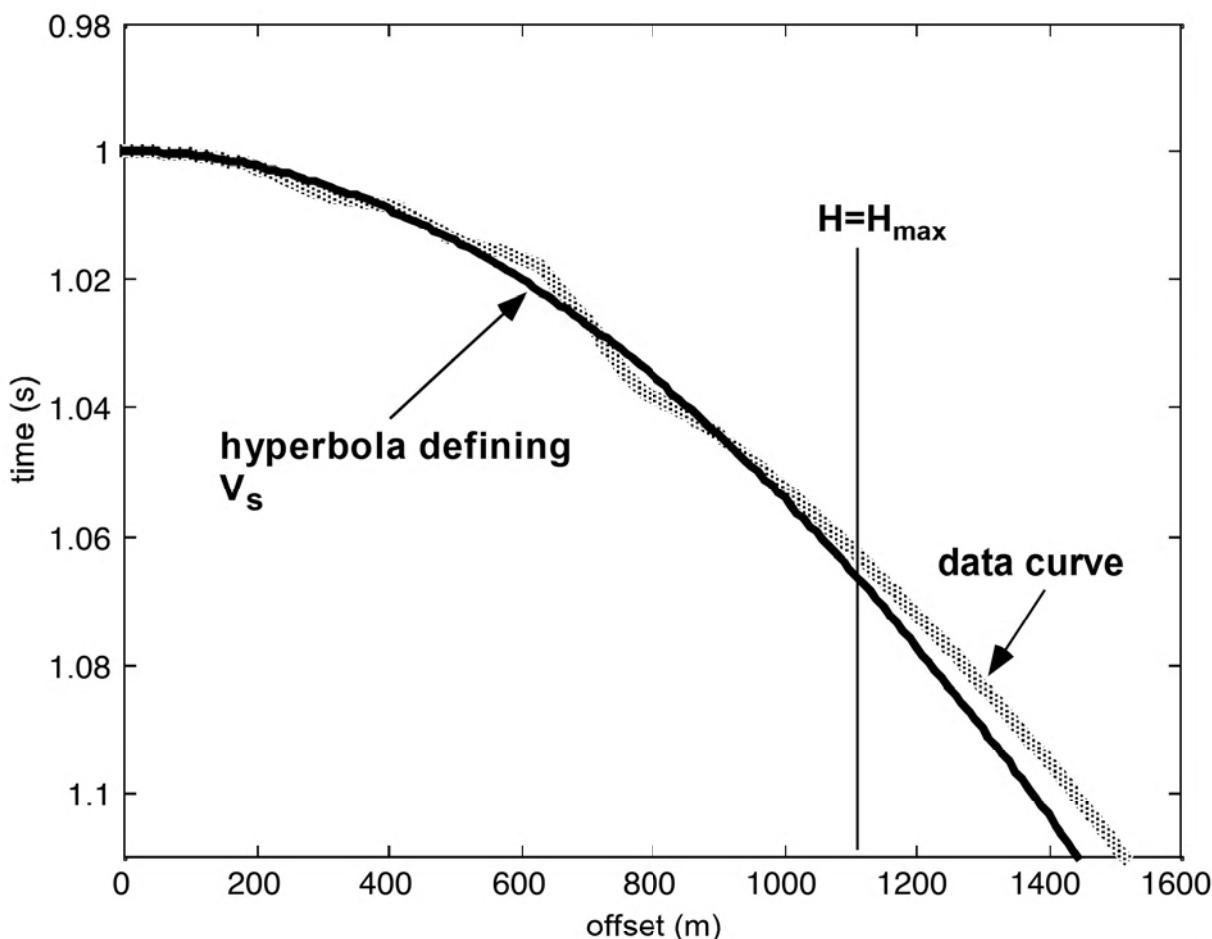
This can be rearranged to give the hyperbolic form:

$$(t_x - \alpha)^2 - \left(\frac{X}{V}\right)^2 = \left(\frac{R}{V}\right)^2 \quad \text{where} \quad \alpha = t_0 - \frac{R}{V}$$

Stacking Velocity

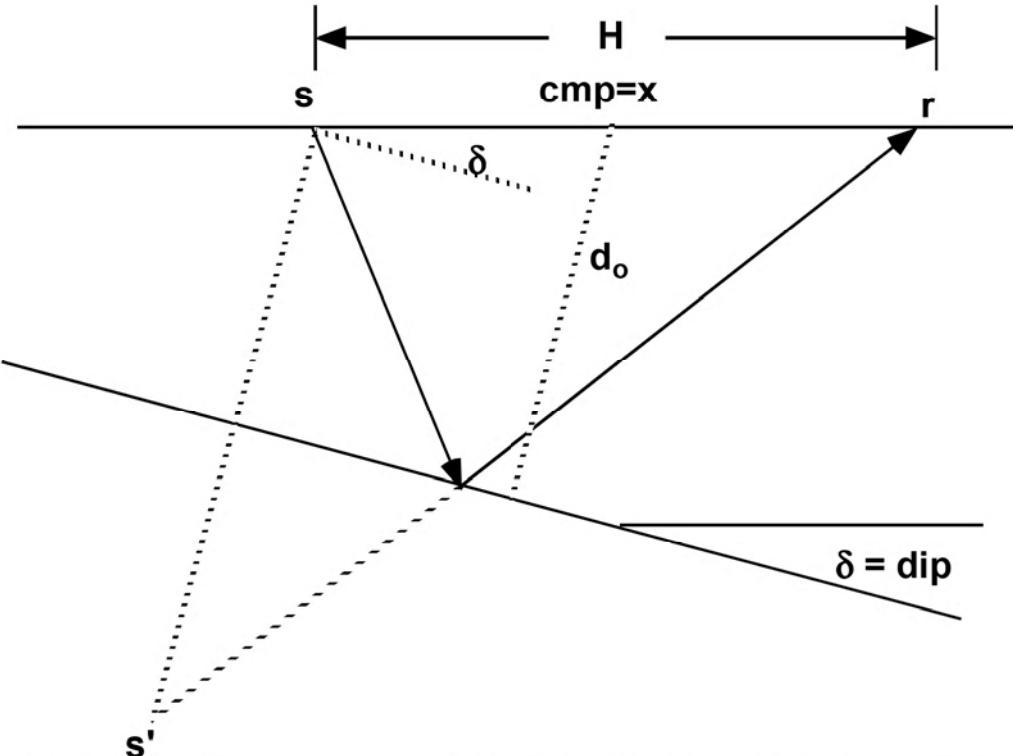
$$v_s$$

Unlike most other velocity representations, stacking velocity is an inherently empirical quantity obtained by optimizing some attribute of a cmp gather. While difficult to define in many realistic situations, it still has a simple intuitive meaning:



v_s can be thought of as the velocity parameter which defines a "best fit" hyperbola to the actual data traveltimes. The precise meaning of best fit depends on the software being used for analysis. In all cases, v_s is a function of the maximum offset used in the analysis and will change if H_{\max} changes.

Normal Moveout and Reflector Dip



Using the concept of an image source, s' , we see that the traveltime from s to r , is equal to the length of $s'r$ divided by v . We can calculate this length using the law of cosines. First note that:

$$\angle rss' = \frac{\pi}{2} + \delta \quad \text{and} \quad \frac{1}{2}ss' = d_o - \frac{H \sin(\delta)}{2}$$

Applying the law of cosines to triangle $ss'r$ gives:

$$(s'r)^2 = (s's)^2 + H^2 - 2(s's)H \cos\left(\frac{\pi}{2} + \delta\right)$$

$$(s'r)^2 = (2d_o - H \sin(\delta))^2 + H^2 - 2H(2d_o - H \sin(\delta))(-\sin(\delta))$$

$$\text{so } (s'r)^2 = 4d_o^2 - 4d_o H \sin(\delta) + H^2 \sin^2(\delta) + H^2 + 4Hd_o \sin(\delta) - 2H^2 \sin^2(\delta)$$

$$\text{and finally } (s'r)^2 = 4d_o^2 + H^2(1 - \sin^2(\delta)) = 4d_o^2 + H^2 \cos^2(\delta)$$

Normal Moveout and Reflector Dip

Thus the desired traveltime expression is:

$$t_H^2 = t_o^2 + \frac{H^2 \cos^2(\delta)}{v^2} \quad \text{where} \quad t_o = \frac{2d_o}{v}$$

So we see that the reflection traveltime curve is still hyperbolic but the characteristic velocity is $v/\cos(\delta)$. Thus, the stacking velocity is:

$$\text{stacking velocity} = v_s = \frac{v}{\cos(\delta)}$$

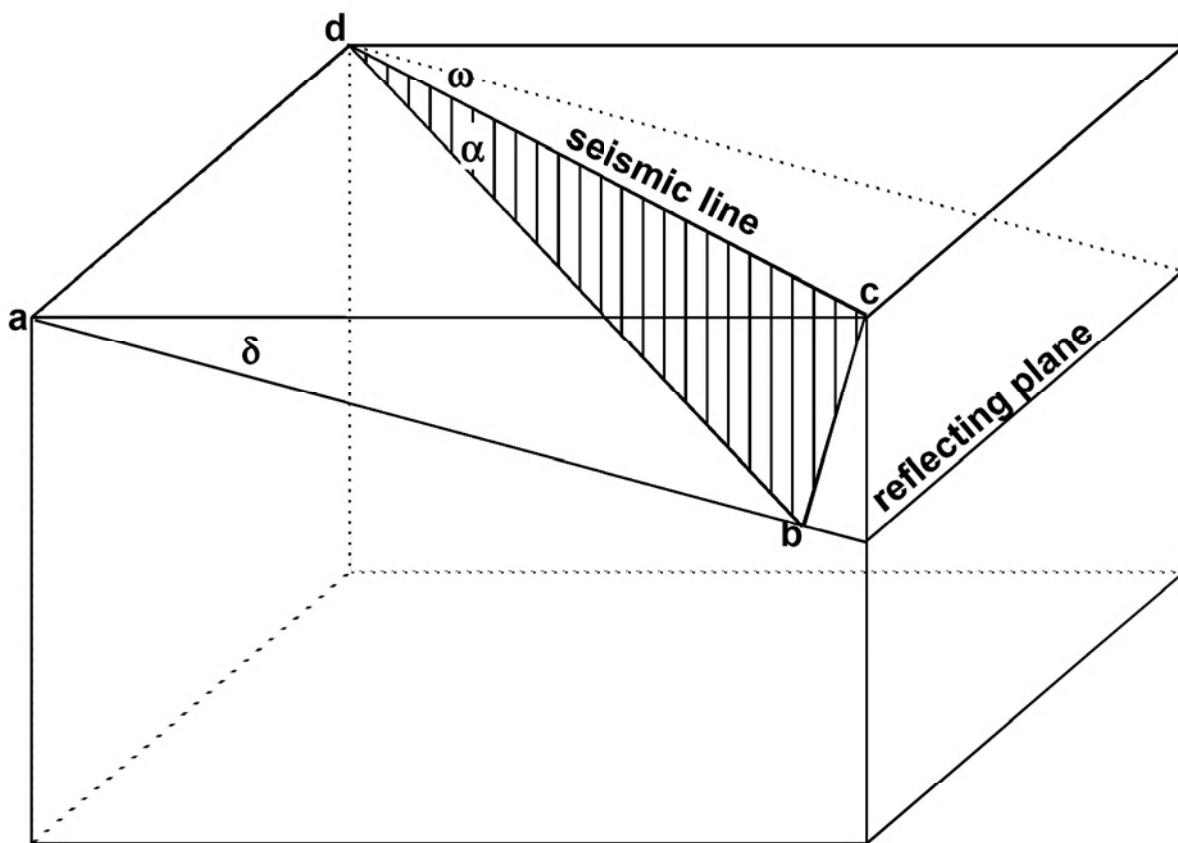
For dipping reflector
beneath a constant
velocity overburden.

The time difference implied by this result is most properly called dip dependent normal moveout (ddnmo) to distinguish it from the zero dip expression:

$$\Delta t = t_H - t_o = t_H \left(1 - \sqrt{1 - \frac{H^2 \cos^2(\delta)}{t_H^2 v^2}} \right) \quad \text{ddnmo}$$

$$\Delta t = t_H - t_o = t_H \left(1 - \sqrt{1 - \frac{H^2}{t_H^2 v^2}} \right) \quad \text{nmo}$$

Normal Moveout and Reflector Dip



When we consider the extension of the dipping reflector result to 3-D, we must use the apparent dip in the direction of the seismic line. (It can be shown that, in a $v(z)$ medium, all raypaths propagate in a vertical plane. See Sherrif and Geldart for more info.) Thus we must use:

$$t_H^2 = t_o^2 + \frac{H^2 \cos^2(\alpha)}{v^2}$$

where $\cos(\alpha) = \sqrt{1 - \sin^2(\delta)\cos^2(\omega)}$

Normal Moveout and Reflector Dip

To derive the relation between angles, consider the diagram and note that:

$$\sin(\delta) = \frac{bc}{ac} \quad \text{and} \quad \cos(\omega) = \frac{ac}{dc}$$

$$\cos(\alpha) = \frac{db}{dc} = \frac{\sqrt{dc^2 - bc^2}}{dc} = \sqrt{1 - \frac{bc^2}{dc^2}}$$

Since $bc=ac \sin(\delta)$ and $dc=ac/\cos(\omega)$, then $bc/dc = \sin(\delta)/\cos(\omega)$ and we obtain:

$$\cos(\alpha) = \sqrt{1 - \sin^2(\delta)\cos^2(\omega)}$$

So we see that stacking velocities depend not only on reflector dip but upon the azimuth of the seismic line with respect to the dipping reflector.

$$\text{stacking velocity} = v_s = \frac{v}{\sqrt{1 - \sin^2(\delta)\cos^2(\omega)}}$$

For dipping reflector beneath a constant velocity overburden in 3D

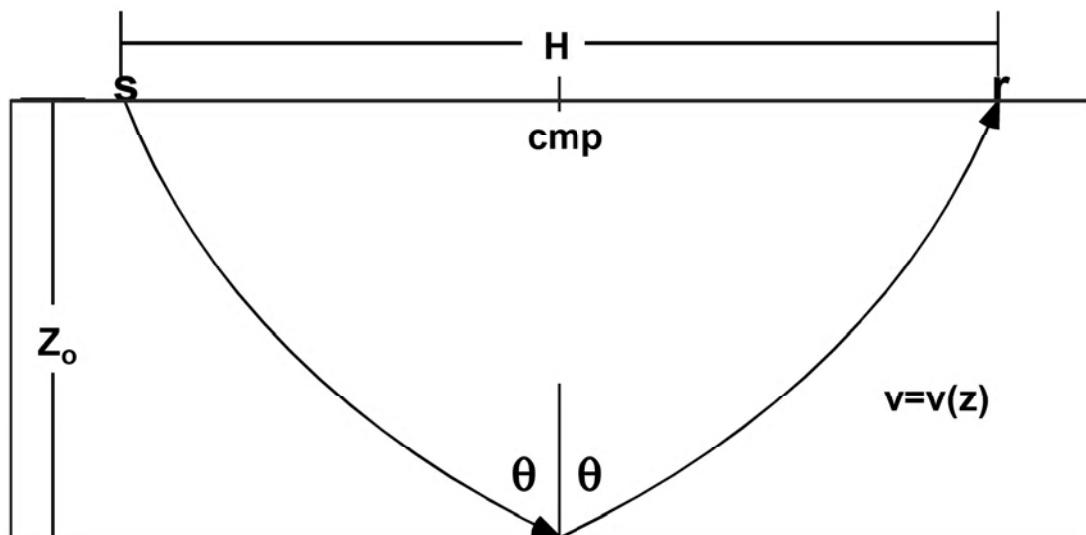
NMO for a $v(z)$ Medium

We now consider the normal moveout problem for a horizontal reflector beneath a $v(z)$ medium. Though this problem has never been solved in closed form, a number of very useful technical analyses have been published which give detailed presentations of the infinite series in powers of p which we present here. For further details consult:

Al Chalabi, M., 1973, Series approximations in velocity and traveltimes computations, Geophys. Prosp., V. 21, pp 783-795

Dix, C.H., 1955, Seismic velocities from surface measurements, Geophysics, v 20, pp 68-86

Taner, M.T., and Koehler, F., 1969, Velocity spectra -digital computer derivation and applications of velocity functions, Geophysics, v 34, pp 859-884



All of these analyses start from the $v(z)$ raypath integrals and seek a relation between t and h of the form:

$$t_h^2 = C_1 + C_2 h^2 + C_3 h^4 + C_4 h^6 + \dots$$

Thus we seek a series relationship which contains only even powers of offset. Odd powers are intuitively forbidden since they would result in an expression which depends on the sign of the offset.

NMO for a $v(z)$ Medium

Dix first derived the value of c_1 and c_2 then later Al Chalabi and Taner and Koehler analyzed a number of the higher order terms. The basic method involved expanding the square roots in the traveltime integrals:

$$x = \int_{z_1}^{z^2} \frac{pv(z)}{\sqrt{1-p^2v^2(z)}} dz \quad t = \int_{z_1}^{z^2} \frac{dz}{v(z)\sqrt{1-p^2v^2(z)}}$$

as power series in p , substituting those expressions into the assumed series form, and equating common powers of p . Please refer to the references for details. The results for the first three terms are:

$$c_1 = t_0^2 = \left(2 \int_0^z \frac{dz}{v(z)} \right)^2 = \left(2 \frac{z}{v_{ave}} \right)^2$$

No surprise here, its just the 2-way zero offset time squared.

$$c_2 = \frac{1}{v_{ave} v_m} = \frac{1}{v_{rms}^2}$$

Here is the basis for the expectation that stacking velocities are rms velocities

where: $v_{ave} = \frac{2z_o}{t_o}$ and $v_m = \frac{1}{Z_o} \int_0^{z_o} v(z) dz$

$$c_3 = \frac{-1t_o \bar{v}_3}{32 Z_o^3 V_m^4} + \frac{1}{16 Z_o^2 V_m^2}$$

You can see why we are always hoping to be able to neglect this one!

where: $\bar{v}_3 = \frac{1}{Z_o} \int_0^{z_o} v^3(z) dz$

NMO for a $v(z)$ Medium

So we see that, to the extent that the terms higher than c_2 are negligible, we are justified in considering stacking velocities to be approximated by rms velocities. Of course we must also assume: no lateral velocity variations and no reflector dip.

The coefficient c_3 is called the "fourth order moveout coefficient" and has been shown (Al Chalabi 1973) to be always negative. Generally c_3 is not significant in P-wave reflection seismology but can have an effect when $h/z_o \gg 1$ and in converted wave or shear wave exploration. It can be shown that c_3 can be neglected whenever:

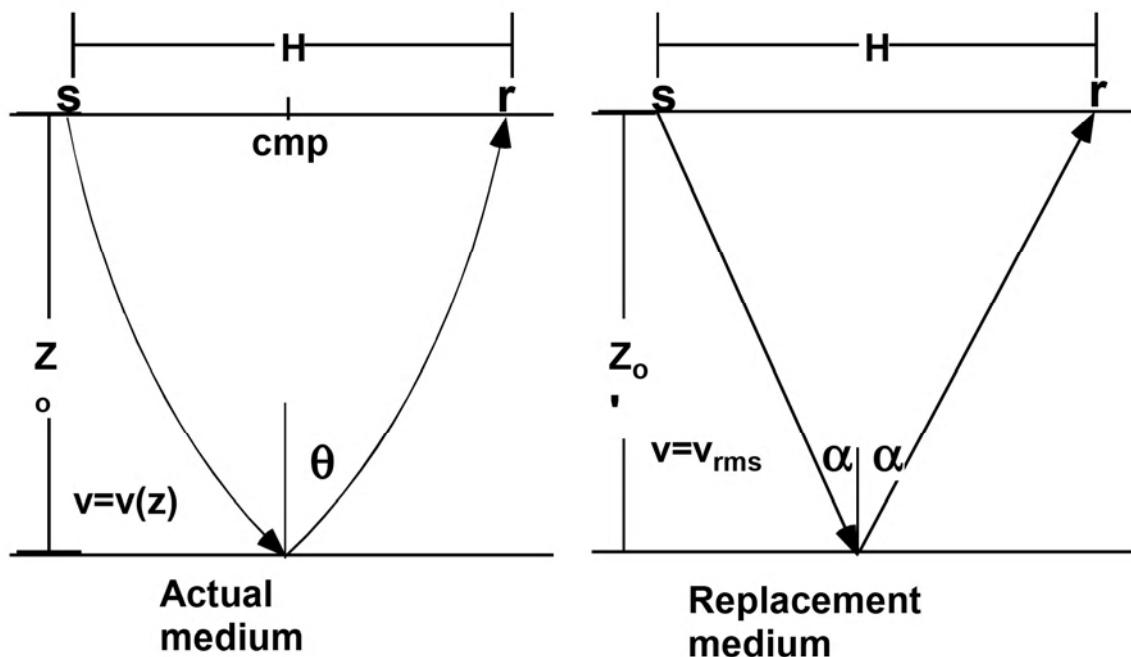
$$\frac{H^2}{z_o^2} \ll \left| \frac{V_{ave}}{16 V_m} - \frac{\bar{V}_3}{16 V_m^3} \right|^{-1}$$

Dix Equation Moveout

When the constant velocity nmo equation is written with the rms velocity of a $v(z)$ medium, the result is often called the Dix equation or Dix equation nmo:

$$t_H^2 = t_o^2 + \frac{H^2}{v_{rms}^2(t_o)} \quad \text{The Dix Equation}$$

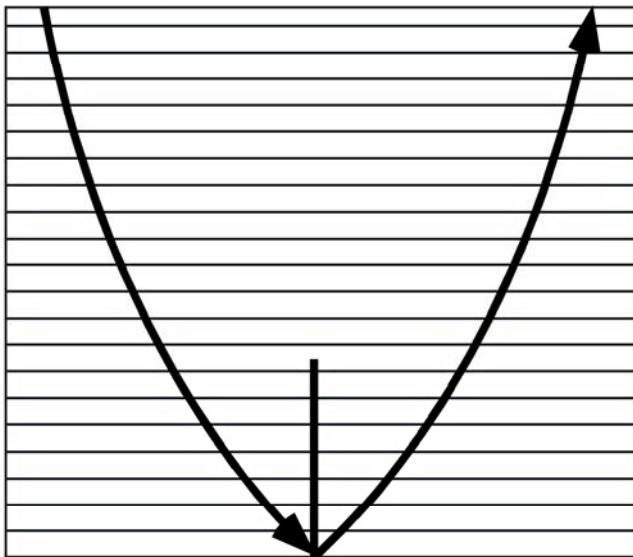
First notice the explicit dependence of v_{rms} on t_o in the notation. Second, consider the physical interpretation of this result. In using an explicit hyperbolic form, we are stating that the wavefronts in a $v(z)$ world are spherical to a high precision. Also, we are stating that we can approximately solve the complicated $v(z)$ raytracing problem by using a "replacement medium" which has a constant velocity equal to the v_{rms} from the surface to the reflector of interest:



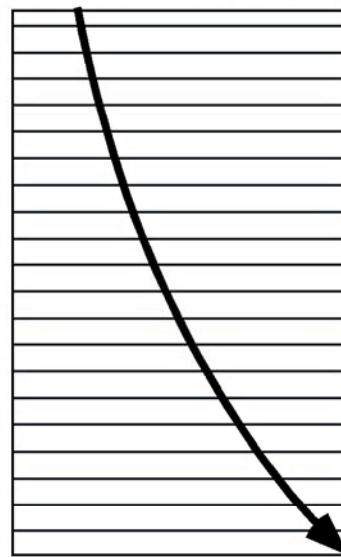
Dix Equation Moveout

Since the problem we've addressed is symmetric, we also now have an accurate, fast way to trace a ray one-way through a stack of layers:

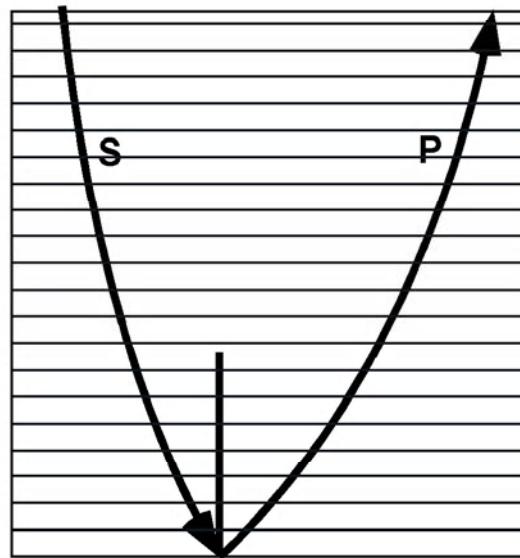
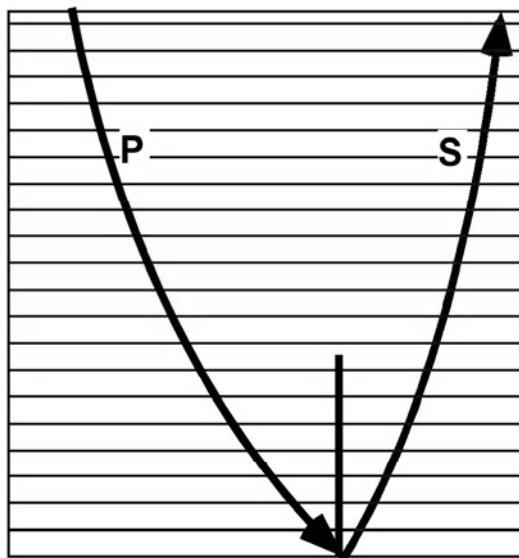
Solved: Two way P-P or S-S raytracing. Use two-way rms velocity.



Also solved: One way P or S raytracing. Use one way rms velocity.

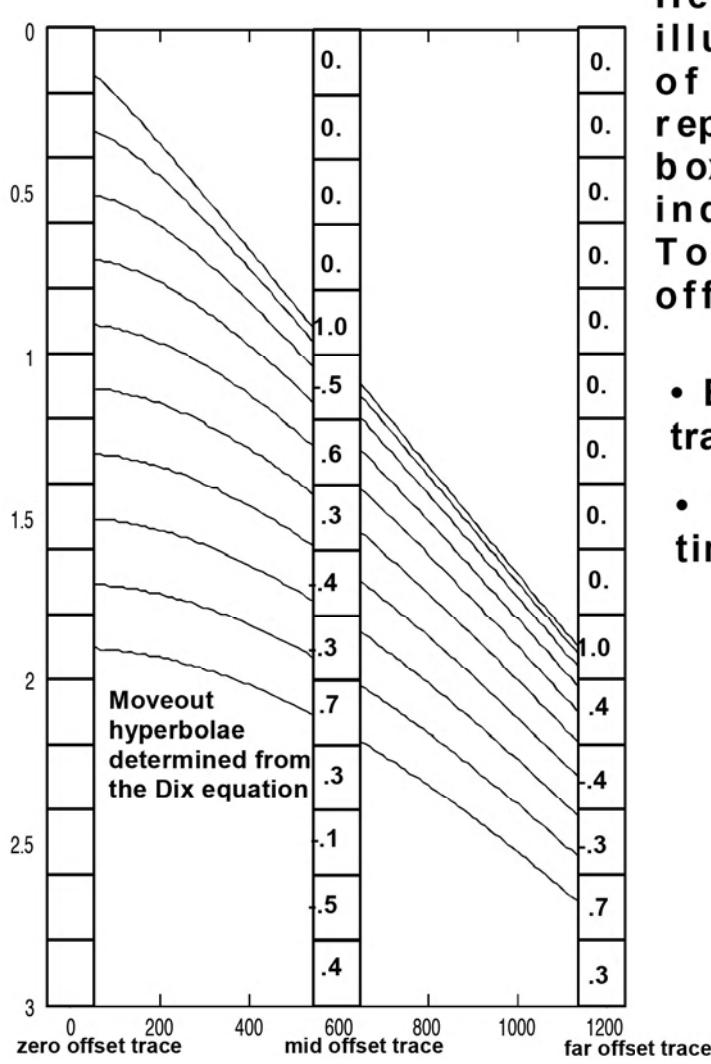


Also solved: Two way P-S or S-P raytracing. Use one way rms velocities for each leg.



Normal Moveout Removal

The removal of normal moveout is a fundamental step in seismic data processing. It amounts to a mapping of a seismic trace recorded at offset h to what would have been recorded at zero offset. Generally, NMOR algorithms require as input the stacking velocity function $v_s(t_o)$, giving v_s as a function of t_o , at the source receiver midpoint of the trace to be corrected. All traces with the same midpoint coordinate get the same NMOR correction.



Here we see a simplified illustration of the process of NMOR. Traces are represented as columns of boxes with numbers indicating sample values. To correct any trace to zero offset:

- Begin with a blank zero offset trace.
- For each zero offset sample time t_o :
 1. Use the NMO equation to determine the sample time t_h at offset which maps t_o .
 2. Interpolate a sample at t_h on the offset trace and map it to t_o . (i.e. place it in the appropriate sample box on the zero offset trace)

Normal Moveout Removal

Several points need elaboration:

- The interpolation of samples on the offset trace must be done with an accurate method. Usually some sort of approximate sinc function method is used. A poor interpolator results in a loss of high frequency information on the corrected trace.
- The convergence of moveout hyperbolae at offset means that approximately the same offset trace sample can be mapped to a number of adjacent zero offset times. This results if a very low frequency result and is known as "moveout stretch". Objectionable levels of stretch are usually removed with a "stretch mute".
- The stacking velocity function (or moveout function) must already be determined. This is generally done in a process called "velocity analysis".

Extension Of NMO and Dip to $v(z)$

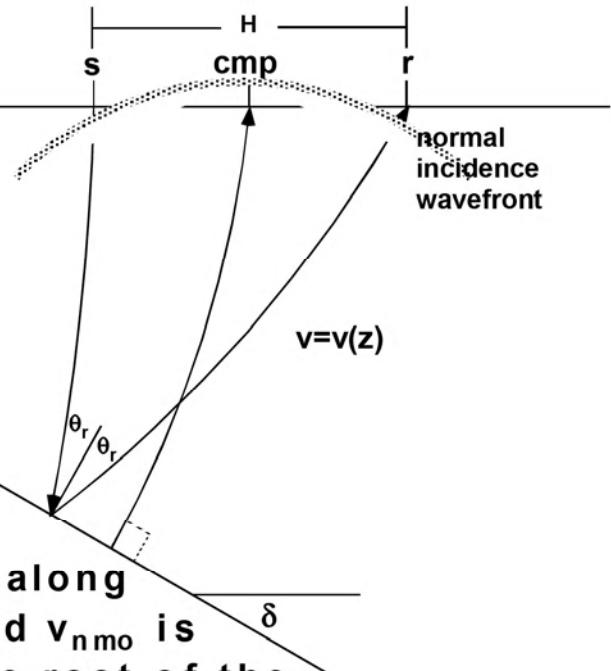
In arbitrarily complex media, the concept of a stacking velocity is difficult to apply since traveltime versus offset curves may be greatly different from hyperbolae. The most extreme situations are better treated with prestack migration but there is a class of problems, between the strictly stratified earth and the extremely complex settings, which can still yield to an extension of the NMO concept. Consider again the expansion of the traveltime expression in powers of offset squared:

$$t_H^2 = c_1 + c_2 H^2 + c_3 H^4 + c_4 H^6 + \dots$$

In most cases, we can also consider this to be a traveltime expansion about the normal incidence ray, since that raypath tends to dominate at zero offset. In this case, we define:

$$c_1 = t_n^2 \quad c_2 = \frac{1}{v_{nmo}^2}$$

So t_N is the two-way traveltime along the normal incidence raypath and v_{nmo} is defined to be the inverse square root of the coefficient of H^2 in the series expansion. Hubral and Krey¹ have performed the most general analysis in this regard and have shown that v_{nmo} can be regarded as a measure of the curvature of the "normal incidence wavefront", a hypothetical wavefront which is emitted at the normal incidence reflection point and arrives at receivers arrayed symmetrically about the cmp after



Extension Of NMO and Dip to v(z)

propagating at half speed. They also derived an expression for v_{nmo} in a medium of constant velocity layers with arbitrary dip though this has not been used extensivly in seismic data processing.

A practical extension of the concepts of the Dix equation and the $\cos(\delta)$ correction, that is widely used, can be formulated as "The Replacement Medium Hypothesis". If we consider the problem of a dipping reflector beneath a $v(z)$ medium, then we might postulate that the raytracing can be approximated by straight rays through a $v=v_{rms}$ medium with a $\cos(\delta)$ correction, but what dip should we use? A reasonable clue is obtained by examining the expression for v_{nmo} from a dipping reflector in constant velocity:

$$t_H^2 = t_o^2 + \frac{H^2 \cos^2(\delta)}{v^2} \quad \text{eqn 1}$$

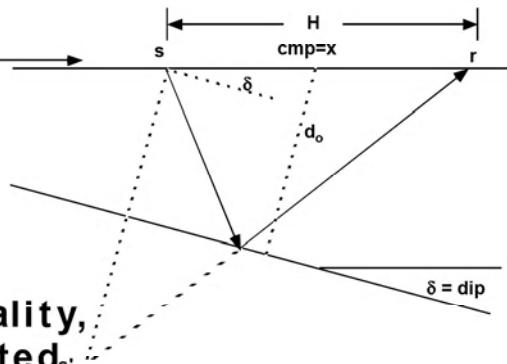
$$\text{where } t_o = \frac{2d_o}{v} \quad \text{eqn 2}$$

We can, with complete generality, take the x origin as the extrapolated point of outcrop of the dipping reflector so that:

$$d_o = x \sin(\delta)$$

and:

$$t_{xH}^2 = \frac{4x^2 \sin^2(\delta)}{v^2} + \frac{H^2 \cos^2(\delta)}{v^2} \quad \text{eqn 3}$$

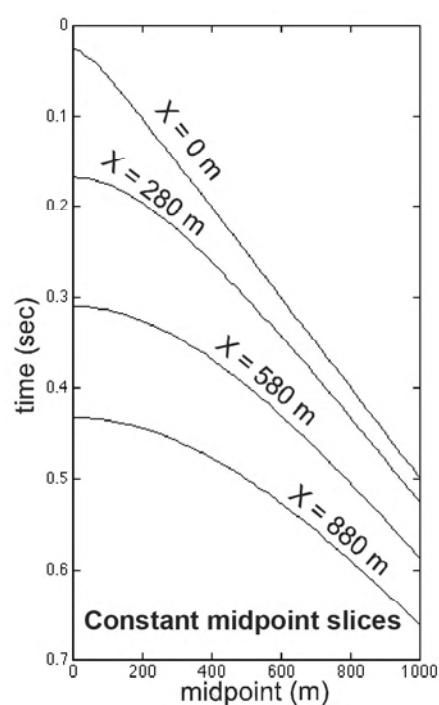
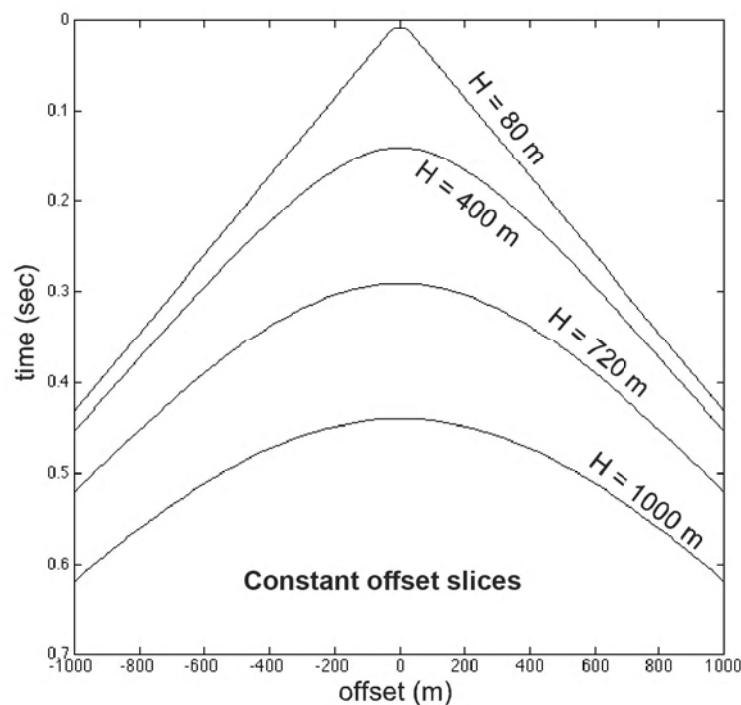
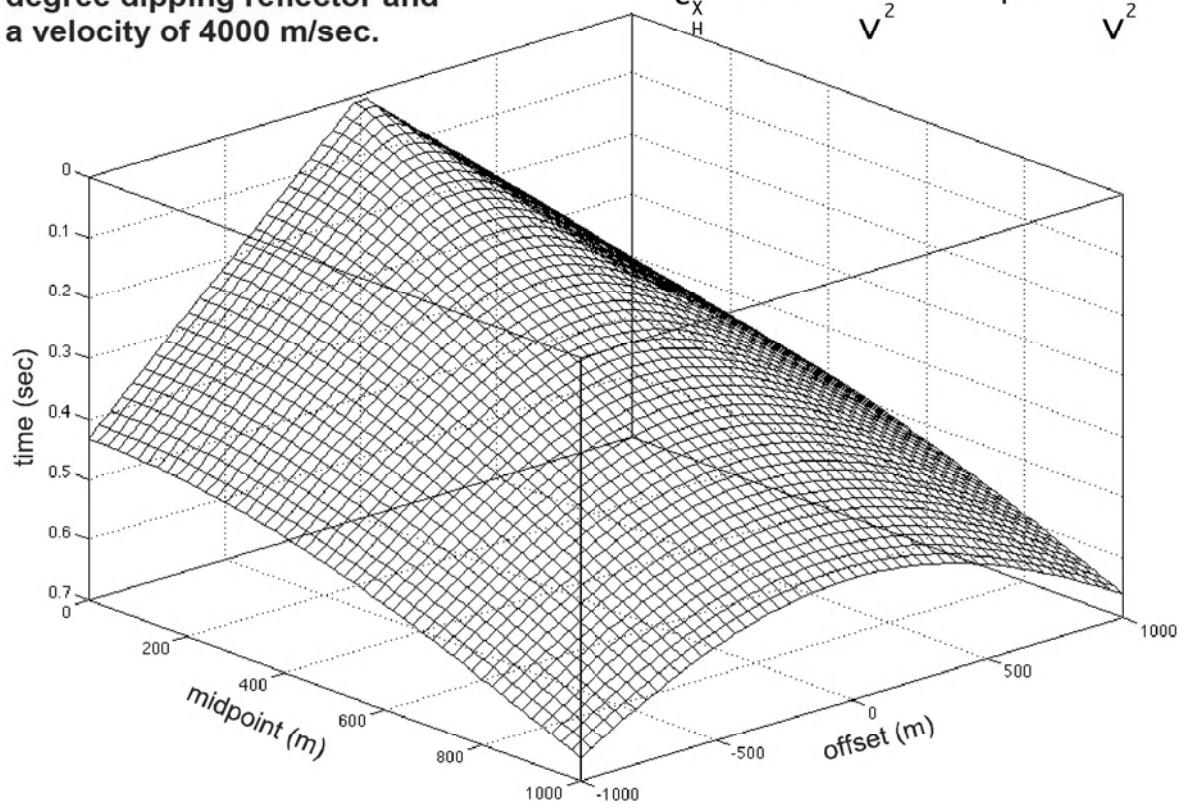


where we now subscript t with x and h to denote the explicit dependence on both.

Extension Of NMO and Dip to v(z)

Graph of equation 3 for a 30 degree dipping reflector and a velocity of 4000 m/sec.

$$t_x^2 = \frac{4x^2 \sin^2(\delta)}{v^2} + \frac{H^2 \cos^2(\delta)}{v^2}$$



Extension Of NMO and Dip to $v(z)$

Consider the gradient of equation 3 $\partial t/\partial x$:

$$\frac{\partial t_{xH}}{\partial x} = \frac{1}{2} \left(\frac{4x^2 \sin^2(\delta)}{v^2} + \frac{H^2 \cos^2(\delta)}{v^2} \right)^{-1/2} \left(\frac{8x \sin^2(\delta)}{v^2} \right)$$

Now setting $H=0$:

$$\left. \frac{\partial t_{xH}}{\partial x} \right|_{H=0} = \frac{1}{2} \left(\frac{v}{2x \sin(\delta)} \right) \left(\frac{8x \sin^2(\delta)}{v^2} \right) = 2 \frac{\sin(\delta)}{v}$$

Or simply:

$$\frac{1}{2} \frac{\partial t_o}{\partial x} = \frac{\sin(\delta)}{v} \quad \text{eqn 4}$$

Thus, we now rewrite equation 1 as:

$$t_H^2 = t_o^2 + \frac{H^2}{v_{nmo}^2} \quad \text{with} \quad v_{nmo} = \frac{v}{\sqrt{1 - \left(\frac{v \partial t_o}{2 \partial x} \right)^2}} \quad \text{eqn 5}$$

Where we have used:

$$\cos(\delta) = \sqrt{1 - \left(\frac{v \partial t_o}{2 \partial x} \right)^2} \quad \text{eqn 6.}$$

So far, these results are exact and are simply a reformulation of the constant velocity result. Note that v_{nmo} as given by equation 5, is now expressed in terms of quantities which can be directly estimated from seismic data.

Extension Of NMO and Dip to v(z)

We now invoke the replacement medium hypothesis and assert that these results generalize to a $v(z)$ medium as:

$$t_H^2 = t_o^2 + \frac{H^2}{v_{nmo}^2} \quad v_{nmo} = \frac{v_{ms}(t_o)}{\sqrt{1 - \left(\frac{v_{ms}(t_o) \partial t_o}{2 \partial x} \right)^2}} \quad \text{eqn 7}$$

$$\cos(\delta) = \sqrt{1 - \left(\frac{v_{ms}(t_o) \partial t_o}{2 \partial x} \right)^2} \quad \text{eqn 8}$$

Several points should be stressed:

- These results are asserted by hypothesis and analogy and have not been formally derived.
- Equation 8 must be regarded as yielding an apparent dip.
- These results are correct in the limit as $v(z) \rightarrow$ constant and in the limit as $\delta \rightarrow 0$. They have been in use for many years in the industry and can be regarded as having empirical backing.

The application of equations 7 and 8 requires the determination of both v_{nmo} and $\partial t_o / \partial x$. v_{nmo} commonly comes from velocity analysis while $\partial t_o / \partial x$ can be obtained by steering for dip as an added dimension in the velocity analysis or by simply picking the dips on the resultant stack. Having obtained estimates of both parameters, the rms velocity can be computed as:

$$v_{ms} = \frac{v_{nmo}}{\sqrt{1 + \left(\frac{v_{nmo} \partial t_o}{2 \partial x} \right)^2}} \quad \text{eqn 9}$$

These results can then be used in Dix interval velocity estimations.

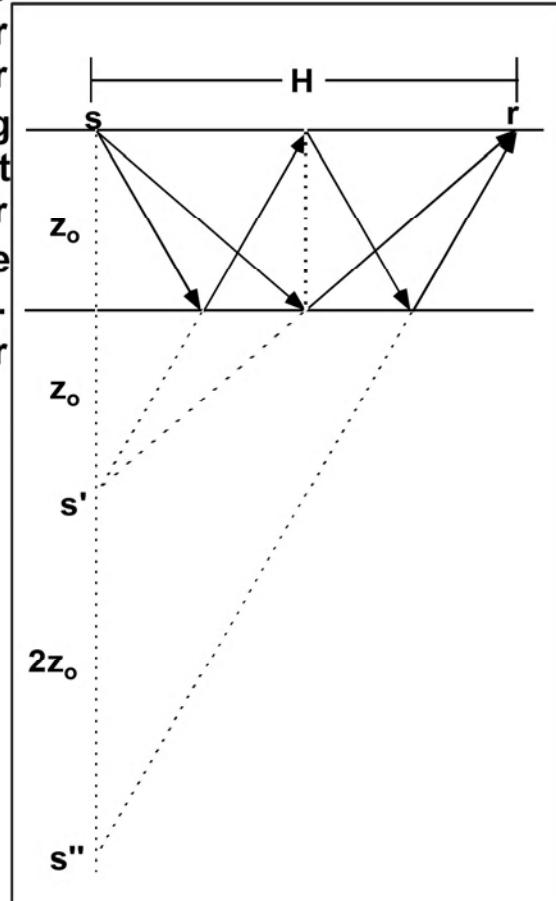
NMO for Multiple Reflections

Consider the constant velocity case of the first order surface related multiple as shown below. The primary is shown arriving at offset h and receiver r while the corresponding multiple arrives at offset h and receiver r after two reflections. We can consider the primary arrival as being from an image source s' at depth $2z_o$ while the first order multiple requires an image source s'' at depth $4z_o$. Accordingly, the traveltime for the multiple is seen to be:

$$t_{\text{mult1}} = \frac{\sqrt{(4z_o)^2 + H^2}}{V}$$

$$\text{or } t_{\text{mult1}}^2 = (2t_o)^2 + \frac{H^2}{V^2}$$

$$\text{where } t_o = \frac{2z_o}{V}$$

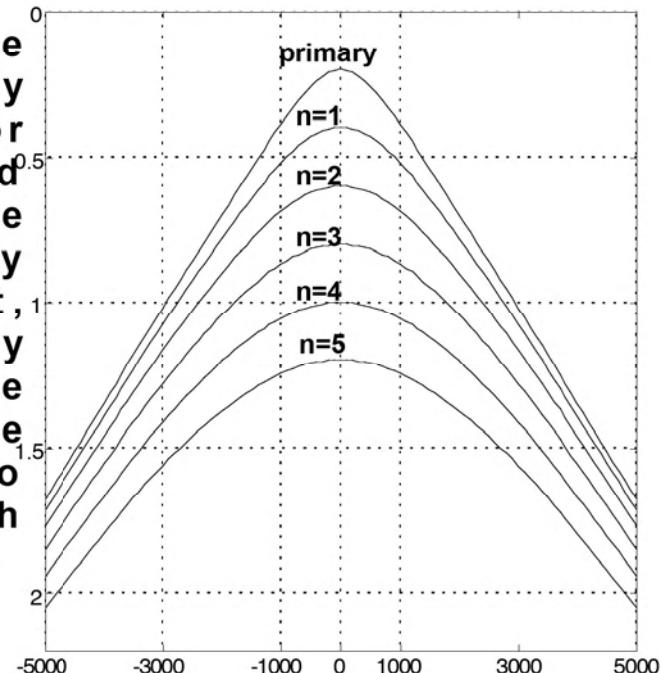


So we see that the multiple has hyperbolic NMO with the same stacking velocity as the primary but a zero offset time that is twice that of the primary. It is not difficult to generalize this to an n bounce multiple to obtain:

$$t_{\text{multn}}^2 = ((n+1)t_o)^2 + \frac{H^2}{V^2} \quad \text{eqn 1 surface coupled multiples when } V=\text{constant.}$$

NMO for Multiple Reflections

Here are the traveltime curves for the primary and first five multiples for the case when $t_0 = .2$ and $v = 3000$. Notice that, while the multiples are perfectly periodic at zero offset, they are not so at any other offset. This is the origin of much of the difficulty in trying to attenuate multiples with predictive deconvolution.

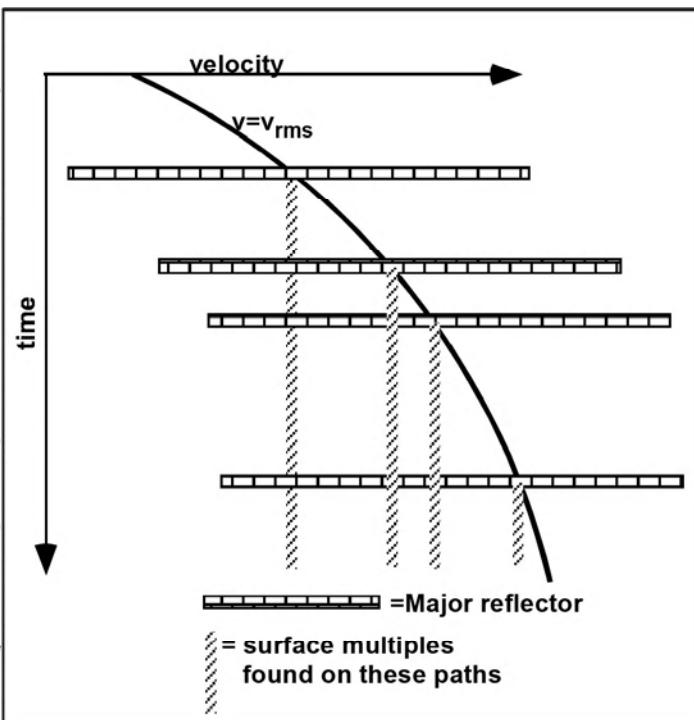


To consider other forms of multiples such as interbeds, we need to generalize our results to a $v(z)$ medium. First, consider the surface multiple. The Dix equation analysis that tells us that v_{nmo} for primaries is v_{rms} equivalently states that we can use v_{rms} to trace an offset ray either one way or two ways through a $v(z)$ medium. Since an n bounce multiple is simply $2n$ one way paths, it follows that the constant velocity result generalizes immediately to:

$$t_{\text{multn}}^2 = ((n+1)t_0)^2 + \frac{h^2}{v_{\text{rms}}^2(t_0)} \quad \text{eqn 2} \quad \begin{matrix} \text{surface coupled} \\ \text{multiples when } v=v(z). \end{matrix}$$

NMO for Multiple Reflections

So we see that surface coupled multiples will always have the same v_{nmo} as the primaries but will occur at progressively later times. In a normal setting where v increases with z , this allows the possibility to distinguish the primary reflections as being those which stack with the fastest velocity at any time. However, in practical settings, it can be difficult to apply this "rule" for several reasons. First, the multiples are often stronger than the primaries they are superimposed on and a stacking velocity analysis may not detect the primary reflections. Second, there are a number of real situations where the stacking velocity can decrease with depth over a short interval.



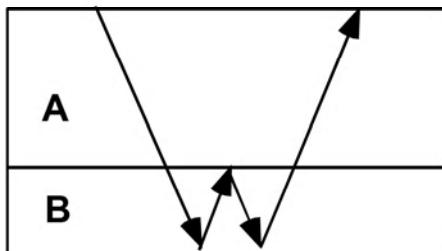
We can use the same techniques to compute v_{nmo} for any given path in a $v(z)$ medium. Recall that we showed that a primary reflection in such a medium has a traveltime with offset curve given by:

$$t_h^2 = t_o^2 + \frac{H^2}{v_{rms}^2(t_o)} + \dots \text{ eqn 3}$$

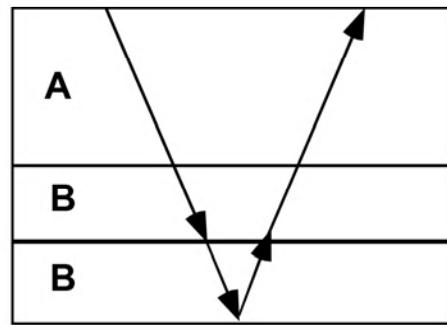
where t_o is the two way vertical traveltime through the layers and v_{rms} is the rms velocity along the two way vertical path.

NMO for Multiple Reflections

Now consider the application of these ideas to the problem of short path multiples. We have noted before that symmetry arguments show that equation 3 can be regarded as giving the slant path one way traveltime through a stack of layers on the down and up paths. Thus it is a general prescription for raytracing through a layered system. Consider the multiple shown on the left below.



Single bounce short path multiple



Equivalent "primary" travelpath

As we see, the multiple occurs in the layered sequence B while sequence A simply refers to the overburden. On the right is an equivalent travelpath for a "primary" which has been formed by simply repeating sequence B. Thus we see that

$$V_{\text{nmo_primary}}^2 = \frac{t_A v_{\text{rms}_A}^2 + t_B v_{\text{rms}_B}^2}{t_A + t_B} \quad t_A \text{ and } t_B \text{ are the one way vertical traveltimes through sequences A and B.}$$

and $V_{\text{nmo_mult 1}}^2 = \frac{t_A v_{\text{rms}_A}^2 + t_B v_{\text{rms}_B}^2 + t_B v_{\text{rms}_B}^2}{t_A + t_B + t_B}$

thus $V_{\text{nmo_mult 1}}^2 = \frac{t_{\text{primary}} V_{\text{nmo_primary}}^2 + t_B v_{\text{rms}_B}^2}{t_{\text{primary}} + t_B} \quad \text{eqn 4}$

where $t_{\text{primary}} = t_A + t_B$.

NMO for Multiple Reflections

The generalization of equation 4 to n bounces is:

$$V_{nmo_multn}^2 = \frac{t_{primary} V_{nmo_primary}^2 + nt_B V_{ms_B}^2}{t_{primary} + nt_B} \quad \text{eqn 5 short path multiples when } v=v(z)$$

As long as the path of the multiple leg in sequence B is short compared to the path through sequence A, we can expect that the multiple v_{nmo} will not differ greatly from that of the primary and that such multiples will appear near the same trajectory in velocity-time space as surface coupled multiples. Also, the multiples will occur at times nt_B later than the primary. However, in a normal setting with v increasing with z , equation 5 also shows that the multiple v_{nmo} will be larger than that of the corresponding primary making it more difficult to distinguish it from other primaries.

CMP Stacking

The process of gathering the data into common midpoint (cmp) gathers and stacking (summing over offset) these has several important properties:

- Random (incoherent) noise is reduced, relative to signal by factor of roughly \sqrt{N} where N is the number of traces in the gather (the "fold").
- The volume of data is reduced by a factor of N making subsequent imaging operations (e.g. migration) much cheaper.
- The data geometry is regularized with the result that imaging operations are more easily formulated.
- The stacked data approximate a simple conceptual model called the zero offset section ("zos") which in turn approximates a normal incidence section ("nis").

Much of prestack processing is dedicated to the construction of a stack with the greatest possible signal bandwidth in both f and k_x . Therefore, examine where the bandwidth of the stack comes from. Consider the representation of the prestack data as an inverse Fourier transform over midpoint and offset wavenumbers:

$$\Psi(x, h, t) = \iint_{-\infty}^{\infty} \phi(k_x, k_h, t) e^{ik_x x + ik_h h} dk_x dk_h \quad \text{eqn 1}$$

Assuming that this expression represents nmo corrected data with statics applied, then we represent the cmp stack as:

$$\Psi_{\text{stk}}(x, t) = \int_{-\infty}^{\infty} \Psi(x, h, t) dh \quad \text{eqn 2}$$

CMP Stacking

Inserting equation 1 into 2 and interchanging the order of integration gives:

$$\Psi_{\text{stk}}(x, t) = \iint_{-\infty}^{\infty} \phi(k_x, k_h, t) e^{ik_x x} \left[\int_{-\infty}^{\infty} e^{ik_h h} dh \right] dk_x dk_h$$

from the definition of a delta function as an inverse Fourier transform, we recognize the term in brackets as a delta function:

$$\Psi_{\text{stk}}(x, t) = \iint_{-\infty}^{\infty} \phi(k_x, k_h, t) e^{ik_x x} [2\pi\delta(k_h)] dk_x dk_h$$

Thus $\Psi_{\text{stk}}(x, t) = 2\pi \int_{-\infty}^{\infty} \phi(k_x, 0, t) e^{ik_x x} dk_x$ eqn 3

Equation 3 proves that the cmp stack passes only the zero k_h wavenumber. This means that it is an extremely effective "f-k" filter when applied to the cmp gathers and almost any "f-k" process applied to those gathers prior to stacking them will fail to improve the stack. It follows, as a corollary, that "f-k" processes applied to source or receiver gathers can significantly alter the stack. Furthermore, more detailed spectral analysis shows that the source and receiver wavenumbers are related to the midpoint and offset wavenumbers through:

$$k_x = k_s + k_r \quad \text{and} \quad k_h = \frac{1}{2}(k_r - k_s)$$

eqns 4

or $k_s = \frac{1}{2}k_x - k_h$ and $k_r = \frac{1}{2}k_x + k_h$

CMP Stacking

These results show that the midpoint wavenumbers are composed equally from the source and receiver wavenumbers. Thus, if an "f-k" filter or similar process is applied to the source gathers with positive results on the stack, it follows that further improvement will be obtained by applying the same filter to the receiver gathers.

As for the temporal frequency spectrum of the stack, it is affected mainly by the suppression of unaligned events in the cmp gathers. Thus, not only does noise tend to stack out, but slight signal misalignment, due to unresolved statics or residual nmo, will tend to reduce the signal band as well.

Post Stack Considerations

Even the most careful and modern seismic processing flow will usually result in a stack that needs further work to approach the ultimate goal of band limited reflectivity. Some major points here are:

- Surface waves and other coherent and random noise will usually result in uncontrolled phase rotations in minimum phase decon. Thus, even if the reflection wavelet is perfectly minimum phase, the deconvolution will be less than perfect.
- If the stratigraphic section has regions of low Q, then the resulting non stationary signal band will further destabilize the deconvolution.
- If Vibroseis has been used, it is unlikely that a perfect deconvolution can be achieved because this is highly dependent on the assumption that the embedded wavelet is the Klauder wavelet. This is an unproven and controversial assumption.
- Even if the prestack deconvolution was aggressive and fully whitened the spectrum, stacking will "dewhitener" it. This is because the random noise levels in the data are strong functions of time and frequency. Since noise "stacks out" relative to signal with a factor of the square root of the fold, the resulting spectrum will be nonwhite.
- Short path multiples may survive the prestack processing and need to be addressed by a post stack deconvolution.
- If AGC or TVSW was used in the prestack processing, then relative amplitude information may be highly distorted.

Post Stack Considerations

Thus we are left with a model for a stack that is largely convolutional with an embedded wavelet with unknown "residual phase" and a non-white amplitude spectrum. Post stack processing techniques which are relevant to improve the situation include:

- **minimum phase, stationary deconvolution:** This is useful to fine tune a target zone which is thought to contain short path multiples.
- **zero phase, stationary deconvolution:** A very safe conservative process used to whiten a target zone without any phase effects.
- **zero phase, non-stationary deconvolution (TVSW):** This is a popular and extremely powerful process. When run iteratively with f-x spatial prediction it can achieve starting improvements in data quality. However, its use (either pre or poststack) effectively ends all possibilities of further minimum phase deconvolutions.
- **minimum phase, non-stationary deconvolution:** Available from specialized processors, this acts like a data dependent inverse Q filter and can achieve remarkable results.
- **f-x spatial prediction:** Though not a deconvolution algorithm, its use in improving the lateral coherence of reflectors is well documented. Used in conjunction with deconvolution algorithms, it moves the data closer to a high resolution reflectivity image.
- **wavelet processing:** Technically speaking, any deconvolution technique is a wavelet processing method; however, the term is usually used to refer either to advanced statistical methods (usually not minimum phase) or to well log matching techniques.
- **migration:** Ya gotta do it.

Post Stack Considerations

The last two need more discussion. First wavelet processing. The two approaches mentioned both have their proponents though its hard to argue against comparison with well control. The derivation of wavelets by matching to wells is a slippery slope that must be trod with caution. This is because the theory of Wiener match filters can match seismic data from Alberta to a Kenyan elephant. It is best to use highly constrained methods which do not allow the full use of the theory of match filters.

A simple example is the method of constant phase rotations. In this case, a simple seismogram with the bandwidth of the data is created and a number of different constant phase rotations are generated. These are compared with the seismic data near the well and the best matching phase is chosen. High quality data with competent processing often works well in this case.

Attempting to determine the details of the residual phase spectrum is a highly non-unique process and subject to a lot of uncertainty. A well known result from geophysical inverse theory is that averages of properties might be very well determined even when the properties themselves are not. Thus the estimation of constant phase rotations might be quite robust even when the determination of residual phase is unstable.

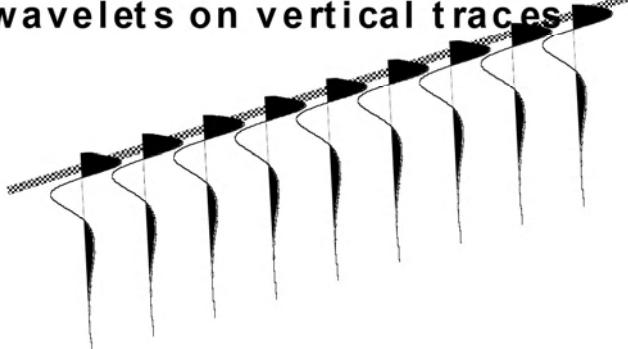
Two general rules of thumb for wavelet processing:

- **ALWAYS** do it.
- Be suspicious of exceptional well ties produced by "wild and ugly" wavelets.

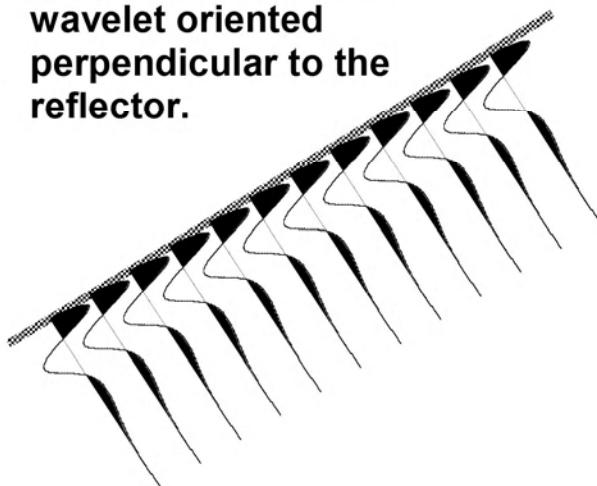
Post Stack Considerations

Migration should generally be done as the last step in the processing sequence. Theoretically, migrated data no longer fits the standard 1-D convolutional model so filters and wavelet processing should be done first. This is because:

The stacked section has wavelets on vertical traces



Migration leaves the wavelet oriented perpendicular to the reflector.



As a result, the relationship between wavelet, reflectivity, and seismic data is no longer a simple 1-D convolution after migration. (Though it is a 2-D or 3-D, space variant convolution.) This means that filtering and deconvolution definitely don't commute with migration and they should usually be done before.

This does leave a bit of a dilemma for structural data. The data must be migrated to the position of the wells to compare them; however, the wavelet estimation and processing should be done before the data is migrated.

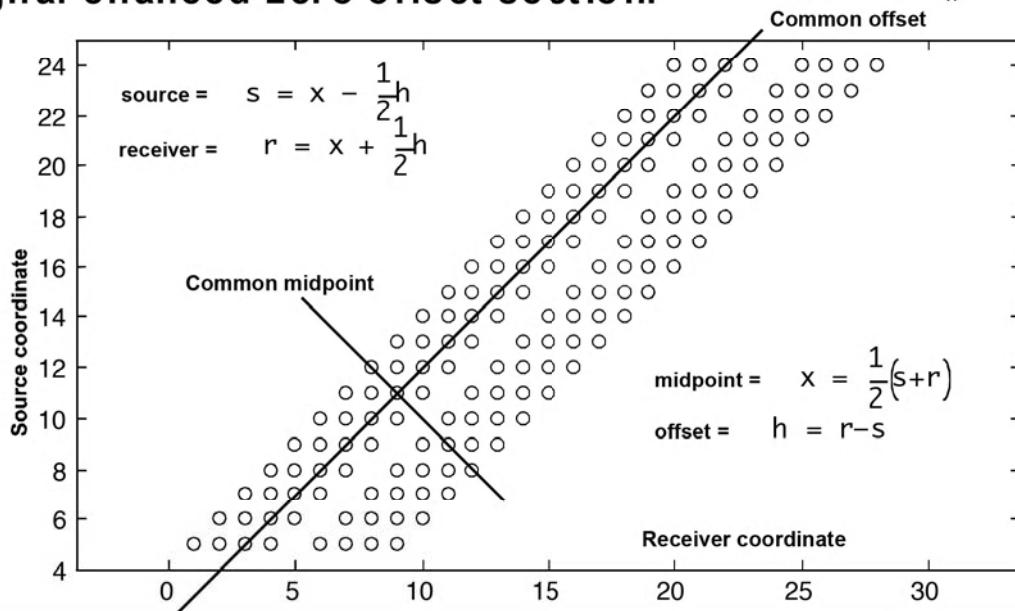
ZOS: A Model for the CMP Stack

A discussion of post stack migration greatly benefits from a firm theoretical model of the cmp stack. We will first examine the "zero offset section" (or ZOS) model which is sufficient for many migration concepts. Later, we will present the "exploding reflector model" which is used to justify most wave equation migration methods.

The ZOS model asserts that the prestack processes of offset dependent gain, normal moveout removal, and stack estimated a signal enhanced version of what we would have recorded had their been a single, coincident source/receiver pair at each cmp.

Steps in the estimation of the signal enhanced ZOS:

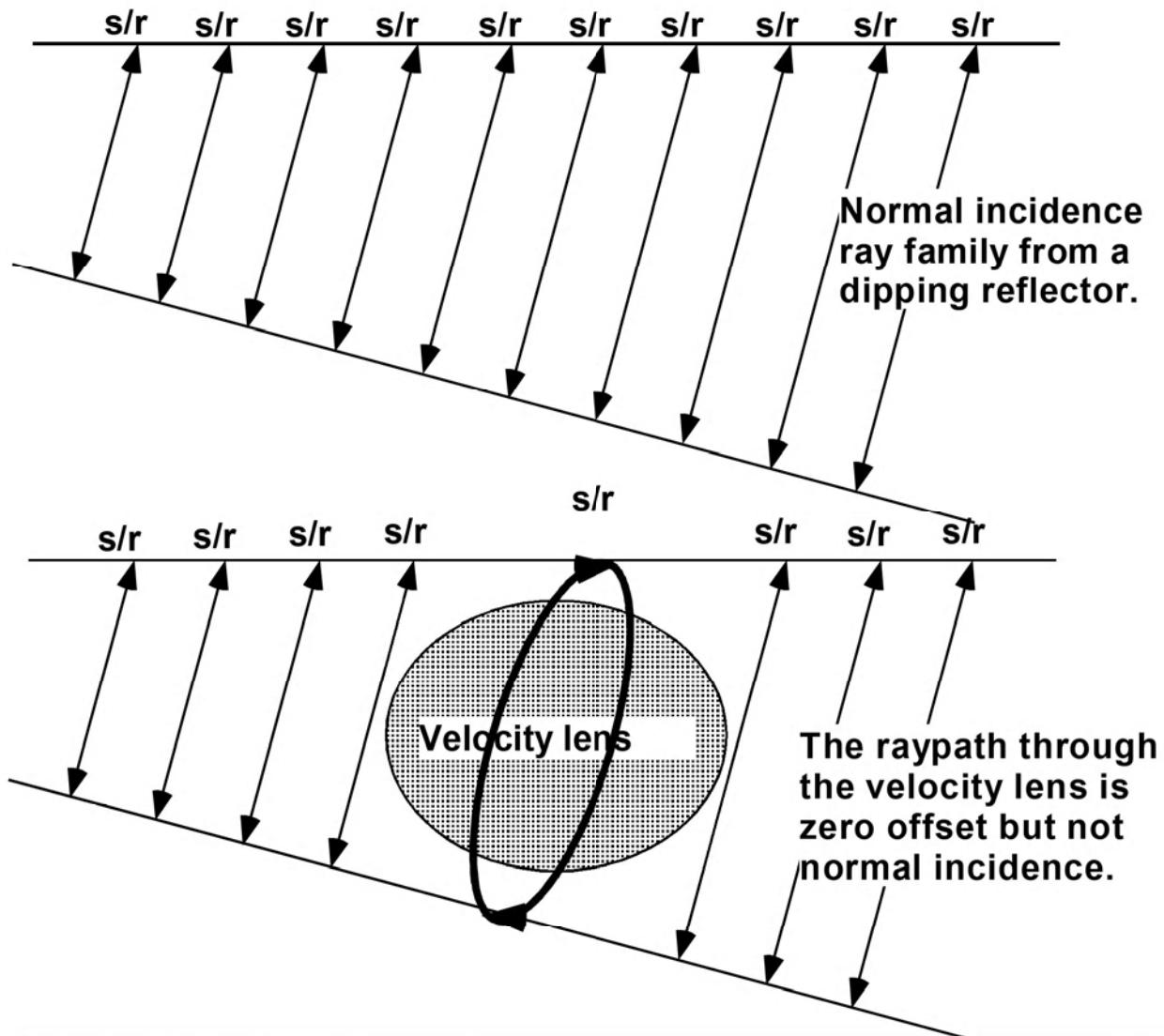
- Sort the data into (x, h) coordinates: $\Psi(s, r, t) \Rightarrow \Psi(x, h, t)$
- Map offset traveltimes to zero offset: $t_o = \sqrt{t^2 - \frac{h^2}{v^2}}$
- Apply NMOR correction: $\Psi(x, h, t) \Rightarrow \Psi(x, h, t_o)$
- Sum (stack) all traces with a "common" midpoint to estimate a signal ehanced zero offset section. $\Psi_o(x, t_o) = \sum_h \Psi(x, h, t_o)$



ZOS: A Model for the CMP Stack

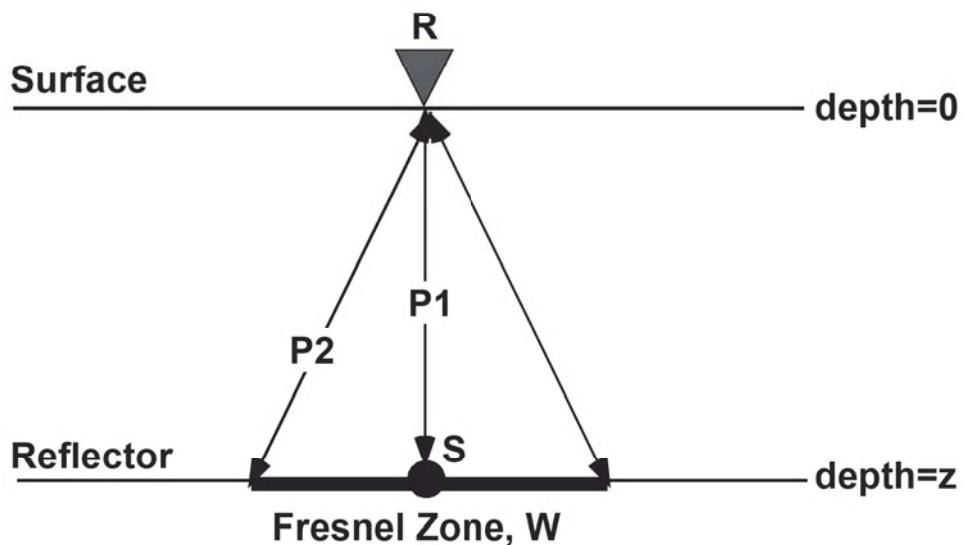
We say that the ZOS is signal enhanced because the stacking process has been designed to select against multiples.

The simplest (and probably dominant) raypath which returns energy to its source is called the normal incidence raypath. We will further assume that all energy on the ZOS may be modeled by normal incidence raypaths. The diagrams below shows a family of normal incidence rays from a dipping reflector and a "pathological" zero offset ray which is not normal incidence and is thus excluded from our model.



Fresnel Zones

Consider the zero offset scattering of waves from a point in the subsurface. High frequency raytracing (e.g. Snell's Law) predicts that the receiver, R, will only record information from the infinitesimal point, S. However, this is only a very gross approximation to the behavior of real waves. If a vertically travelling, monochromatic wave, of frequency f , strikes the reflector, it is actually "backscattered" in all directions and from all points.



The receiver at R will respond to the total backscattered wavefield whose amplitude will be the sum of all individually scattered wavefields. Since waves interfere both constructively and destructively, we seek a measure of the width, W, for energy arriving at R is "in phase". This is usually called the Fresnel Zone and is defined by the requirement that the difference in lengths of the paths P1 and P2 be one-quarter wavelength (assuming constant velocity). Thus:

$$\sqrt{z^2 + \left(\frac{W}{2}\right)^2} - z = \frac{\lambda}{4}$$

or, solved for W,

$$W = 2\sqrt{\left(\frac{\lambda}{4} + z\right)^2 - z^2}$$

Fresnel Zones

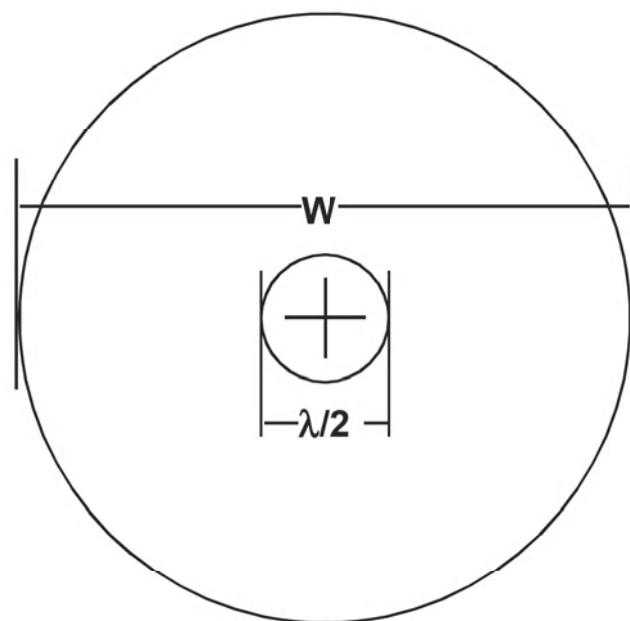
Other, alternate expressions for W are:

$$W = \sqrt{\frac{\lambda^2}{4} + 2z\lambda} = \sqrt{\frac{v^2}{4f^2} + 2z\frac{v}{f}}$$

$$W \approx \sqrt{2z\lambda} = \sqrt{2z\frac{v}{f}} \quad \text{for } z \gg \lambda$$

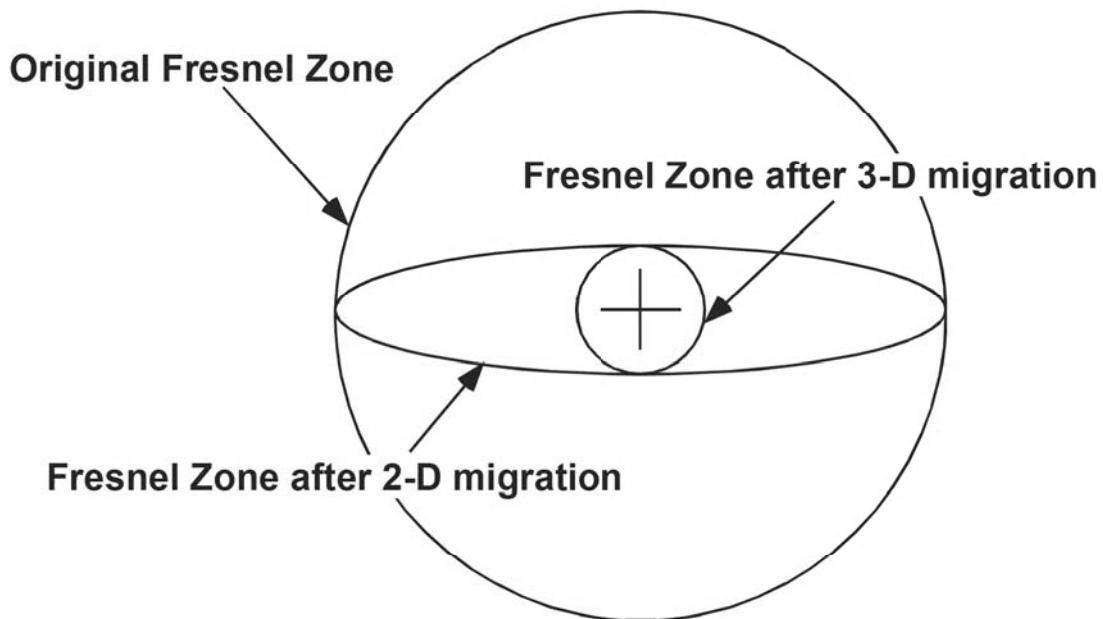
Migration, which can be conceptualized as lowering the receiver to the point S, can be said to shrink the Fresnel zone to its theoretical minimum of $\lambda/2$. (Note that it does not shrink to zero as is sometimes erroneously stated when the approximate expression for W is used.)

In 3-D, the Fresnel zone is a disk which defines the inherently unfocused nature of stacked data.

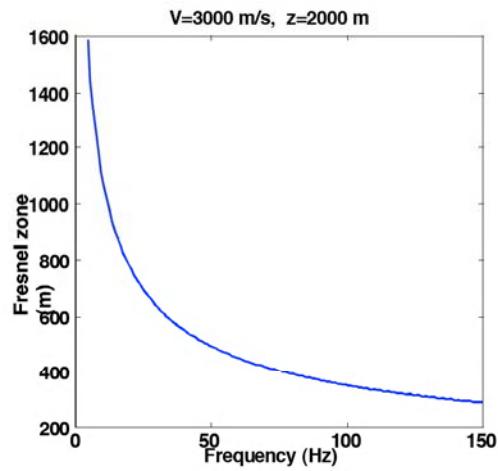
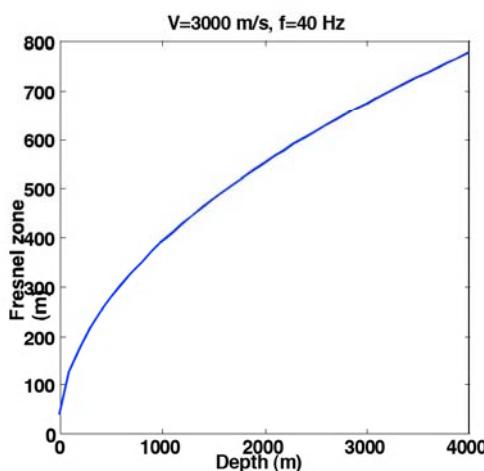


Fresnel Zones

Real earth seismic data is always 3-D regardless of whether the acquisition geometry is 2-D or 3-D. Thus, when 2-D migration is run, the Fresnel zone is only collapsed in the inline direction.



This is one of the most compelling justifications for 3-D imaging techniques, even in the case of sedimentary basins with flat-lying structures. A 2-D migration of a seismic line must be considered as giving the average reflectivity over a Fresnel zone width in the cross line direction.



Methods of Seismic Data Processing

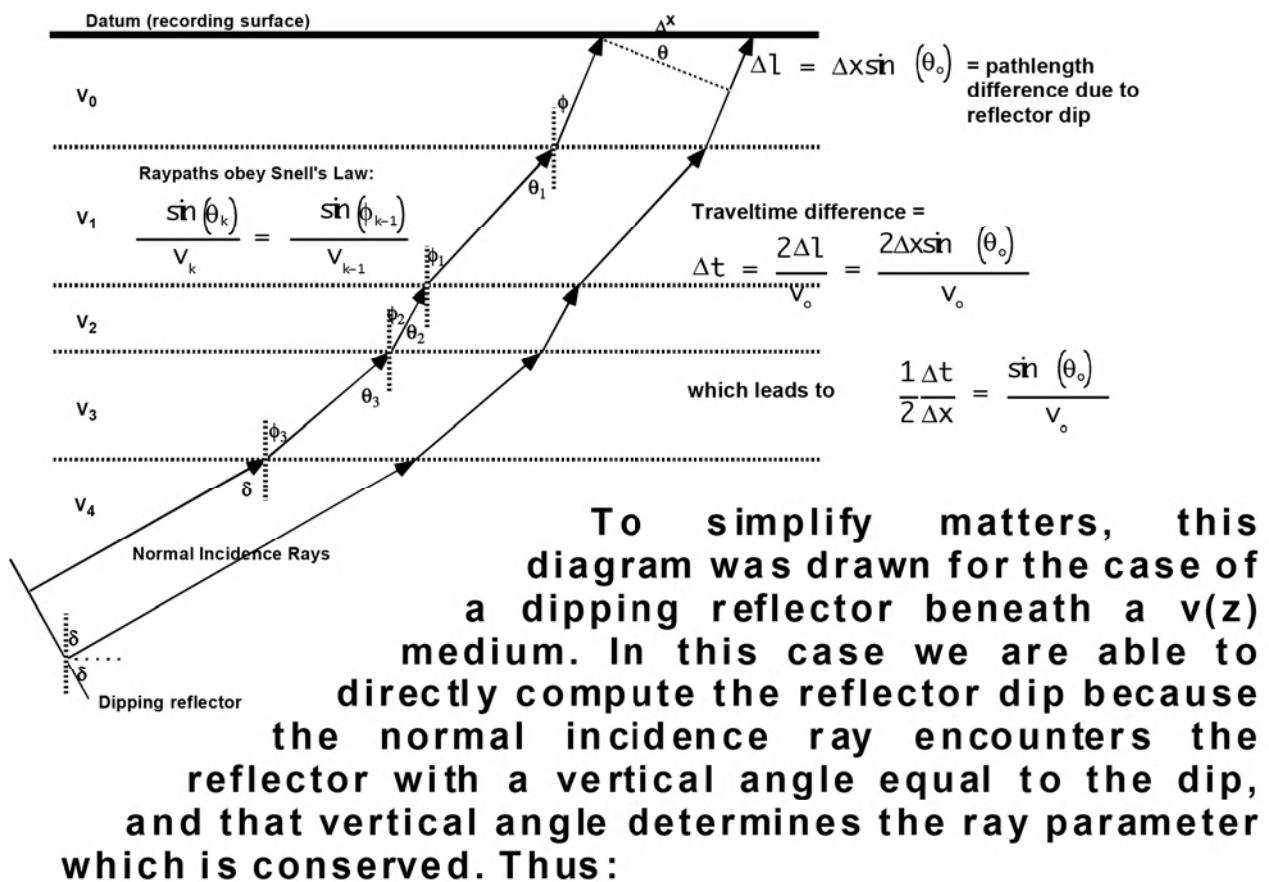
Lecture Notes Geophysics 557

Chapter 8 Migration Concepts

Raytrace Migration of Normal Incidence Seismograms

We are now in a position to develop our first example of a migration technique. It should be obvious by now that the cmp stack display, in which all energy is plotted in time beneath its surface arrival point, does not display reflections in their correct spatial positions. The mapping of reflections from (x, t) to (x_m, z) is a general definition of the migration of seismic data.

One of the oldest, and most intuitive, migration technique is based on normal incidence raytracing. In the diagram below, we illustrate the link between reflector dip and the measurable traveltime gradient on a ZOS.



$$\sin(\delta) = \frac{v_4 \Delta t}{2 \Delta x}$$

Raytrace Migration of Normal Incidence Seismograms

In more complex cases, where the velocity above the reflector is a general function of (x, y, z) , the link between reflector dip and traveltime gradient is still there though it is not so direct. To deduce the dip, and the position of the reflector in the general case, we use the normal incidence raytrace migration algorithm:

Definition: A pick is a triplet of values $(x, t_o, \Delta t/\Delta x)$ measured from a ZOS having the meaning:

x = inline coordinate at which the measurement is made

t_o = normal incidence traveltime (2-way) measured at x .

$\Delta t/\Delta x$ = Normal incidence traveltime gradient (time dip) measured at x and t_o

To perform the raytrace migration, the following materials are necessary:

i) A set of picks to be migrated: $P_{ij} = \{x_i, t_{oij}, \Delta t/\Delta x_{ij}\}$

ii) A velocity model: $v(x, y, z)$

iii) A calculator, compass, protractor, straightedge, and a brain

Or, if one or more of the elements of iii) is unavailable:

iv) A computer program

Raytrace Migration of Normal Incidence Seismograms

Normal incidence raytrace migration algorithm:

For all locations x_i , $i=1,2 \dots nlocs$, and for all picks at x_i , $(t_{oij}, \Delta t / \Delta x_{ij})$, $j=1,2,\dots npicks$, then do the following six steps:

Step 1) Determine the emergence angle of the ij^{th} ray:

$$\sin(\theta_{oij}) = \frac{v_o(x_i) \Delta t}{2 \Delta x_{ij}}$$

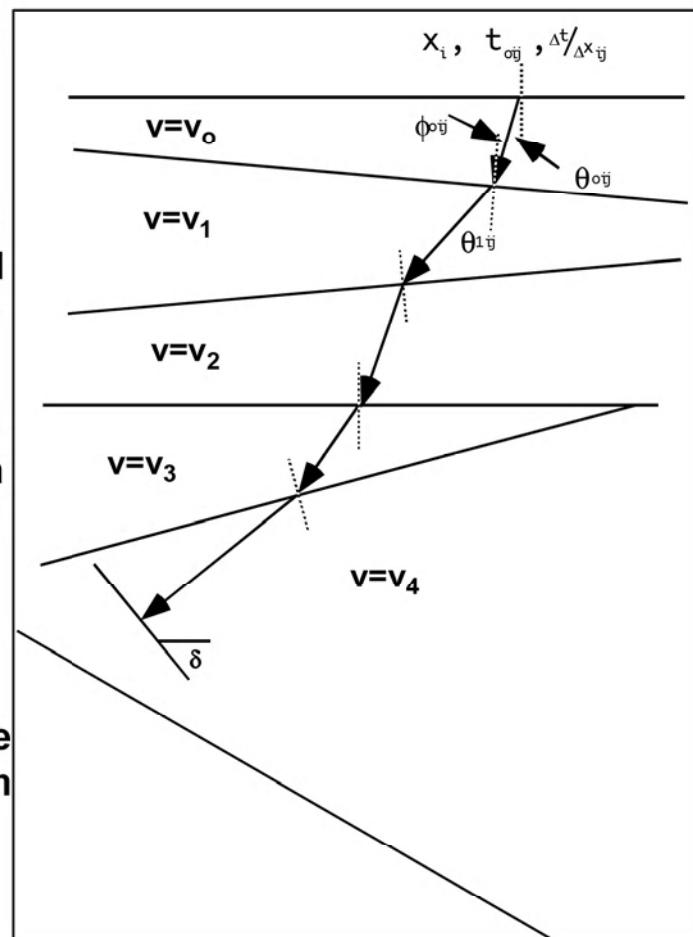
Step 2) Project ray down to first interface and determine incident angle ϕ_{oij} .

Step 3) Determine the length of the raypath in the layer, l_o , and compute layer traveltime:

$$\delta t_{ij} = \frac{l_{oij}}{v_o}$$

Step 4) Determine the refraction angle, θ_1 , from Snell's law:

$$\sin(\theta_{1ij}) = \frac{v_1}{v_o} \sin(\phi_{oij})$$



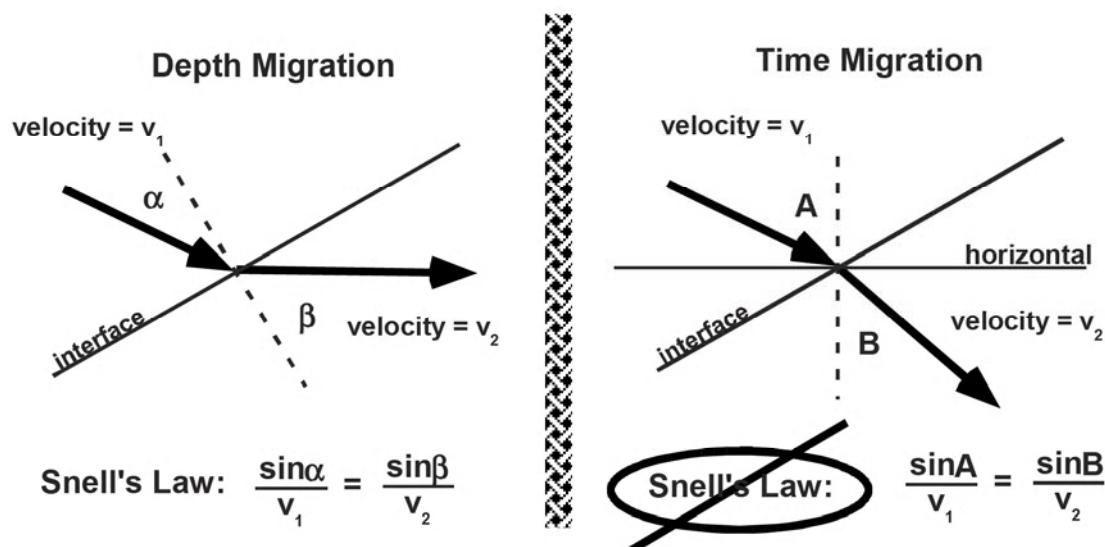
Step 5) Repeat steps 2->4 for each velocity layer interface until the accumulated ray traveltime is half of t_{oij} .

$$t_{rayij} = \sum_{k=1}^n \frac{l_{kij}}{v} = \frac{t_{oij}}{2}$$

Step 6) When the accumulated traveltime is half the measured t_{oij} , then stop raytracing and draw in the reflector perpendicular to the raypath at this point.

Time and Depth Migrations, A First Look

Depth Migration downward continues data along raypaths which obey Snell's Law regardless of velocity complexity. Time migration always assumes a locally horizontal velocity interface and hence violates Snell's law when lateral gradients are present.



Questions:

- Does a "successful" time migration confirm the validity of the velocity model? Does a successful depth migration? If you said no to both, which provides the stronger constraint?
- Which type of migration is normal incidence raytrace migration?
- Time migration is biased to leave unchanged a certain class of rays. What are they and what are some geological reflectors that can produce such rays as normal incidence reflections?
- Which type of migration do you expect to be more "forgiving" of velocity errors? Is this a good thing?

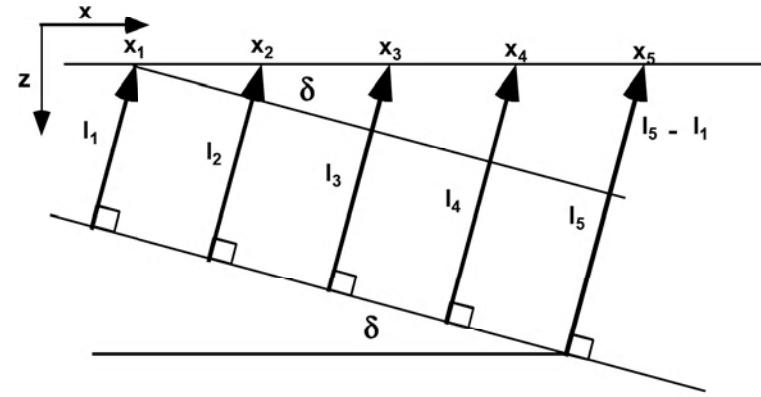
For $v=v(z)$ there is no meaningful difference between time and depth migration

Elementary Constant Velocity Migration

-1-

Consider a ZOS image of a dipping reflector overlain by a constant velocity. The dip is related to the raypath lengths by:

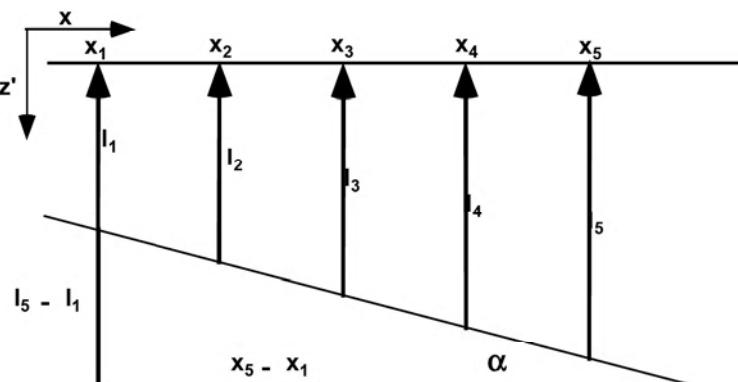
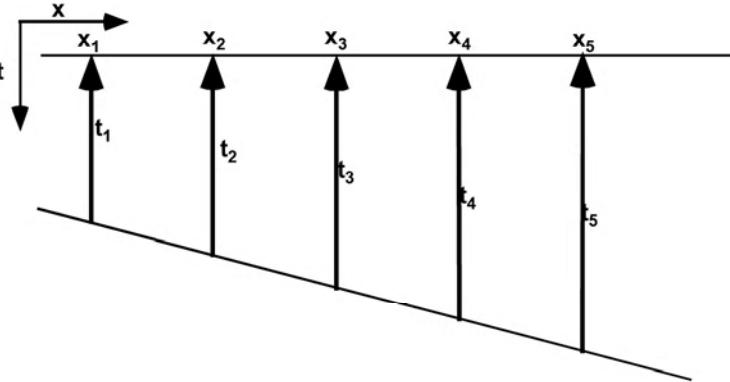
$$\sin(\delta) = \frac{l_5 - l_1}{x_5 - x_1}$$



The time display of this data forms what is called the "zos image". The resultant time dip is:

$$\frac{\Delta t}{\Delta x} = \frac{t_5 - t_1}{x_5 - x_1}$$

We can plot the zos image as an apparent depth section by stretching the arrival times according to $z' = vt/2$. This results in an apparent dip, α , which is related to the real dip through:

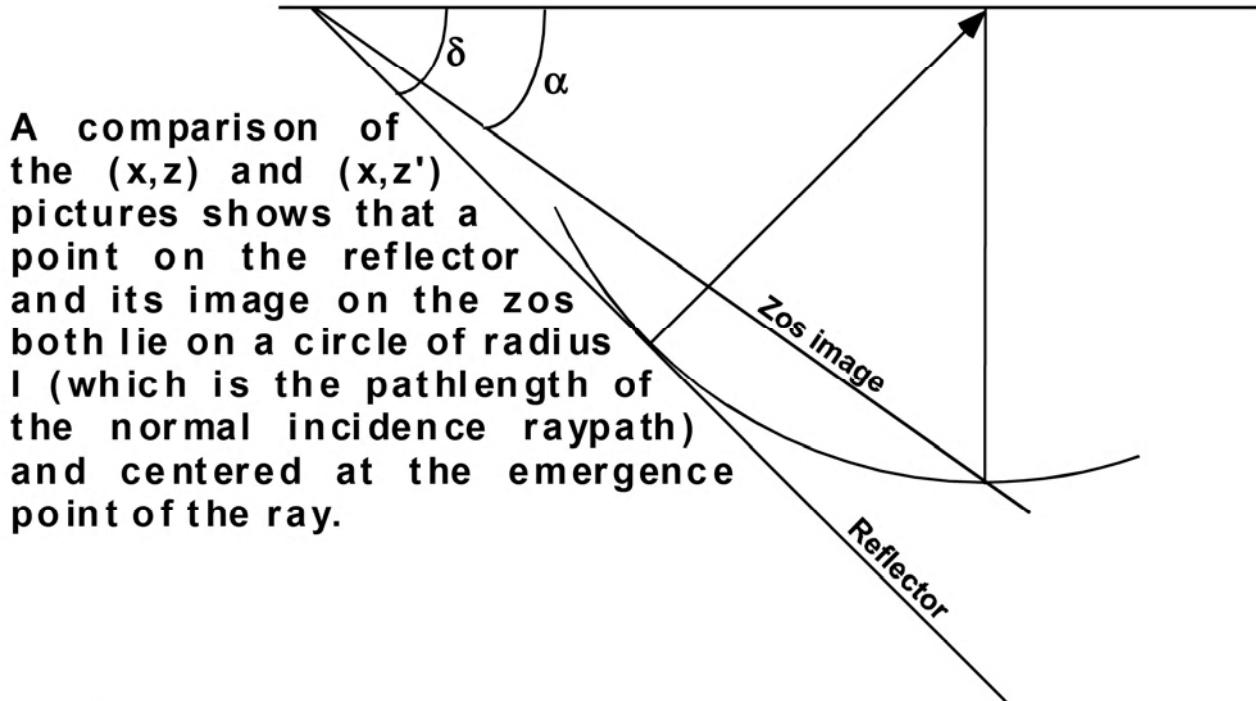


$$\tan(\alpha) = \frac{l_5 - l_1}{x_5 - x_1}$$

$\sin(\delta) = \tan(\alpha)$

This is called the "migrator's formula"

Elementary Constant Velocity Migration



A comparison of the (x, z) and (x, z') pictures shows that a point on the reflector and its image on the zos both lie on a circle of radius l (which is the pathlength of the normal incidence raypath) and centered at the emergence point of the ray.

Note that:

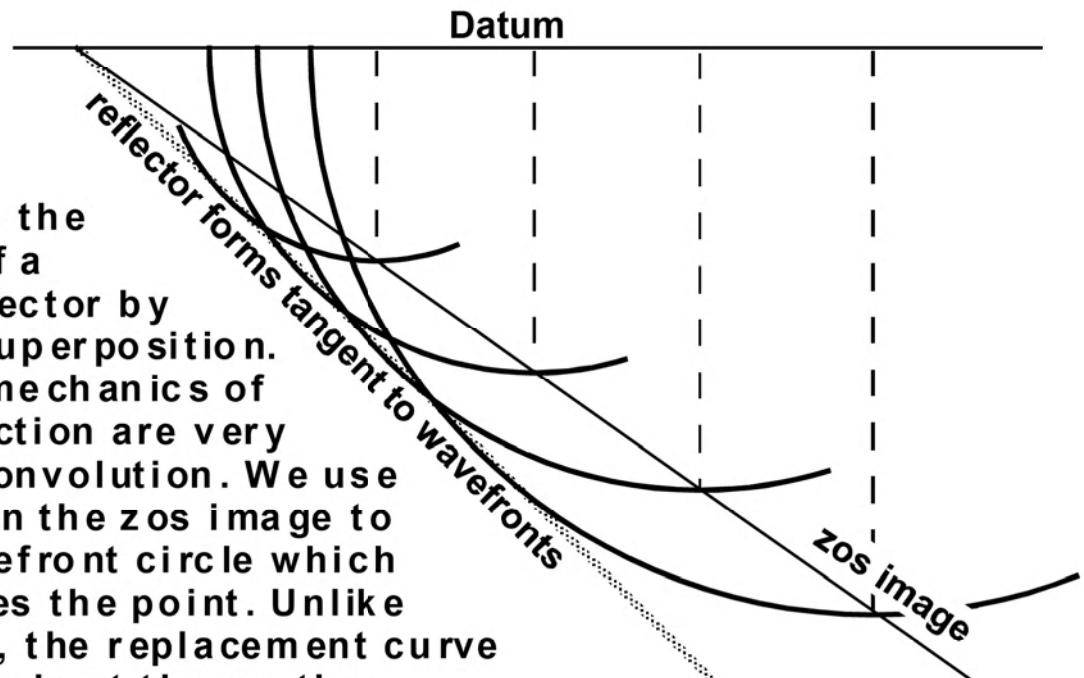
- The zos image lies at the nadir (lowest point) of the circle.
- The reflection is tangent to the circle.
- The zos image and the reflector agree on the datum (recording plane).

These observations form the basis of one of the earliest "wavefront" migration methods which is exact for constant velocity:

- i) Stretch the zos image from t to z'
- ii) Replace each point in the (x, z') picture with a "wavefront circle" of radius z' .
- iii) The superposition of all such wavefronts will form the desired migrated depth section.

Elementary Constant Velocity Migration

Here we see the migration of a dipping reflector by wavefront superposition. The actual mechanics of the construction are very similar to convolution. We use each point in the zos image to scale a wavefront circle which then replaces the point. Unlike convolution, the replacement curve varies throughout the section.

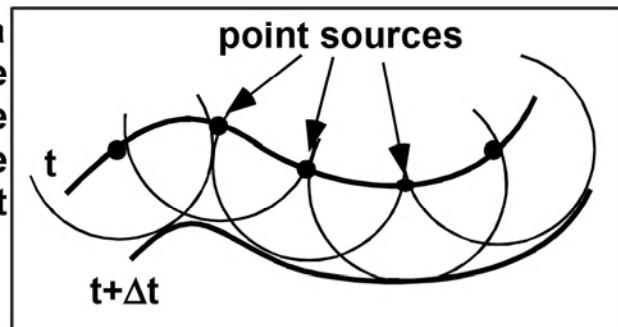


So we deduce that the "impulse response" of migration is a wavefront circle. We can regard the wavefront circles as defining "curves of equal probability". That is, considered in complete isolation, a single arrival at time t_o at x can only be said to lie somewhere along the locus of points whose traveltime from x is t_o . We rely upon the constructive interference of a great many wavefronts to form the migrated image.

What would be the result of the above migration if one of the traces is unbalanced with respect to the others such that its amplitude is 10^6 times too large?

Huygens Principle and Point Diffractors

Christian Huygens was an early physicist and astronomer (contemporary with Newton) who made a number of advances in the understanding of waves. Most notable was his principle that, if one knows the position of a wavefront at time t , its position at $t+\Delta t$ can be computed by considering it to be the superposition of many small point sources of waves. For the seismic problem, we can adapt Huygens principle to argue that the response of a continuous reflector can be considered as the superposition of the responses of many point scatterers or point diffractors.

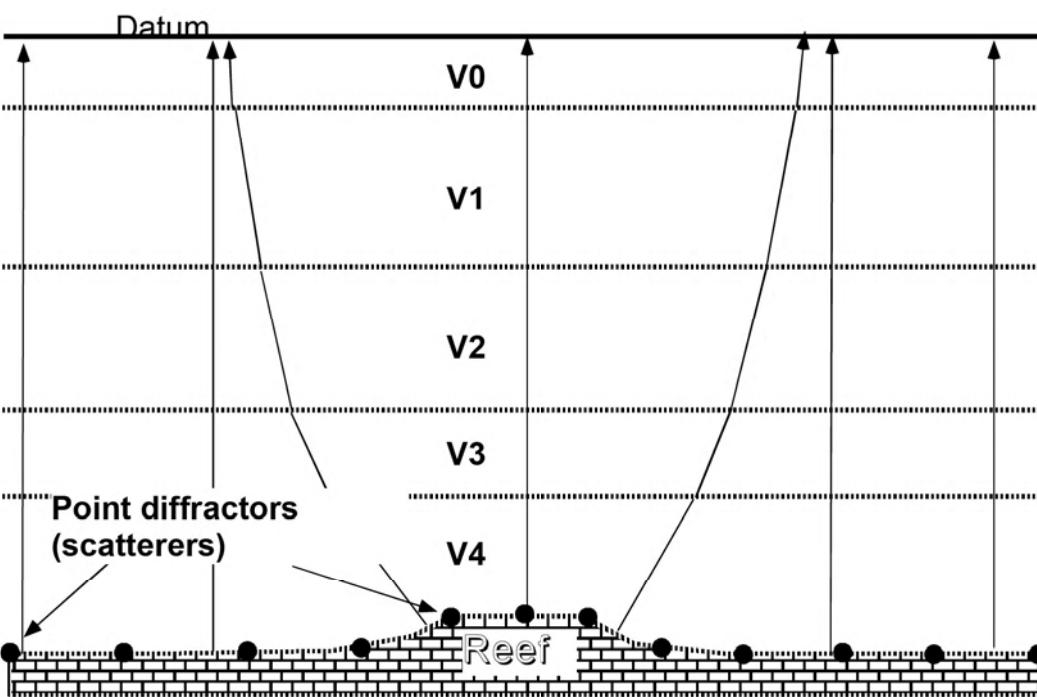


Huygens (paraphrased):

Response of the
continuous structure

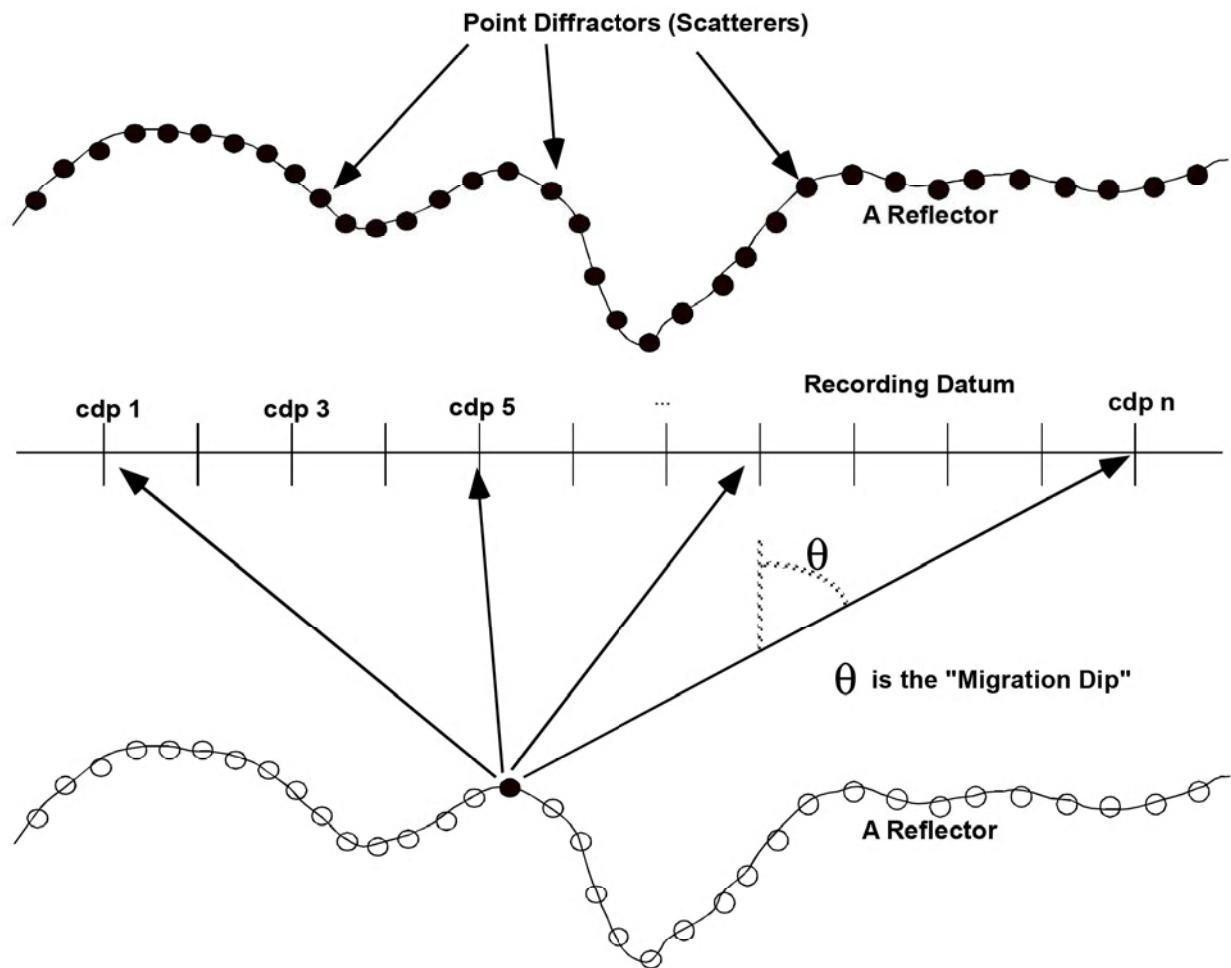


Superposition of a suitable set
of point diffractors



Huygens Principle and Point Diffractors

The seismic response of an arbitrary reflector can be considered as the linear superposition of the responses of its constituent point diffractors.

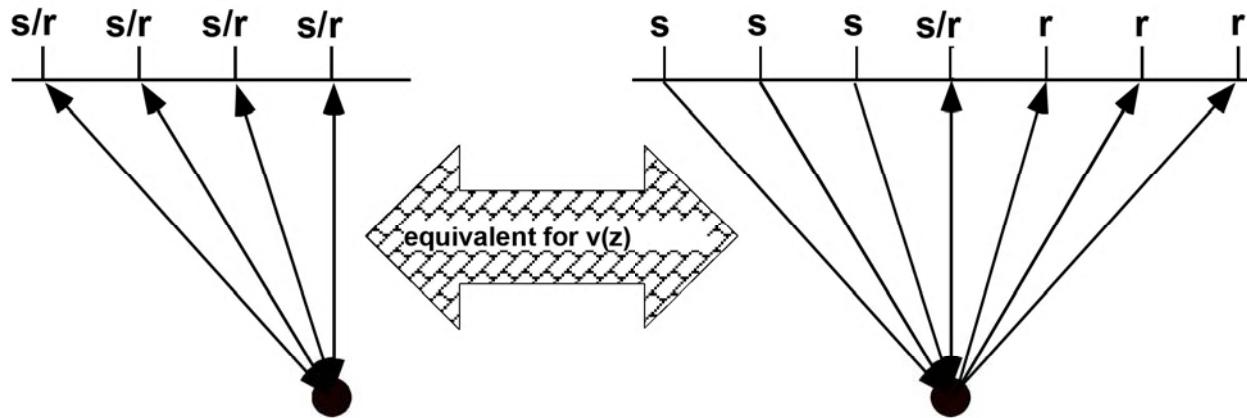


The complete imaging of any point diffractor requires the ability to migrate all scattered energy regardless of raypath angle. This angle is usually a program parameter referred to as "Migration Dip".

- A steep dip algorithm is a high resolution algorithm
- This concept is independent of geologic dip

Huygens Principle and Point Diffractors

The traveltime curve of a point diffractor for the zero offset section has a raypath geometry nearly identical to the NMO problem.

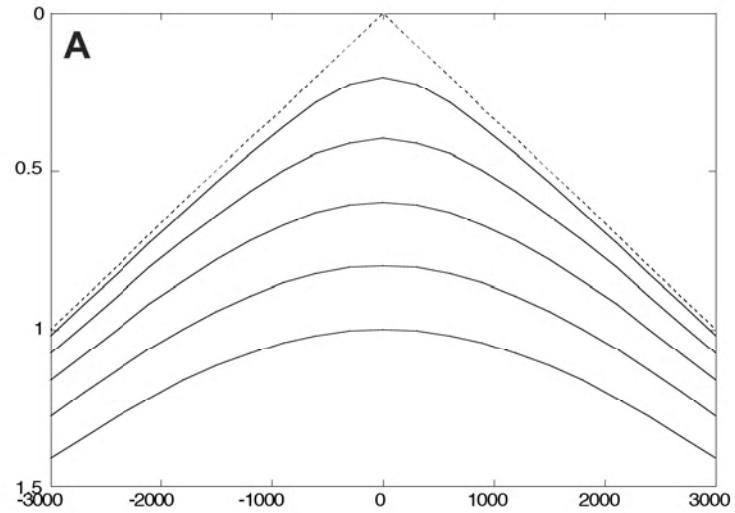


On the left we see the raypaths needed to compute half of the response of a point diffractor assuming coincident source and receiver pairs. On the right are the raypaths needed to compute the NMO curve for a reflection point at the point diffractor. As long as there are no lateral velocity variations, these geometries give equivalent traveltimes. Thus we immediately conclude:

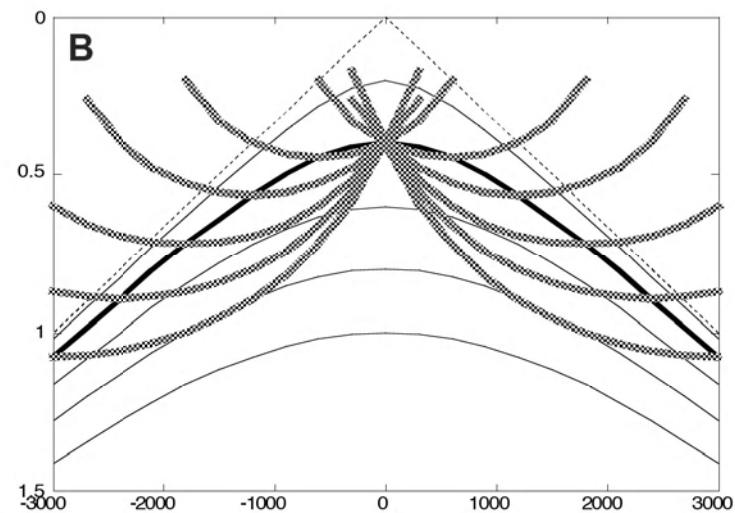
- For a constant velocity, the traveltime curve of a point diffractor is a constant velocity hyperbola whose apex is at the two way vertical traveltime to the diffractor.
- For $v=v(z)$, the traveltime curve of a point diffractor is an approximate hyperbola characterized by $v_{rms}(t_o)$ with apex at t_o , which is the vertical traveltime to the point diffractor.

Huygens Principle and Point Diffractors

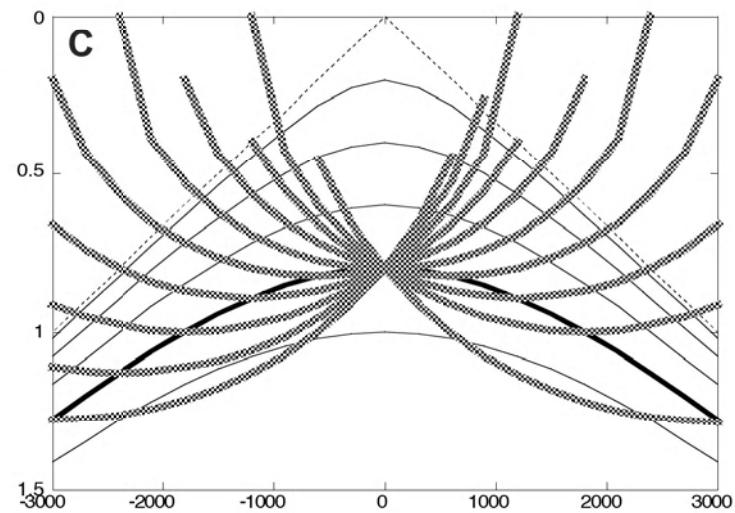
A) Here is a "diffraction chart" which shows the traveltimes for five different point diffractors in a constant velocity medium.



B) In this panel, we have migrated the second hyperbola by replacing its points with the appropriate constant velocity wavefront. Notice how the wavefronts focus at the apex of the hyperbola.

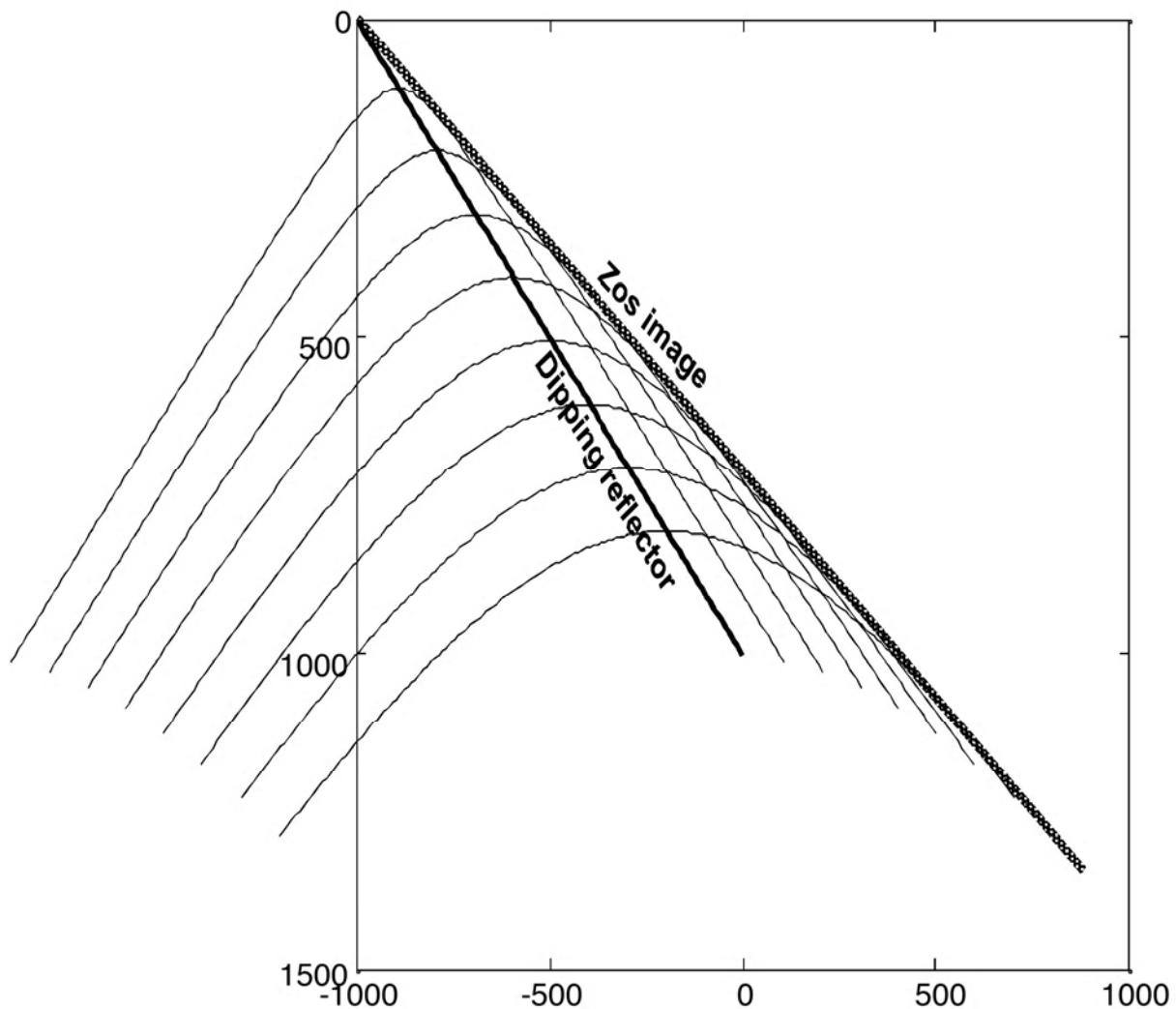


C) Here we migrate the fourth hyperbola by the same method. It is evident that this method will focus all five hyperbolae simultaneously. From Huygens principle, we conclude that this method will exactly migrate any shape reflector.



Huygens Principle and Point Diffractors

We can use the diffraction curves to do the inverse of migration which is forward modeling. If we replace each point on a dipping reflector with an appropriate hyperbola from the diffraction chart, the zos image is formed above the reflector where the diffraction tails constructively interfere.



The Exploding Reflector Model

We have developed a number of tools and concepts for analyzing and manipulating wavefields, most notably the F-K transform, and it is natural to think of applying them to stacked data. However, what kind of wavefield is the cmp stack? We have seen that a zos is a good approximation to the cmp stack, but a zos is fundamentally the composite of many individual experiments and their resultant wavefields. There is no single physical experiment which could record a zos and so a zos cannot be a single physical wavefield.

There is a useful thought experiment, called the exploding reflector model (ERM), which does yield something very similar to a zos and serves as the basis for most of our post stack wavefield migration methods. As shown below in A, we imagine a geology identical to the real earth in all respects except that velocities are halved and the reflectors are lined with explosives. Receivers are placed at the cmp locations and at $t=0$ the explosives are detonated. This creates a wavefield morphologically identical to the geology at the instant of explosion.

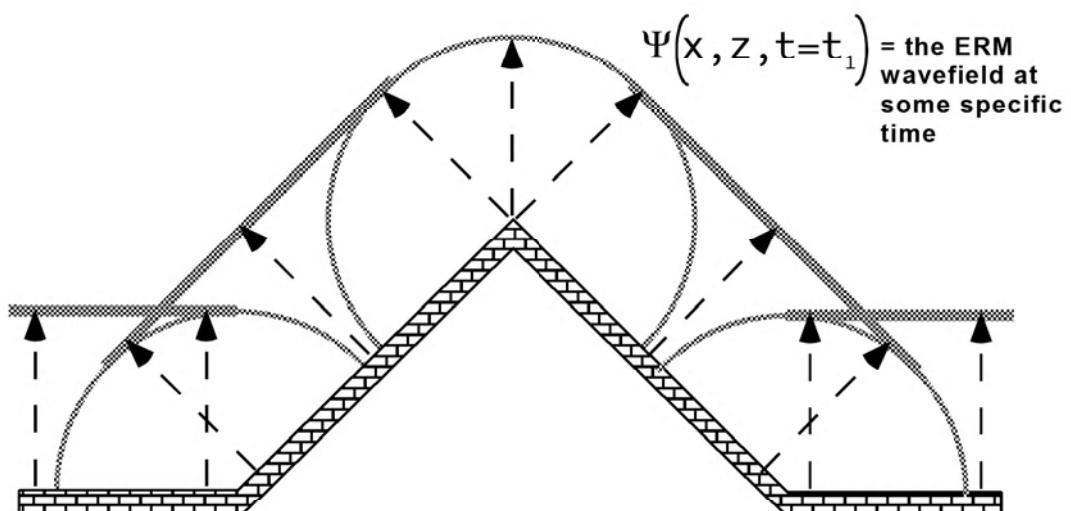
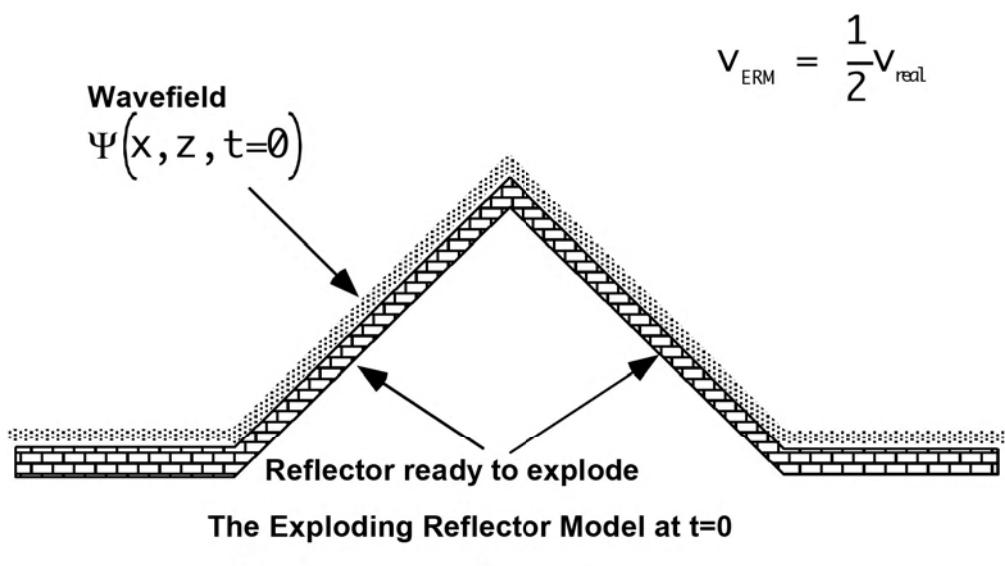
The ERM model thus identifies the mathematical goal of the migration problem:

$$\Psi(x, z, t=0) = \begin{array}{l} \text{a) The ERM wavefield at } t=0 \\ \text{b) The "geology"} \\ \text{c) The migrated section} \end{array}$$

The ERM wavefield is allowed to propagate only in the $-z$ direction (upwards) without reflections, mode conversions, or multiples. It obeys Snell's law in a medium in which the velocities are half the real velocities so that the traveltimes will agree with those of the zos.

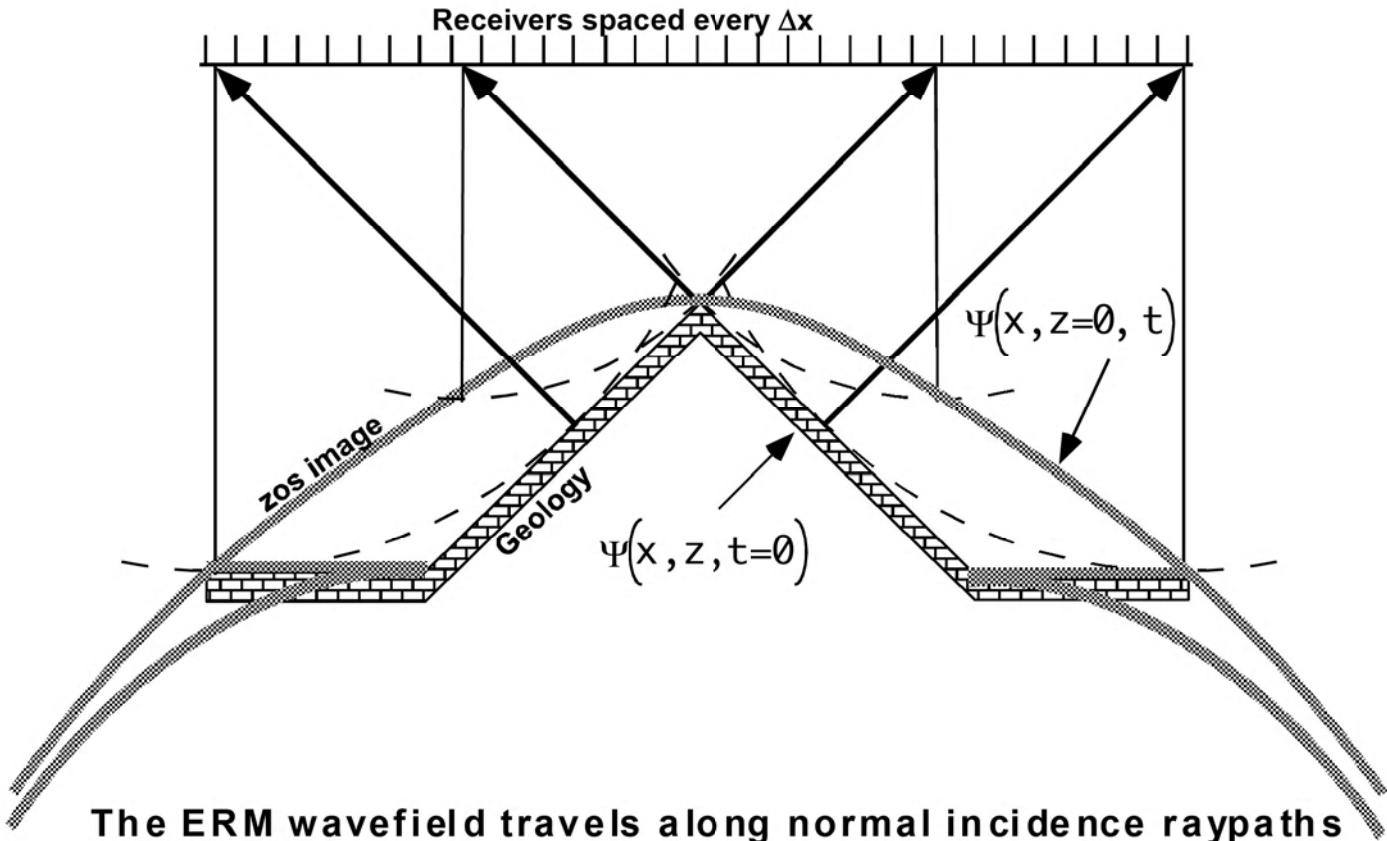
The Exploding Reflector Model

Below in B, we see a snapshot of the ERM wavefield at some time before it arrives at the recorders.



The Exploding Reflector Model

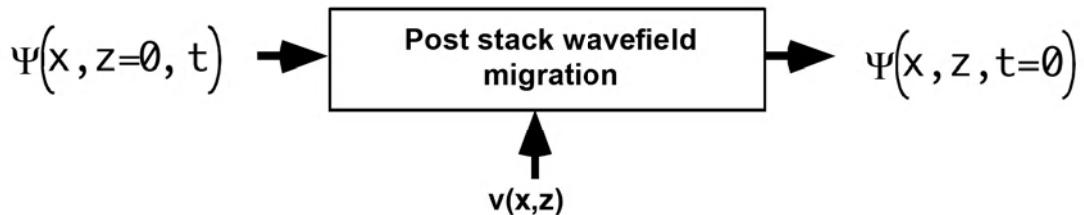
-3-



The ERM wavefield travels along normal incidence raypaths and is recorded at the recording datum ($z=0$) with arrival times equivalent to normal incidence times. When stretched to apparent depth, we see the picture above. Thus the recorded wavefield is:

- $\Psi(x, z=0, t)$ = a) the ERM wavefield at $z=0$
- b) the wavefield model of a zos
- c) the starting point of the post stack migration computation

We can now view migration as a process of wavefield transformation:



The Exploding Reflector Model

- What can be said of the amplitude information in $\Psi(x, z, t=0)$? Is it reflectivity?
- Does the ERM represent time or depth migration?
- Can this theory handle multiples? Is this a good feature?

We can use the exploding reflector model to define yet one more wavefield. So far we have:

$\Psi(x, z, t=0)$ = the geology or the migrated depth section

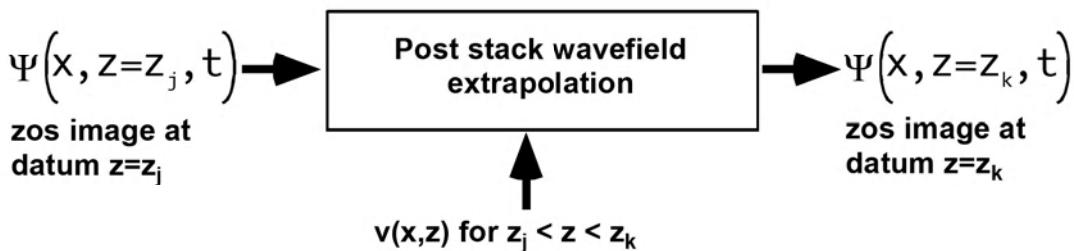
$\Psi(x, z=0, t)$ = the recorded zos image, a model for the cmp stack

$\Psi(x, z, t=t_k)$ = a snapshot of the wavefield at the arbitrary time t_k

Noting that the first and third of these are essentially similar (both are depth sections), we might ask whether we can attach meaning to $\Psi(x, z=z_k, t)$? In fact its meaning is quickly evident: it represents a recording on a datum plane $z=z_k$.

$\Psi(x, z=z_k, t)$ = a zos image as it would be recorded at $z=z_k$

The process of deducing one zos image from another is known as wavefield extrapolation. Though related to migration, extrapolation is mathematically simpler and can serve as a building block for advanced migration algorithms:

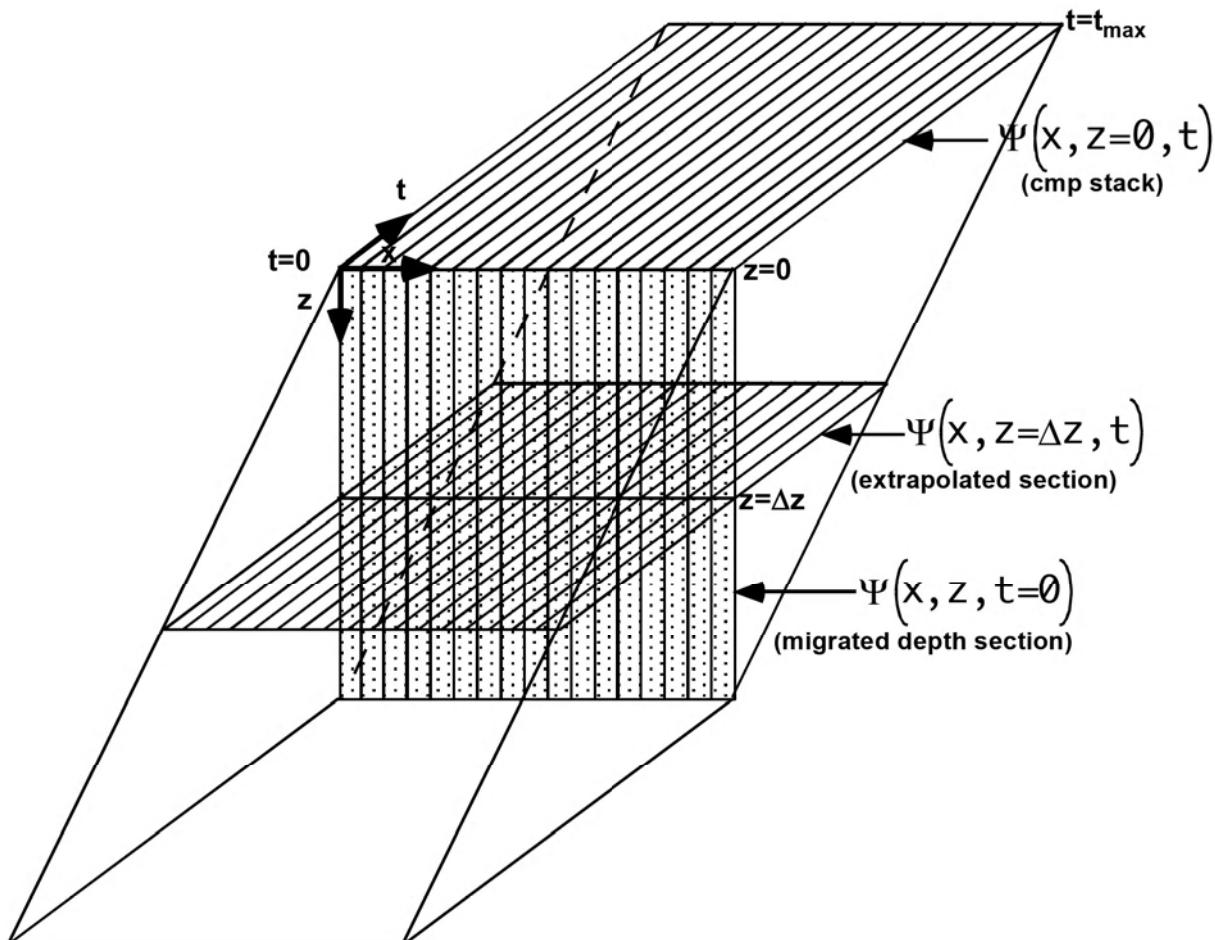


The Exploding Reflector Model

Note that extrapolation requires only local velocity information (that between the two datums involved). Also, it is important to note:

$$\Psi(x, z=z_k, t) \Big|_{t=0} = \Psi(x, z, t=0) \Big|_{z=z_k}$$

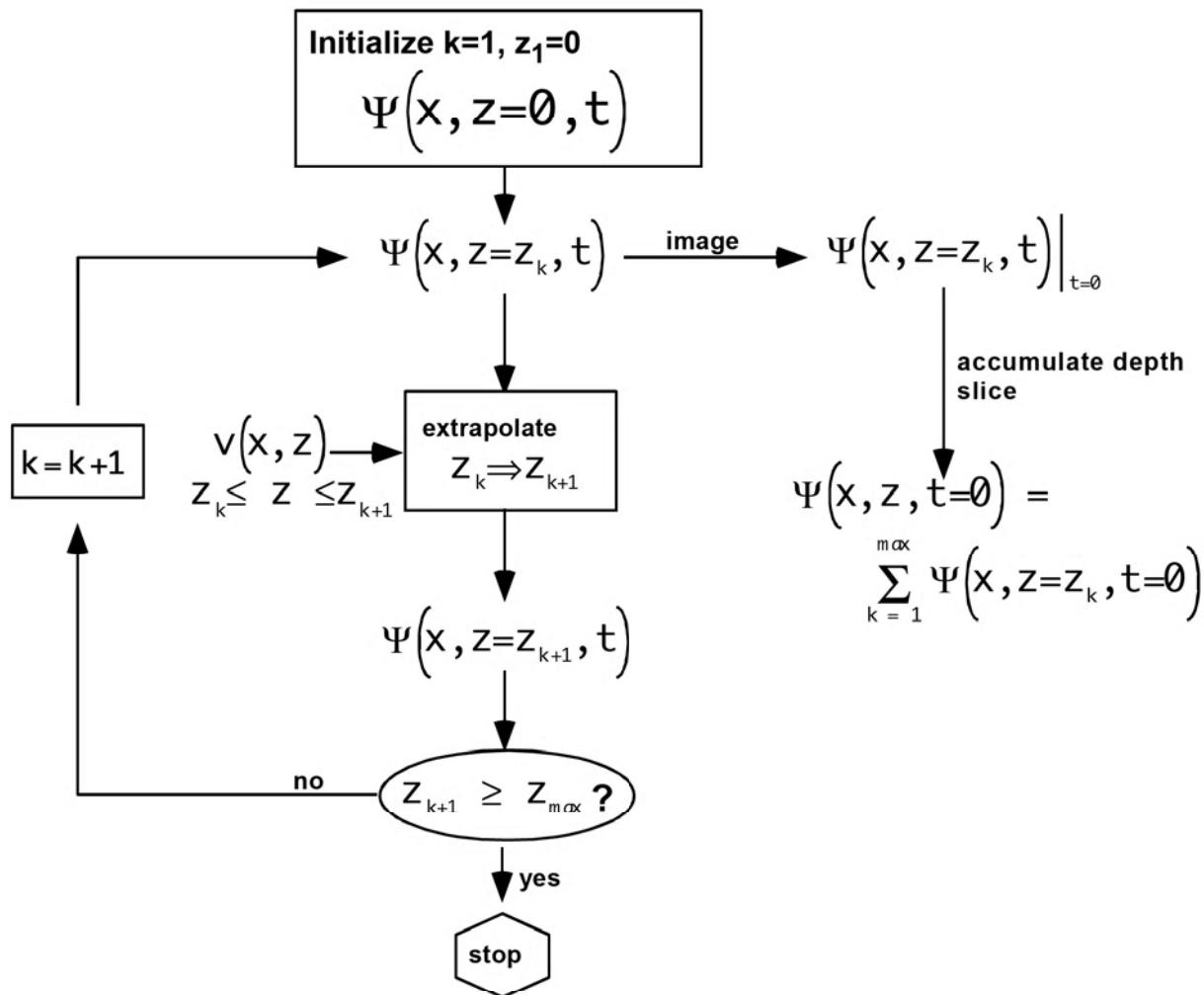
That is, any zos image may be evaluated on its datum ($t=0$) to provide a single depth slice of the migrated depth section. The process of evaluating an extrapolated section on its datum is called imaging.



Here we see the geometric relationship of the various zos images and the migrated depth section for a 2-D problem. The snapshot wavefield has not been drawn but would be a vertical plane at positive time.

The Exploding Reflector Model

Though we have not yet seen precisely how wavefield extrapolation can be accomplished, we can appreciate how it can be used in a recursive "downward continuation" migration algorithm:



Usually, the z_k are regularly spaced, one depth sample apart, and the algorithm gives a regularly sampled depth section directly.

F-K Migration, Geometric Approach

F-K migration was introduced by Stolt (Stolt, R.H., 1978, Migration by Fourier Transform, *Geophysics*, 43, pp23-48) and is an exact solution to the constant velocity migration problem as posed by the exploding reflector model. Stolt also proposed a stretching technique as a preprocessor to handle variable velocity but this is now recognized as a low accuracy approach. F-K migration remains a valuable tool for the role it often plays as a building block in more sophisticated algorithms and as an unsurpassed conceptual and teaching device.

Recall that a propagating Fourier plane wave has its wavenumbers and temporal frequency coupled through the dispersion relation:

$$k_x^2 + k_z^2 = \frac{f^2}{v^2} \quad \text{eqn 1}$$

(For simplicity, the theory will be developed in 2-D. The 3-D case is a relatively simple extension as x and y are treated identically and is left as an exercise for the reader.) This result was derived by studying the relationship between the apparent wavelengths, λ_x and λ_z , and the true wavelength, λ , and was shown to be a geometric consequence of wavefront geometry. It is also true that equation 1 is the mathematical equivalent if the scalar wave equation when expressed in the Fourier domain.

The dispersion relation says that only two of the three quantities: k_x , k_z , and f , are independent and the third is specified by equation 1. Thus if a wavefield is to be built by a superposition of Fourier plane waves (of the form $\exp[2\pi i(k_x x + k_z z + ft)]$) then any two of the frequencies may be specified arbitrarily but the third must be computed from eqn 1 or the result will not be a propagating wave.

F-K Migration, Geometric Approach

Let us apply the exploding reflector model (ERM) to the problem of migrating a single Fourier plane wave. Then, the general problem of migrating a seismic section follows by linear superposition. Let:

$$\Psi(x, z, t) = e^{2\pi i (k_x x + k_z z - ft)} = \text{the ERM wavefield}$$

Then we have:

$$\Psi(x, z=0, t) = e^{2\pi i (k_x x + k_z z)} = \text{the migrated depth section}$$

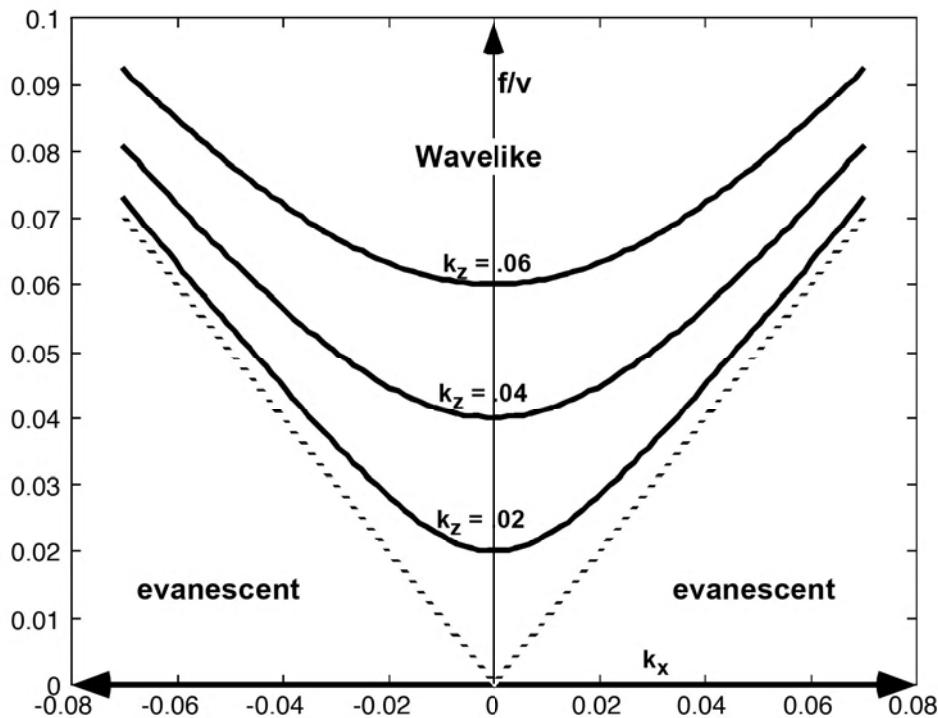
$$\Psi(x, z=0, t) = e^{2\pi i (k_x x - ft)} = \text{the cmp stack}$$

We see that knowledge of the Fourier components of cmp stack allows us to deduce the Fourier components of the migrated depth section through:

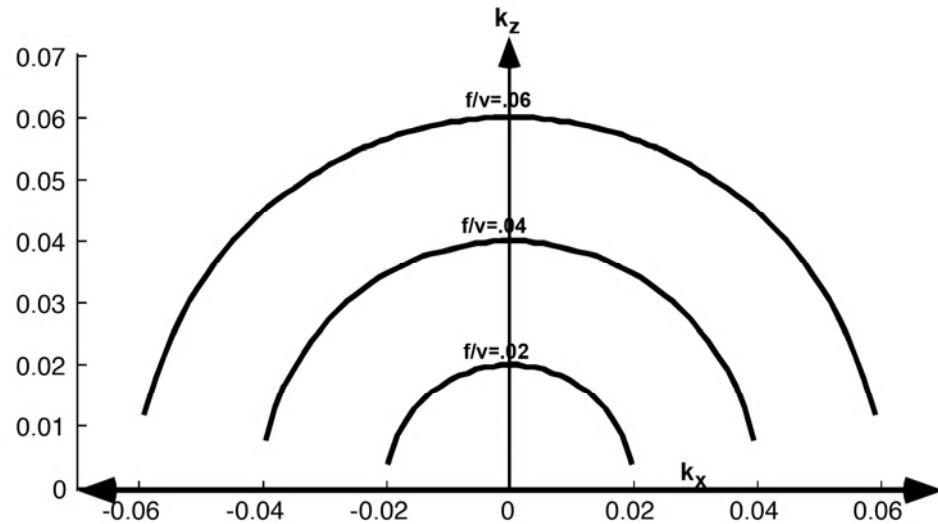
$$k_z = \pm \sqrt{\frac{f^2}{v^2} - k_x^2} \quad \text{eqn 2}$$

The choice of sign in eqn 2 corresponds to either upward or downward traveling waves. In general, both must exist in a real wavefield; however, by the ERM hypothesis, we select only the upward traveling waves (-z direction) by choosing the + sign. We further note that the square root in equation 2 must always evaluate to a real number or we will compute exponentially decaying waves rather than propagating ones. Thus the portion of f-k_x space satisfying $\text{abs}(k_x) < f/v$ is the wavelike portion while the remainder is called the evanescent region. Lines of constant k_z satisfy $(f/v)^2 - k_x^2 = \text{constant}$ and are hyperbolae. Lines of constant f/v in (k_x, k_z) space satisfy $k_x^2 + k_z^2 = \text{constant}$ and are circles.

F-K Migration, Geometric Approach



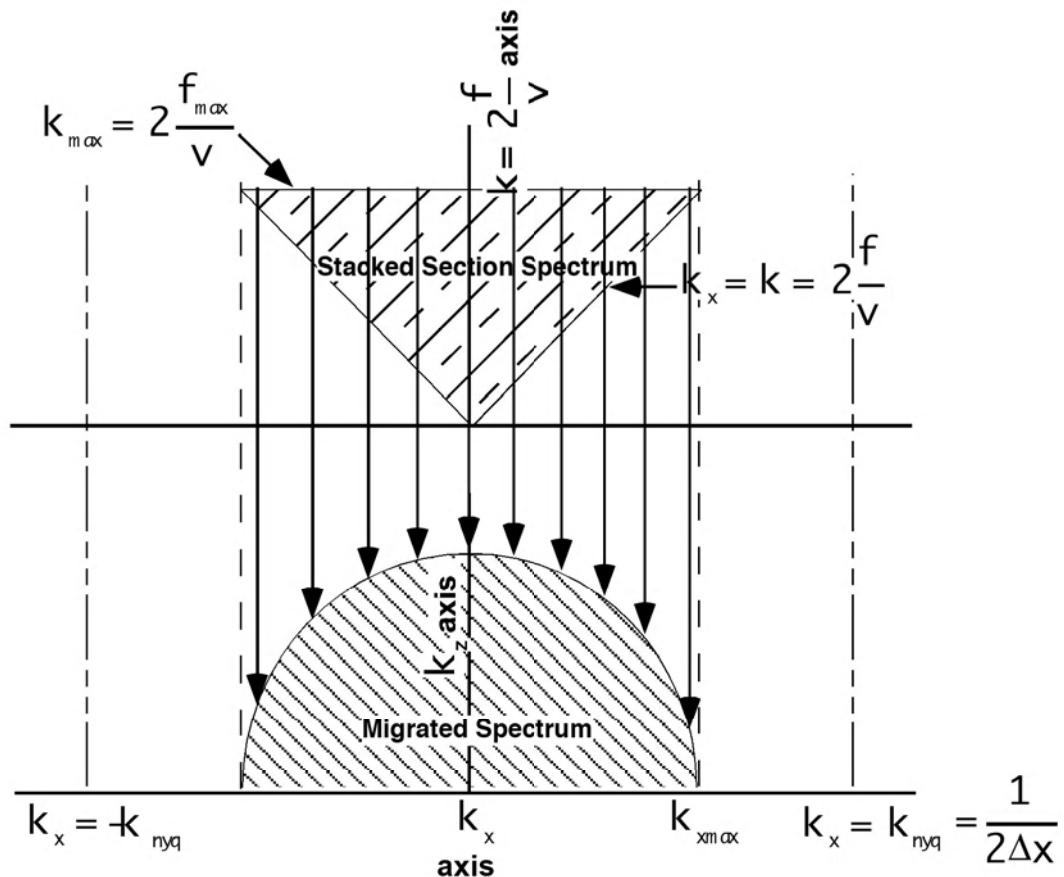
Above we see (k, k_x) space plotted with contours of constant k_z . Note that these are hyperbolae.



This is (k_x, k_z) space plotted with contours of constant f/v . Note that these are circles.

F-K Migration, Geometric Approach

In the Fourier domain, the migration process reduces to a mapping of samples from (k_x, f) to (k_x, k_z) as shown below. The mapping preserves k_x and maps a constant f (a horizontal line in the wavelike region of (k_x, f)) to a circle in (k_x, k_z)



Notice that the maximum k_x value after migration is not necessarily the Nyquist. Instead, k_{\max} is determined by the maximum frequency as:

$$k_{\max} = 2 \frac{f_{\max}}{v} \quad \text{eqn 3}$$

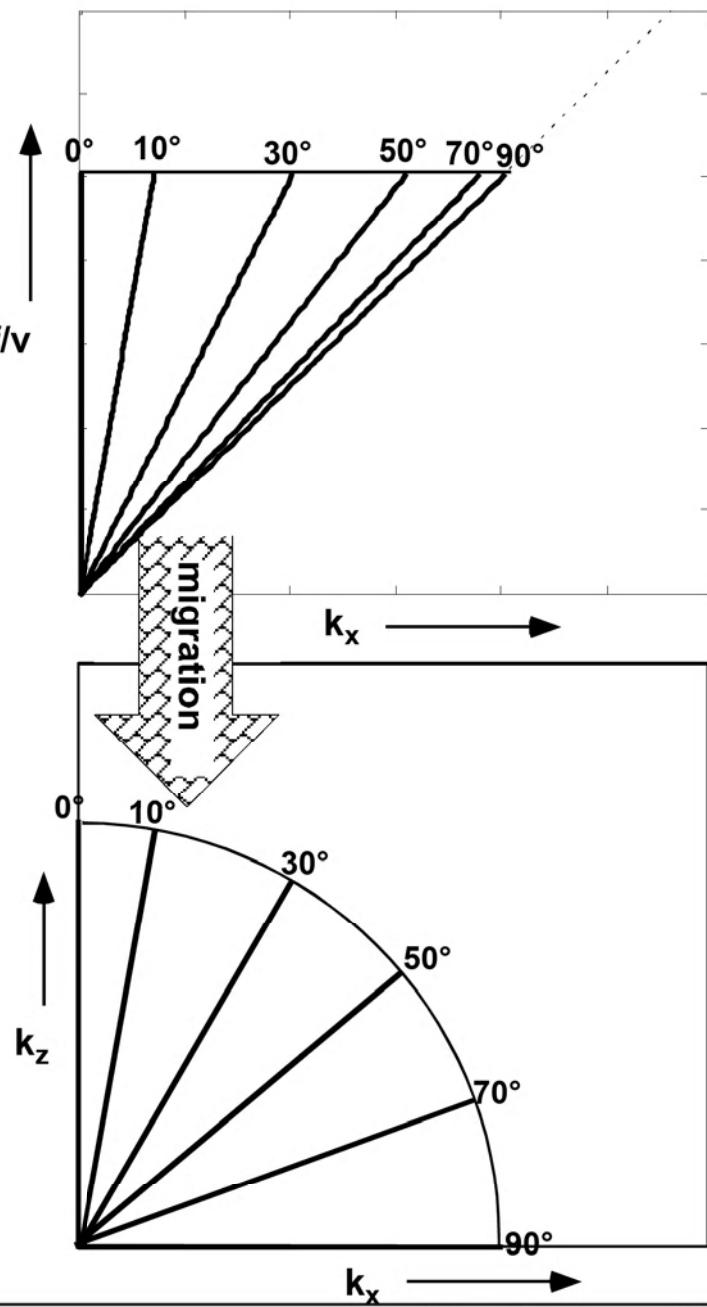
F-K Migration, Geometric Approach

Equation 3 assumes that all "dips" up to and including 90° were migrated. (Recall that we are really talking about the vertical angle of a ray which equals geologic reflector dip for a normal incidence raypath but not for a diffraction.) Thus we should examine how dips map from apparent dip in (k_x, k) space to real dip in (k_x, k_z) space. This mapping is described by the migrators formula which we derived as:

$$\sin(\delta) = \tan(\alpha)$$

where α is the apparent dip and δ is the real dip. The diagrams here show lines of constant δ in $k=f/v$ both spaces. Then in (k_x, k) space this is called the "folded fan" and the process of mapping to (k_x, k_z) space unfolds the fan. We can also see that if the migration algorithm has a maximum dip limit, δ_{\max} , then equation 3 must be modified to:

$$k_{x\max} = 2 \frac{f_{\max}}{v} \sin(\delta_{\max}) \quad \text{eqn 4}$$



F-K Migration, Mathematics

We represent the general ERM wavefield as:

$$\Psi(x, z, t) = \iint_{-\infty}^{\infty} \varphi(k_x, f) e^{2\pi i (ft - k_x x + k_z z)} dk_x df \quad \text{eqn 1}$$

where

$$k_z = \sqrt{\frac{f^2}{v^2} - k_x^2} \quad \text{eqn 2}$$

Then the stacked section is:

$$\Psi(x, z=0, t) = \iint_{-\infty}^{\infty} \varphi(k_x, f) e^{2\pi i (ft - k_x x)} dk_x df \quad \text{eqn 3}$$

Thus we identify $\varphi(k_x, k)$ as the 2-D Fourier transform of the CDP stack. We also identify the migrated solution as:

$$\Psi(x, z, t=0) = \iint_{-\infty}^{\infty} \varphi(k_x, f) e^{2\pi i (k_z z - k_x x)} dk_x df \quad \text{eqn 4}$$

where k_z is given by eqn 2. As it stands, the f integration is not an inverse Fourier transform. To make it so, we substitute:

$$f = v \sqrt{k_x^2 + k_z^2} \quad \text{and} \quad df = \frac{v}{\sqrt{k_x^2 + k_z^2}} k_z dk_z$$

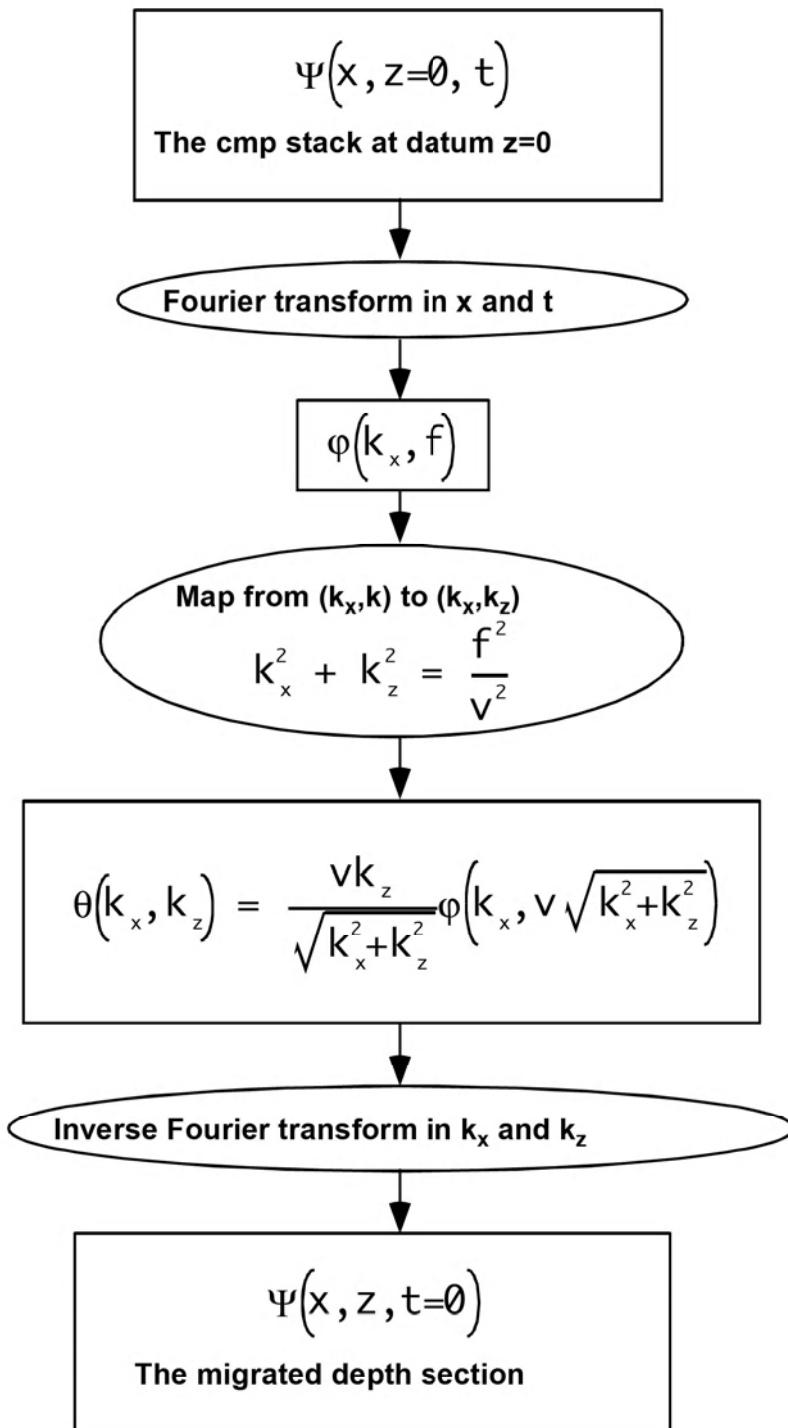
Then eqn 4 becomes:

$$\Psi(x, z, t=0) = \iint_{-\infty}^{\infty} \theta(k_x, k_z) e^{2\pi i (k_z z - k_x x)} dk_x dk_z$$

where $\theta(k_x, k_z) = \frac{vk_z}{\sqrt{k_x^2 + k_z^2}} \varphi(k_x, v \sqrt{k_x^2 + k_z^2})$

eqn 5

F-K Migration, Mathematics



Flow chart for FK migration

F-K Wavefield Extrapolation

Suppose the amplitude and phase of a sine wave of specific wavelength are known at $z=z_0$ and we wish to know its value at $z_0 + \Delta z$. In general the equation for this sine wave is:

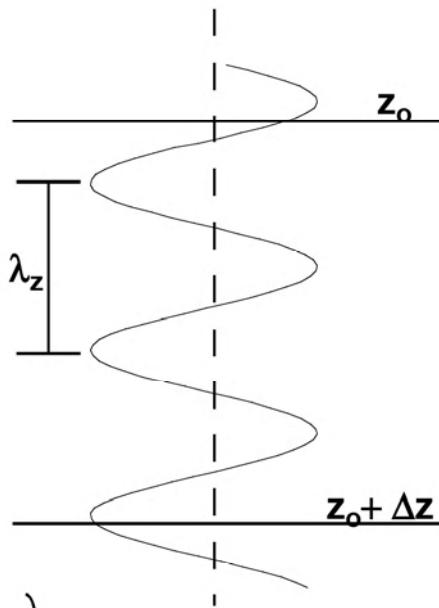
$$f(z) = A \sin(2\pi k_z z + \beta)$$

Where A and β are known constants. For this case:

$$\begin{aligned} f(z_0) &= A \sin(2\pi k_z z_0 + \beta) \\ f(z_0 + \Delta z) &= A \sin(2\pi k_z z_0 + 2\pi k_z \Delta z + \beta) \end{aligned}$$

So we see that $f(z_0 + \Delta z)$ can be computed from $f(z_0)$ by adding an amount $2\pi k_z \Delta z$ to the argument of the sine function. This term, called a phase shift, is simply computed from known values and is related to λ_z by:

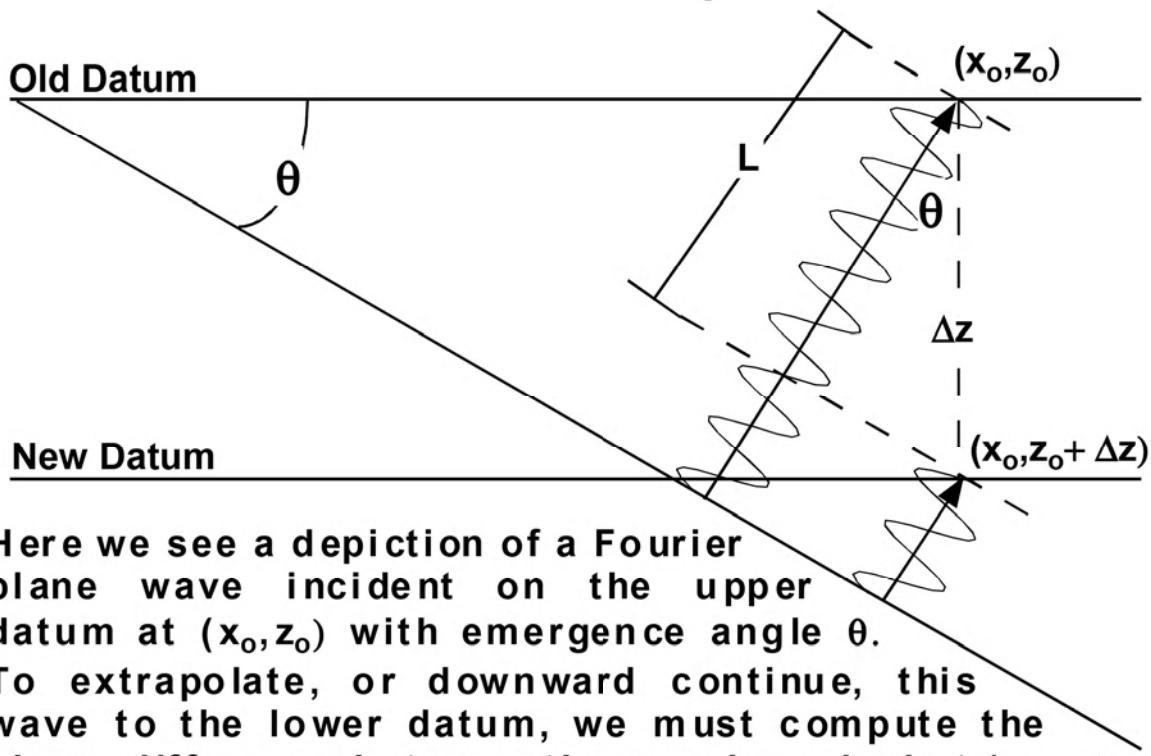
$$\Phi = 2\pi k_z \Delta z = \frac{2\pi \Delta z}{\lambda_z} \quad \text{eqn 1}$$



So periodic functions are trivially easy to extrapolate by a simple phase shift. What's more, the extrapolation is accurate for any value of Δz so long as k_z is known.

The downward continuation of upward traveling waves is of great importance in migration theory, as we have seen from our examination of the exploding reflector model (ERM). The phase shift required for this purpose is always given by equation 1, however, the form of k_z changes depending on the number of dimensions. That is, the phase shift is always $2i$ times the fractional number of z wavelengths that fit in the extrapolation step.

F-K Wavefield Extrapolation



$$\Phi = \frac{2\pi L}{\lambda} \quad \text{but:} \quad \lambda = \frac{v}{f} \quad \text{and} \quad L = \Delta z \cos(\theta)$$

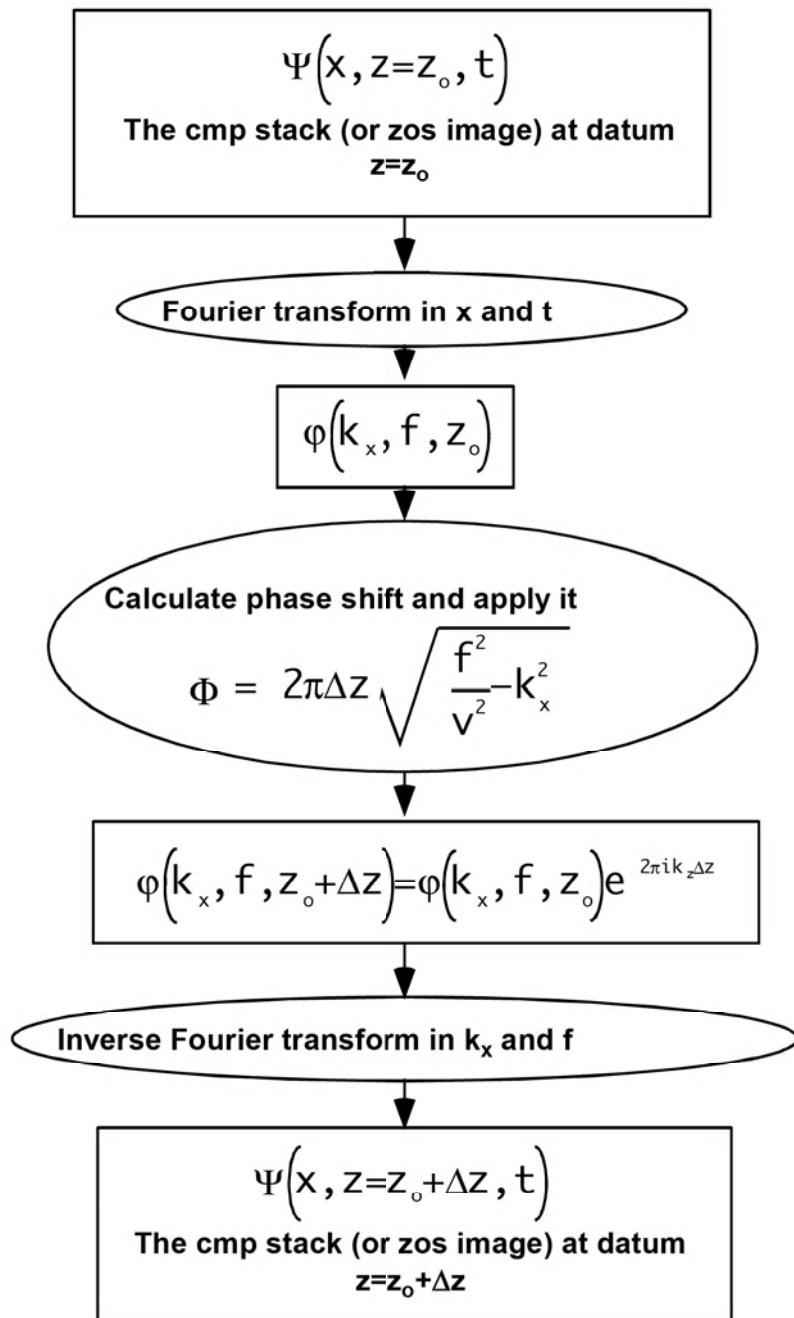
$$\text{so } \Phi = \frac{2\pi f \Delta z \cos(\theta)}{v} \quad \text{Now recall that} \quad \sin(\theta) = \frac{vk_x}{f}$$

$$\text{so} \quad \boxed{\Phi = \frac{2\pi f \Delta z}{v} \sqrt{1 - \frac{v^2 k_x^2}{f^2}} = 2\pi \Delta z \sqrt{\frac{f^2}{v^2} - k_x^2}} \quad \text{eqn 2}$$

and, we recognize the square root in the last term as k_z so

$\Phi = 2\pi k_z \Delta z$ which is the same as for the 1-D case (eqn 1).

F-K Wavefield Extrapolation



Flowchart for F-K wavefield extrapolation.

F-K Wavefield Extrapolation

The mathematical derivation of F-K extrapolation is a straight forward consequence of the general expression for the ERM wavefield:

$$\Psi(x, z, t) = \iint_{-\infty}^{\infty} \varphi(k_x, f) e^{2\pi i (ft - k_x x + k_z z)} dk_x df \quad \text{eqn 3}$$

where $k_z = \sqrt{\frac{f^2}{v^2} - k_x^2}$

We simply evaluate this expression on two different datums and deduce the relation:

$$\Psi(x, z=z_o, t) = \iint_{-\infty}^{\infty} \varphi(k_x, f) e^{2\pi i (ft - k_x x + k_z z_o)} dk_x df$$

$$\Psi(x, z=z_o + \Delta z, t) = \iint_{-\infty}^{\infty} \varphi(k_x, f) e^{2\pi i (ft - k_x x + k_z z_o + k_z \Delta z)} dk_x df$$

Comparing these two expressions shows that they are related by an extrapolation operator applied in the frequency domain given by:

$$e^{2\pi i k_z \Delta z} = \text{wavefield extrapolation operator}$$

Thus, in the Fourier domain, the downward extrapolation of the upward traveling ERM wavefield is expressed by:

$$\varphi(k_x, f, z_o + \Delta z) = \varphi(k_x, f, z_o) e^{2\pi i k_z \Delta z}$$

eqn 4

Recursive F-K Wavefield Extrapolation for $v=v(z)$

We will now apply the F-K extrapolation formalism for the case of a $v(z)$ medium. We will develop a recursive algorithm for the construction of $\Psi(x,z,t=0)$ with the additional assumption that we will ignore transmission losses at velocity boundaries. If $\phi(k_x, f, z=0)$ is the transform of the cmp stack, then we extrapolate it downwards by Δz with:

$$\phi(k_x, f, \Delta z) = \phi(k_x, f, 0) \exp \left(2\pi i \sqrt{\frac{f^2}{v_o^2} - k_x^2} \Delta z \right) \quad \text{eqn 1}$$

where v_o is the near surface velocity. Next, we extrapolate another step down through $v=v_1$ with:

$$\begin{aligned} \phi(k_x, f, 2\Delta z) &= \phi(k_x, f, \Delta z) \exp \left(2\pi i \sqrt{\frac{f^2}{v_1^2} - k_x^2} \Delta z \right) \\ \phi(k_x, f, 2\Delta z) &= \phi(k_x, f, 0) \exp \left(2\pi i \Delta z \left(\sqrt{\frac{f^2}{v_o^2} - k_x^2} + \sqrt{\frac{f^2}{v_1^2} - k_x^2} \right) \right) \end{aligned}$$

So we see that the phase shifts simply sum and we immediately deduce the result for n layers as:

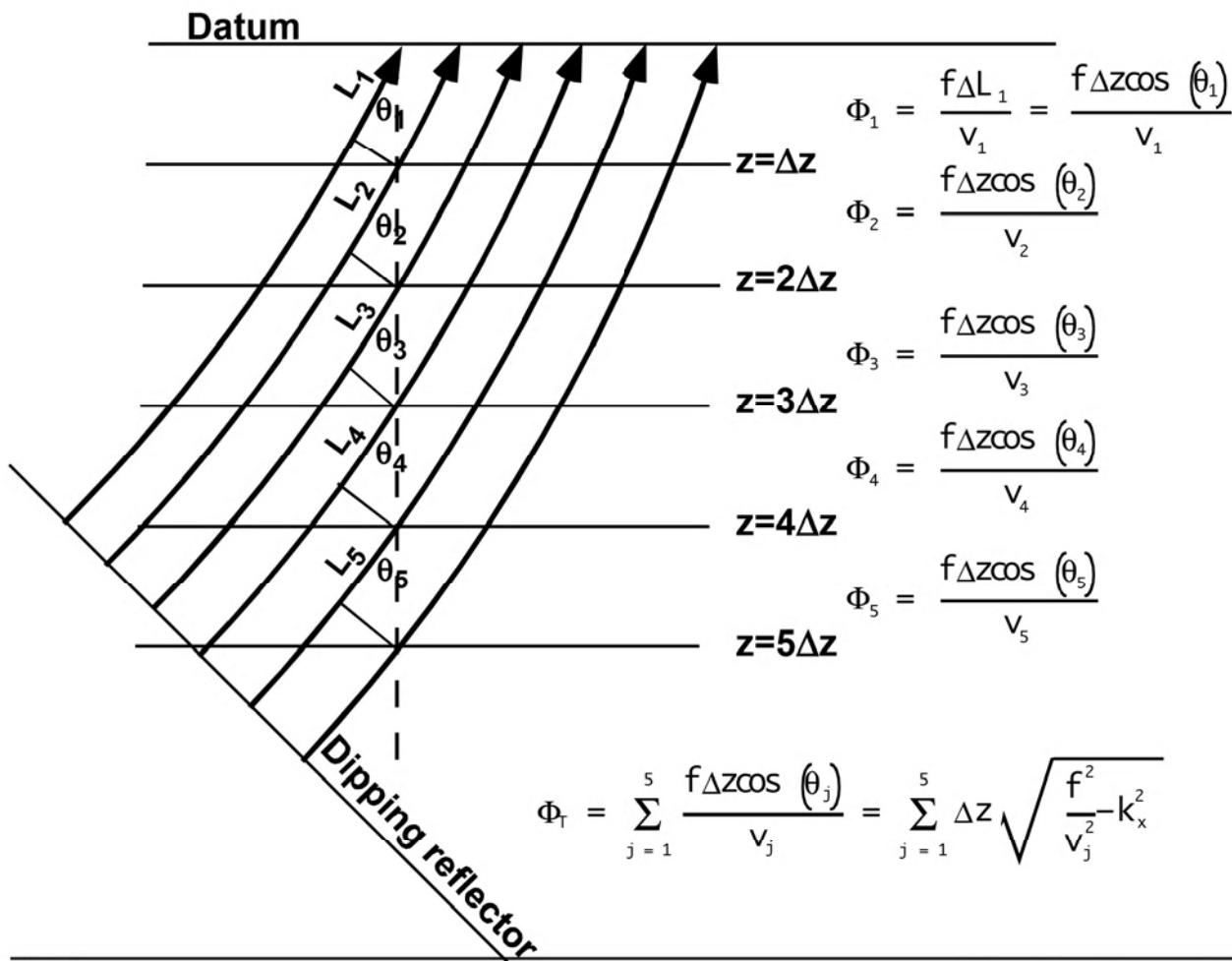
$$\phi(k_x, f, n\Delta z) = \phi(k_x, f, 0) \exp \left(2\pi i \Delta z \sum_{j=1}^n \sqrt{\frac{f^2}{v_j^2} - k_x^2} \right) \quad \text{eqn 2}$$

Recursive F-K Wavefield Extrapolation for $v=v(z)$

In the limit as the Δz step shrinks to zero, the summation becomes an integral and we obtain:

$$\varphi(k_x, f, z) = \varphi(k_x, f, 0) \exp \left(2\pi i \int_0^z \sqrt{\frac{f^2}{v(z')^2} - k_x^2} dz' \right) \quad \text{eqn 3}$$

This result has been derived by many scientists (most notably in Quantum Mechanics) and is known as a WKBJ solution to the wave equation. As a seismic migration technique, the method was first published by Gazdag (Gazdag, J, 1978, Wave-equation migration by phase shift, Geophysics, 43, 1342-1351) and is known as the Gazdag phase shift method.



The Extrapolation Operator

We have shown that wavefield extrapolation can be accomplished in the F-K domain by phase shift:

$$\varphi(k_x, f, z_o + \Delta z) = \varphi(k_x, f, z_o) \exp\left(2\pi i \Delta z \sqrt{\frac{f^2}{v_o^2} - k_x^2}\right) \quad \text{eqn 1}$$

Let us examine the phase shift in more detail:

phase shift = $\Phi = \Delta z \sqrt{\frac{f^2}{v_o^2} - k_x^2} = \frac{f \Delta z}{v_o} \sqrt{1 - \frac{k_x^2 v_o^2}{f^2}}$

rewrite this as: $\Phi = \mu_s + \mu_f \quad \text{eqn 2}$

where $\mu_s = \frac{f \Delta z}{v_o} = f \Delta \tau$

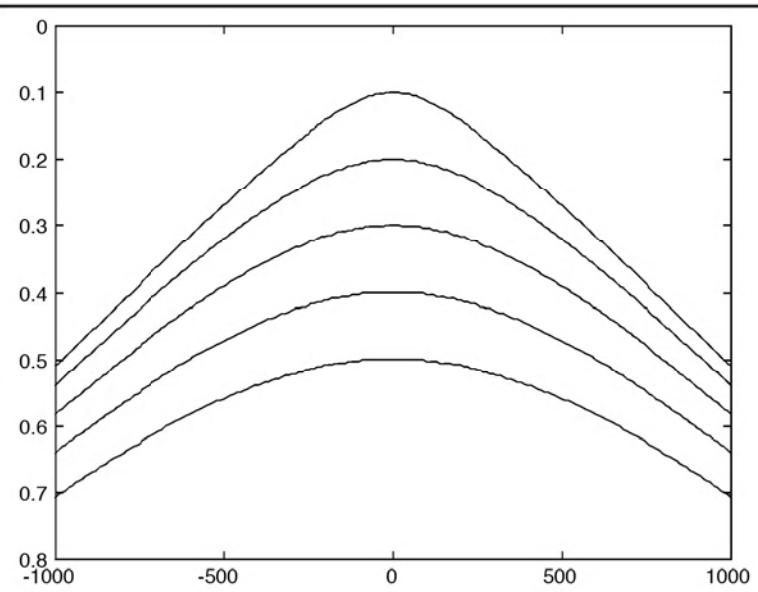
and $\mu_f = \frac{f \Delta z}{v_o} \left(\sqrt{1 - \frac{k_x^2 v_o^2}{f^2}} - 1 \right) = f \Delta \tau \left(\sqrt{1 - \frac{k_x^2 v_o^2}{f^2}} - 1 \right)$

μ_s is called the "static phase shift" or "thin lens term" and represents a constant time shift independent of dip. μ_f is called the "focussing phase shift" because it causes the motion of dipping events. Note that μ_f vanishes for $k_x = 0$. This form of the extrapolation phase shift (eqn 2) is convenient because it emphasizes both the time shift and focussing aspects of extrapolation.

The Extrapolation Operator

Since the extrapolation operator is applied via multiplication in the Fourier domain, it must be a multi-dimensional convolution in space-time. Consider the diffraction chart below:

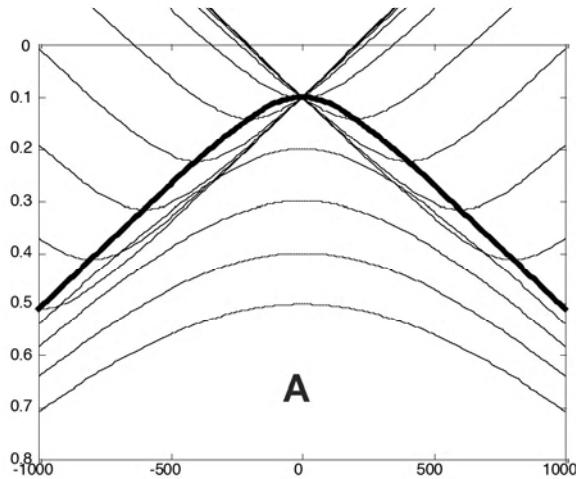
Each hyperbola in this chart represents the response of a point diffractor as seen from the datum $z=0$. The chart was drawn for a constant velocity of 2000m/s and has point diffractors at depths: 200m 400m 600m 800m & 1000m. (We regard the velocity as an exploding reflector velocity.) If we



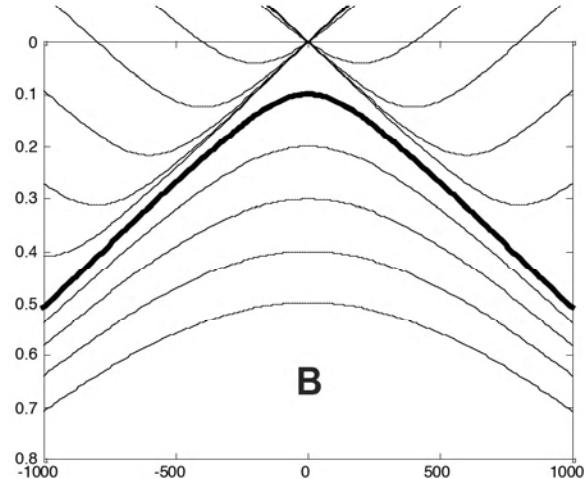
downward continue this chart to the depth of the first diffractor, we expect to see the top hyperbola focussed and the remaining four should be shifted upward in time and focussed somewhat. In fact, the second hyperbola should be transformed into the first, the third into the second, and so on.

After some deliberation, we might guess that cross correlating the chart with the first hyperbola would do the trick. Certainly, it would focus the first hyperbola but what would it do to the others? Recalling that a cross correlation is a convolution with the time reverse of the operator, we can do the operation as convolution by replacement with the first hyperbola upside down as the convolution operator.

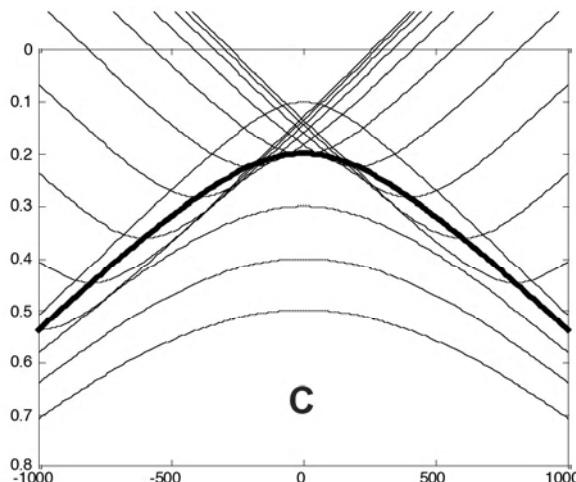
The Extrapolation Operator



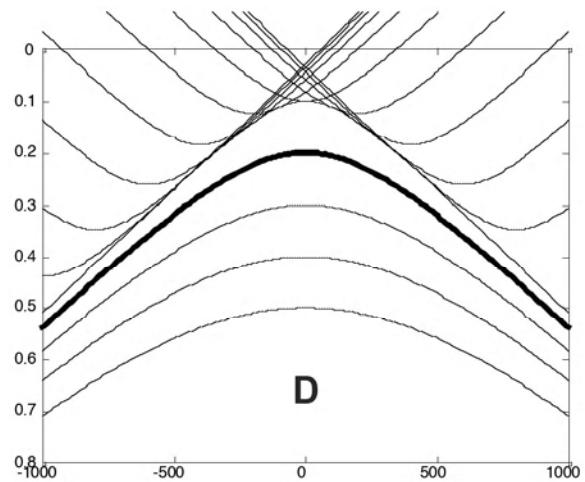
A) Focussing term only. First hyperbola convolved with its time reverse without the time shift.



B) Time shift and Focussing term. First hyperbola convolved with its time reverse including the time shift.



C) Focussing term only. Second hyperbola convolved with the time reverse of the first without the time shift.



D) Time delay and Focussing term. Second hyperbola convolved with the time reverse of the first including the time shift. Note that the image formed is the first hyperbola.

Vertical Time-Depth Conversions

It is quite common to convert a single trace from time to depth, or the reverse, by a simple 1-D conversion. That is, we map samples from the time trace to the depth trace by:

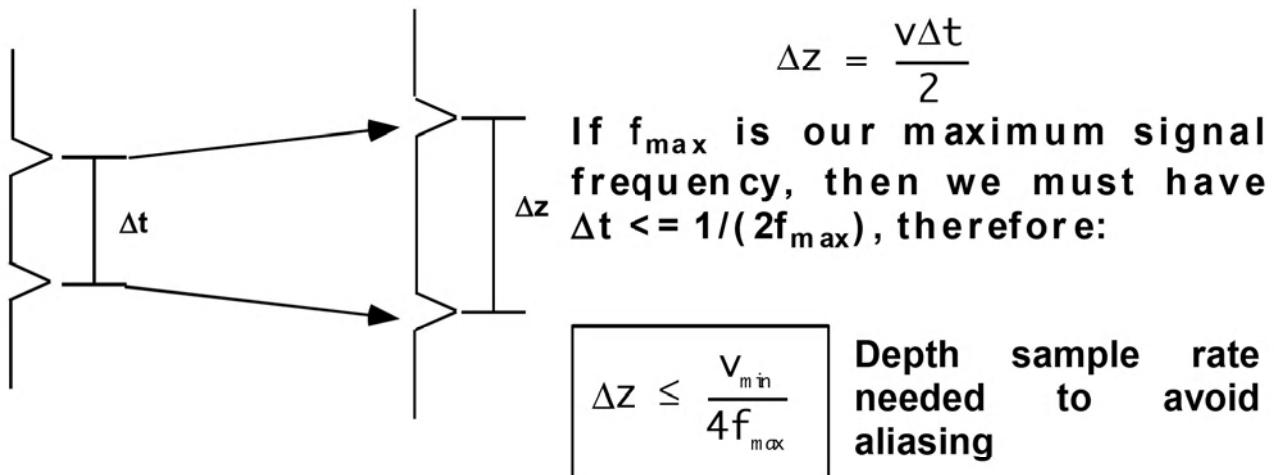
$$z(\tau) = \int_0^\tau v_{\text{ns}}(\tau') d\tau'$$

and from depth to time by:

$$\tau(z) = \int_0^z \frac{dz'}{v_{\text{ns}}(z')}$$

These 1-D operations are NOT migrations and are commonly called stretches. They are very useful for converting from migrated depth to migrated time (or the reverse) or from unmigrated time to unmigrated depth (or the reverse). Thus it is possible to have time or depth displays both before and after migration and the interpreter must be careful to understand what is being presented.

A form of aliasing is possible in all such operations (including migrations) though it is most easily analyzed in 1-D. Consider the conversion of two samples from time to depth:



Time and Depth Migration in Depth

Thus far, we have developed a theory for constructing a migrated depth section $\Psi(x,z,t=0)$ from a zero offset section, $\Psi(x,z=0,t)$. It is often desired to express the migrated data with a vertical time coordinate for interpretation. Called τ , this migrated time is created from z by a simple 1-D stretch using the migration velocity model. It is always correct to do this following migration with $v(x,z)$ where the stretch is defined by:

$$\tau(x,z) = \int_0^z \frac{dz'}{v(x,z')} \quad \text{eqn 1}$$

A display created from a migration depth section using equation 1 is called a "migrated time" display which is a distinct term from the common "time migration". The two are equivalent only when lateral velocity variations are absent.

Time migration seeks to create $\Psi(x,\tau,t=0)$ directly without first creating $\Psi(x,z,t=0)$. To see how this might be done recall that the Gazdag phase shift extrapolation can be written as:

$$\varphi(\Delta z) = \varphi_o e^{2\pi i (\mu_s + \mu_f)} \quad \text{eqn 2}$$

where $\mu_s = \frac{f \Delta z}{v_o} \quad \text{eqn 3a}$

$$\mu_f = \frac{f \Delta z}{v_o} \left(\sqrt{1 - \frac{k_x^2 v_o^2}{f^2}} - 1 \right) \quad \text{eqn 3b}$$

Time and Depth Migration in Depth

As mentioned previously, μ_s is called the static phase shift and causes the upward (in time) movement of extrapolated data while μ_f causes focussing but no bulk time shift. Thus, we can implement a focussing only "extrapolation" by omitting the μ_s term from the extrapolation operator and letting $\Delta z/v_o = \Delta\tau$:

$$\varphi(\Delta\tau) = \varphi_o e^{2\pi i \mu_{f\tau}} \quad \text{eqn 4}$$

where $\mu_{f\tau} = f\Delta\tau \left(\sqrt{1 - \frac{k_x^2 v_o^2}{f^2}} - 1 \right)$ eqn 5

Recursive extrapolation with equation 4 (or some approximation to it) is a form of time migration while recursive extrapolation with equation 2 is a depth migration. As mentioned previously, there is no essential difference between the two if there are no lateral velocity variations. This is because the time migration can be stretched to depth or the depth migration to time with:

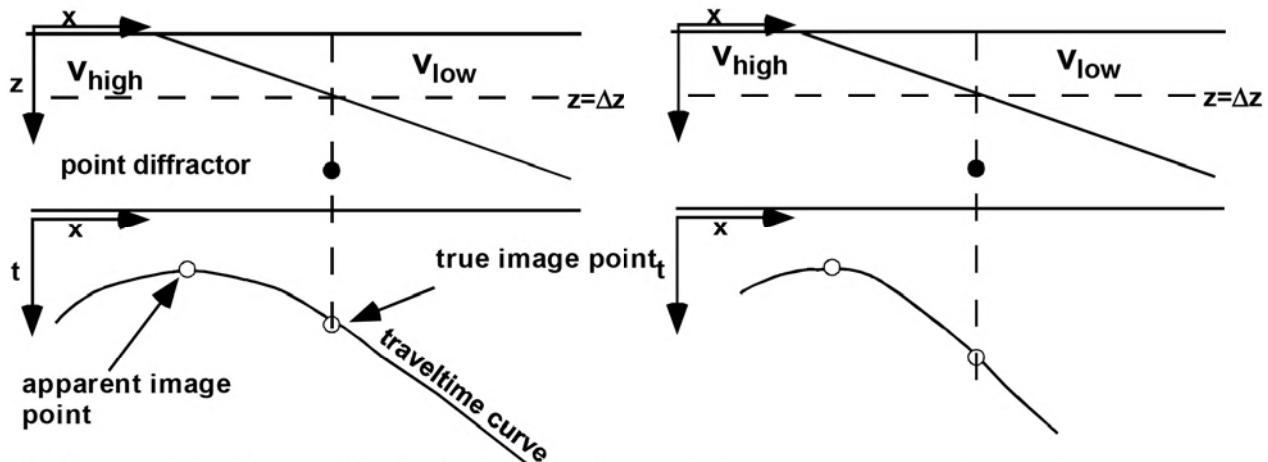
$$\tau(z) = \int_0^z \frac{dz'}{v(z')} \quad \text{or} \quad z(\tau) = \int_0^\tau v(t)dt \quad \text{eqn 6}$$

We see that this is the case because the essential difference between the two techniques is the term:

$$\exp \left(2\pi i \int_0^z \mu_z(z') dz' \right) = \exp \left(2\pi i f \int_0^z \frac{dz'}{v(z')} \right) \quad \text{eqn 7}$$

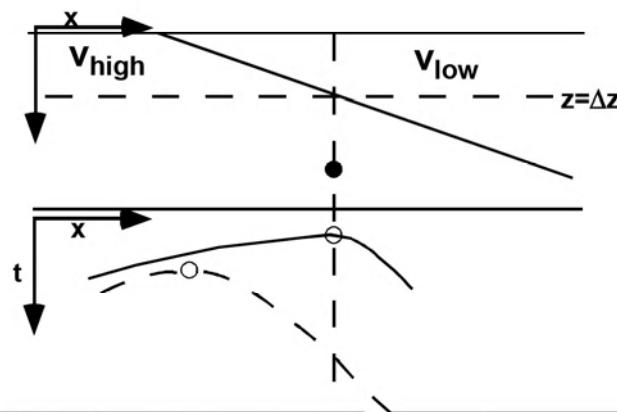
Time and Depth Migration in Depth

Equation 7 causes a stretch equivalent to equations 6 when applied in the frequency domain. Time migration is thus equivalent to depth migration when $v=v(z)$ because the static delay, $\Delta z/v(z) = \Delta\tau$, does not depend on x . Since data moves all around during migration following various raypaths, any accumulated time delays in $\Psi(x, \tau, t=0)$ will be complicated functions of dip and position if v varies with x . When v varies only with z , then all data at fixed τ have the same accumulated delay regardless of migration raypath. When v varies laterally, time migration actually refracts data in a way that systematically violates Snell's law.



1) Response of a point diffractor under lateral velocity variations. Note that the apparent image point is displaced to the left because of the high velocity layer.

2) After extrapolation to $z=\Delta z$ using only the focussing phase shift. Since it vanishes when time dip is zero, it cannot change the apparent image point.



3) After application of the static shift term, the apparent image point has moved to coincide with the real image point. This is because the time delay varies with x and so changes the shape of the traveltime curve.

Kirchhoff Migration

Kirchhoff migration refers to any method of reflector imaging which follows from the integral solution to the (constant velocity) scalar wave equation. Schneider (1978, Integral formulation for migration in two and three dimensions: Geophysics, 43, 49-76) derived the basic equations directly from Green's theorem for the scalar wave equation. For a 3-D migration, the result for the exploding reflector wavefield can be expressed:

$$\Psi(x,y,z,t) = \frac{1}{2\pi} \int dx_0 dy_0 \left[\underbrace{\frac{\cos(\theta)}{r^2} \Psi(x_0, y_0, z=0, t-r/v)}_{\text{near-field term}} + \underbrace{\frac{\cos(\theta)}{vr} \frac{\partial}{\partial t} \Psi(x_0, y_0, z=0, t-r/v)}_{\text{far-field term}} \right] \quad (1)$$

where $r = \sqrt{(x-x_0)^2 + (y-y_0)^2 + z^2}$ $\cos(\theta) = \frac{\sqrt{(x-x_0)^2 + (y-y_0)^2}}{r}$

Equation (1) can be regarded as the space-time equivalent of Stolt's f-k migration algorithm. It is exact for constant velocity and an excellent approximation for variable velocity.

Since the near field term decays as r^{-2} while the far field term goes as r^{-1} , the latter is generally dominant and (1) is usually approximated as:

$$\Psi(x,y,z,t) \approx \frac{1}{2\pi} \int \frac{\cos(\theta)}{vr} \frac{\partial}{\partial t} \Psi(x_0, y_0, z=0, t-r/v) dx_0 dy_0 \quad (2)$$

Kirchhoff Migration

Setting $t=0$ in (2) gives the migrated section:

$$\Psi(x,y,z,t=0) \approx \frac{1}{2\pi} \int \frac{\cos(\theta)}{vr} \frac{\partial}{\partial t} \Psi(x_o, y_o, z=0, r/v) dx_o dy_o \quad (3)$$

obliquity factor
spreading factor

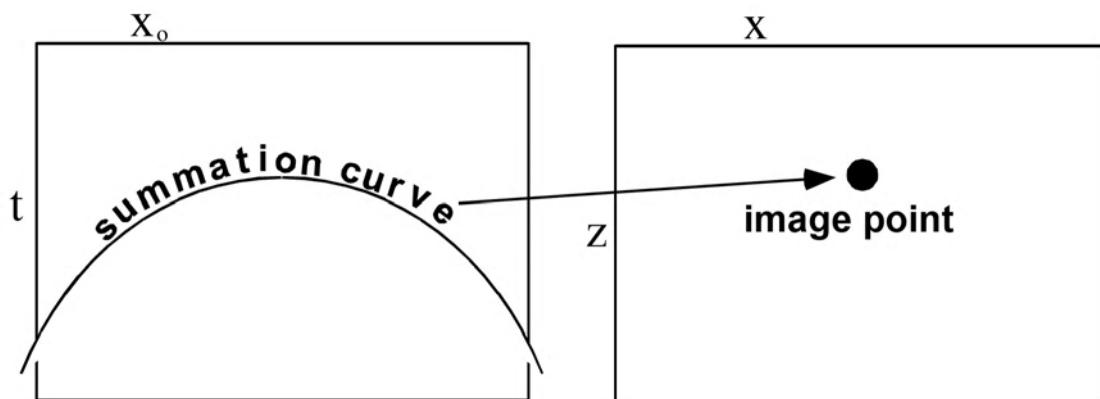
The implementation of equation (3) can proceed as follows:

A) Differentiate the input section with respect to time.

B) For each output point:

- 1) compute the traveltime surface $t=r/(v(x,y,z))$
- 2) sum through the differentiated data along the traveltime surface. Apply the obliquity factor and spreading factor at the same time.

Accurate implementation requires a sophisticated raytracing method for step B.1 and a high fidelity interpolator for step B.2. In 2-D the result is similar though the time derivative must be raised to the 1/2 power and the summation is along a diffraction curve.



Kirchhoff Migration

Kirchoff migration is one of the most adaptable migration schemes available. It can be easily modified to account for such difficulties as:

- topography
- irregular recording geometry
- prestack migration
- converted wave imaging

When formulated as a depth migration, it tends to be a slow method because great care must be taken in the raytracing. When formulated as a time migration, straight ray calculations using rms velocities can be done to greatly speed the process.

Another advantage of Kirchoff methods is the ability to perform "target oriented" migrations. That is, equation (3) need only be evaluated for those points in (x,z) which comprise the target. Since the cost of computation is directly proportional to the number of output points, this can greatly reduce the run times and makes migration parameter testing very feasible.

Finite Difference Concepts

Finite difference techniques are perhaps the most direct, but often the least intuitive, of migration methods. As the name implies, they involve the approximation of analytic derivatives of the wave equation with finite difference expressions. To see how this might be done, consider the definition of the derivative of an arbitrary function $f(z)$:

$$\frac{df(z)}{dz} = \lim_{\Delta z \rightarrow 0} \frac{1}{\Delta z} (f(z + \Delta z) - f(z)) \quad (1)$$

A simple finite difference approximation for this derivative is:

$$\frac{df(z)}{dz} \approx \frac{1}{\Delta z} (f(z + \Delta z) - f(z)) \quad (2)$$

Here Δz is assumed small with respect to length scales of interest but is still finite.

To see how finite difference operators can be used to predict or extrapolate a function, suppose that values for $f(z)$ and its derivative are known at z , then equation (2) can be rearranged to predict $f(z + \Delta z)$:

$$f(z + \Delta z) \approx f(z) + \frac{df(z)}{dz} \Delta z \quad (3)$$

This can be regarded as a truncated Taylor series. Taylor's theorem says that:

$$f(z + \Delta z) = f(z) + \frac{df(z)}{dz} \Delta z + \frac{1}{2} \frac{d^2 f(z)}{dz^2} \Delta z^2 + \frac{1}{6} \frac{d^3 f(z)}{dz^3} \Delta z^3 + \dots \quad (4)$$

Finite Difference Concepts

It is reasonable to ask what is the relationship between extrapolation of a function with equation (3) and the phase shift extrapolation operator which we have derived from the scalar wave equation. Recall that if $f(z)$ is a solution to the wave equation in the Fourier domain, then it can be extrapolated by:

$$f(z+\Delta z) = f(z)e^{ik_z\Delta z} \quad (5)$$

Here, k_z is the vertical wavenumber whose value may be determined from the scalar wave dispersion relation:

$$k_z = \pm \sqrt{\frac{\omega^2}{v^2} - k_x^2} \quad (6)$$

Suppose that $f(z)$ in equation (4) is the complex exponential: $\exp(ik_z z)$; then the nth order derivative is equivalent to multiplying by (ik_z) raised to the nth power:

$$f(z+\Delta z) = f(z) + ik_z f(z) \Delta z + \frac{1}{2} (ik_z)^2 f(z) \Delta z^2 + \frac{1}{6} (ik_z)^3 f(z) \Delta z^3 + \dots$$

$$f(z+\Delta z) = f(z) \left(1 + ik_z \Delta z + \frac{1}{2} (ik_z \Delta z)^2 + \frac{1}{6} (ik_z \Delta z)^3 + \dots \right) \quad (7)$$

The series expression in (7) is the well known exponential series which sums to $\exp(ik_z \Delta z)$. Thus is is equivalent to equation (5). We can conclude that:

- Extrapolation by phase shift is equavalent to extrapolation with an infinite Taylor series
- Extrapolation by finite difference approximations is equavalent to extrapolation with a truncated Taylor series.

Finite Difference Concepts

The finite difference approximation used in equation (2) is the most simple possible. Many other approximations have been explored which are much more accurate. Note that equation (2) estimates a derivative using just two points and is therefore considered a "local" approximation. Higher order expressions generally use more points and so become increasingly less local.

$$\frac{df(z)}{dz} \approx \frac{1}{\Delta z} (f(z + \Delta z) - f(z))$$

forward first order first derivative approximation

$$\frac{df(z)}{dz} \approx \frac{1}{\Delta z} (f(z) - f(z - \Delta z))$$

backward first order first derivative approximation

$$\frac{df(z)}{dz} \approx \frac{1}{2\Delta z} (f(z + \Delta z) - f(z - \Delta z))$$

centered first order first derivative approximation

$$\frac{d^2f(z)}{dz^2} \approx \frac{1}{\Delta z^2} (f(z + \Delta z) - 2f(z) + f(z - \Delta z))$$

centered first order second derivative approximation

Finite Difference Migration

Most finite difference migrations methods are formulated in the space-time domain though space-frequency methods are also common. Also, they are all recursive extrapolation techniques. Though there are many equivalent ways to develop any particular method, a consistent approach can be formulated by beginning with in the frequency domain with phase shift theory as follows:

1) Develop a suitable rational approximation to the one way dispersion relation for scalar waves. A rational approximation is one which eliminates the square root though it may involve multiplication, division, and powers of frequencies and wavenumbers.

2) Construct a space-time domain differential equation from the approximate dispersion relation by replacing:

$$k_z \Rightarrow -i\frac{\partial}{\partial z}; \omega \Rightarrow i\frac{\partial}{\partial t}; k_x \Rightarrow -i\frac{\partial}{\partial x}; k_y \Rightarrow -i\frac{\partial}{\partial y}$$

3) Choose an appropriate form for your finite difference operators (i.e. forward, backward, or central differences)

4) Develop the difference equation which corresponds to your differential equation found in step 2.

5) Solve the difference equation for the extrapolated wavefield.

6) Implement a migration by recursive extrapolation computer algorithm using step 5.

Finite Difference Migration

As an illustration, we will develop the "15 degree" finite difference time migration algorithm which was one of the first to be developed. Recall that in the Fourier domain, time migration by recursive extrapolation proceeds by using the focussing phase shift at each extrapolation step:

$$\mu_f = \frac{\omega \Delta z}{v} \left(\sqrt{1 - \frac{v^2 k_x^2}{\omega^2}} - 1 \right) \quad (1)$$

Since the general form of an extrapolation phase shift is

$$\Phi = k_z \Delta z \quad (2)$$

We seek a "wave equation" whose dispersion relation approximates:

$$k_z = \frac{\omega}{v} \left(\sqrt{1 - \frac{v^2 k_x^2}{\omega^2}} - 1 \right) \quad (3)$$

The second term in the square root is $\sin(\theta)$ where θ is the migration "dip" or scattering angle. We expect that an approximate form for (3) can be obtained by expanding the square root using the binomial theorem and keeping only the first few terms:

$$\sqrt{1 - \frac{v^2 k_x^2}{\omega^2}} = 1 - \frac{v^2 k_x^2}{2\omega^2} + \frac{3v^4 k_x^4}{8\omega^4} + \dots \quad (4)$$

Finite Difference Migration

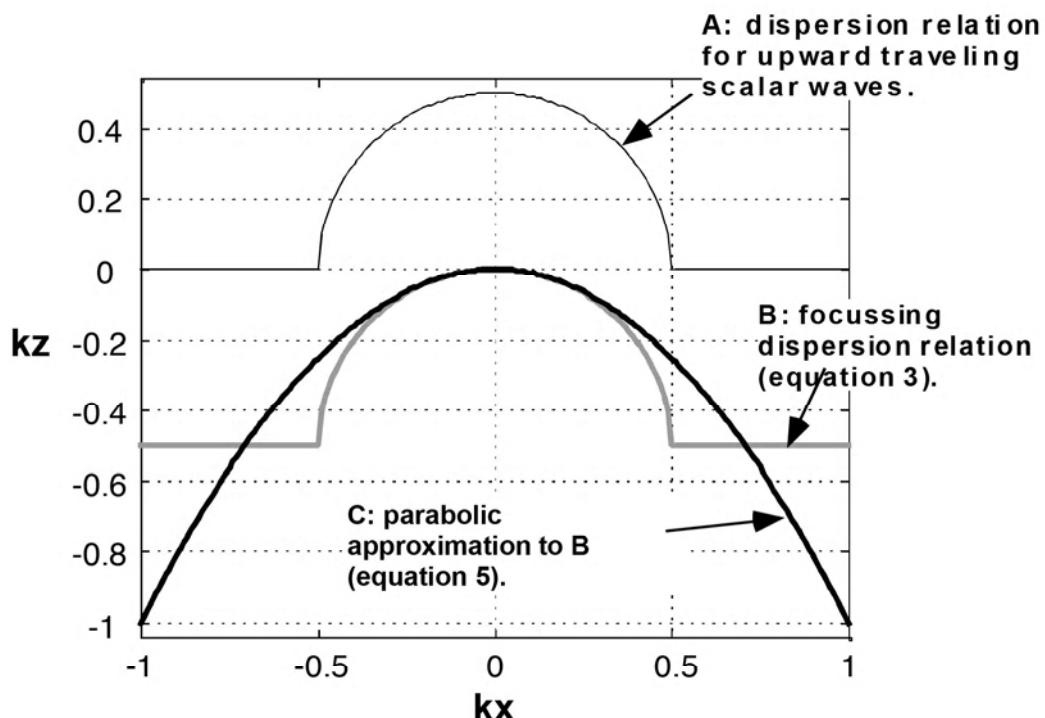
Truncating (4) at two terms and substituting into (3) gives:

$$k_z \approx \frac{\omega}{v} \left(1 - \frac{v^2 k_x^2}{2\omega^2} - 1 \right) = -\frac{vk_x^2}{2\omega} \quad (5)$$

To construct a partial differential equation with this dispersion relation, First multiply both sides by ω/v and then replace the spectral multiplications with partial derivatives as already discussed. The result is:

$$\frac{2}{v} \frac{\partial^2}{\partial z \partial t} \Psi(x, z, t) + \frac{\partial^2}{\partial x^2} \Psi(x, z, t) = 0 \quad (6)$$

Equation 6 is known as the 15° or parabolic wave equation. The reason for the name should be apparent from the figure below:



Finite Difference Migration

Before proceeding with the finite difference approximations, it is appropriate to change variables in (6):

Let: $\tau = z/v$ and $\frac{\partial}{\partial z} = \frac{\partial \tau}{\partial z} \frac{\partial}{\partial \tau} = \frac{1}{v} \frac{\partial}{\partial \tau}$

This leads to:

$$\frac{2}{v^2} \frac{\partial^2}{\partial \tau \partial t} \Psi(x, \tau, t) + \frac{\partial^2}{\partial x^2} \Psi(x, \tau, t) = 0 \quad (7)$$

Equation (7) is an approximate wave equation which can be used to migrate stacked data from zero offset time to migrated time. (Since the dispersion relation concepts which we began with are based on the exploding reflector model, the velocity in (7) is half the real earth velocity.)

The implementation of a finite approximation of (7) is usually done in the time domain and is well treated in the literature (see Claerbout, Fundamentals of Geophysical Data Processing, 1976, page 211). A summary solution procedes like:

Let: $\Delta x^{-2} T = \frac{\partial^2}{\partial x^2}$ and $\Psi_j^k = \Psi(x, k\Delta\tau, j\Delta t)$

Then, if δ symbolizes a finite difference operator, (7) becomes:

$$\delta_\tau \delta_t \Psi_j^k = - \frac{V^2}{2\Delta x^2} T \Psi_j^k \quad (8)$$

Finite Difference Migration

The t difference operator is implemented as:

$$\delta_t (\Psi_{j+1}^k - \Psi_j^k) = -\frac{\Delta t v^2}{4 \Delta x^2} \mathbf{T} (\Psi_{j+1}^k + \Psi_j^k) \quad (9)$$

and then the τ operator as:

$$(\Psi_{j+1}^{k+1} - \Psi_j^{k+1}) - (\Psi_j^{k+1} - \Psi_j^k) = -\frac{\Delta t \Delta \tau v^2}{8 \Delta x^2} \mathbf{T} (\Psi_{j+1}^{k+1} + \Psi_{j+1}^k + \Psi_j^{k+1} + \Psi_j^k) \quad (10)$$

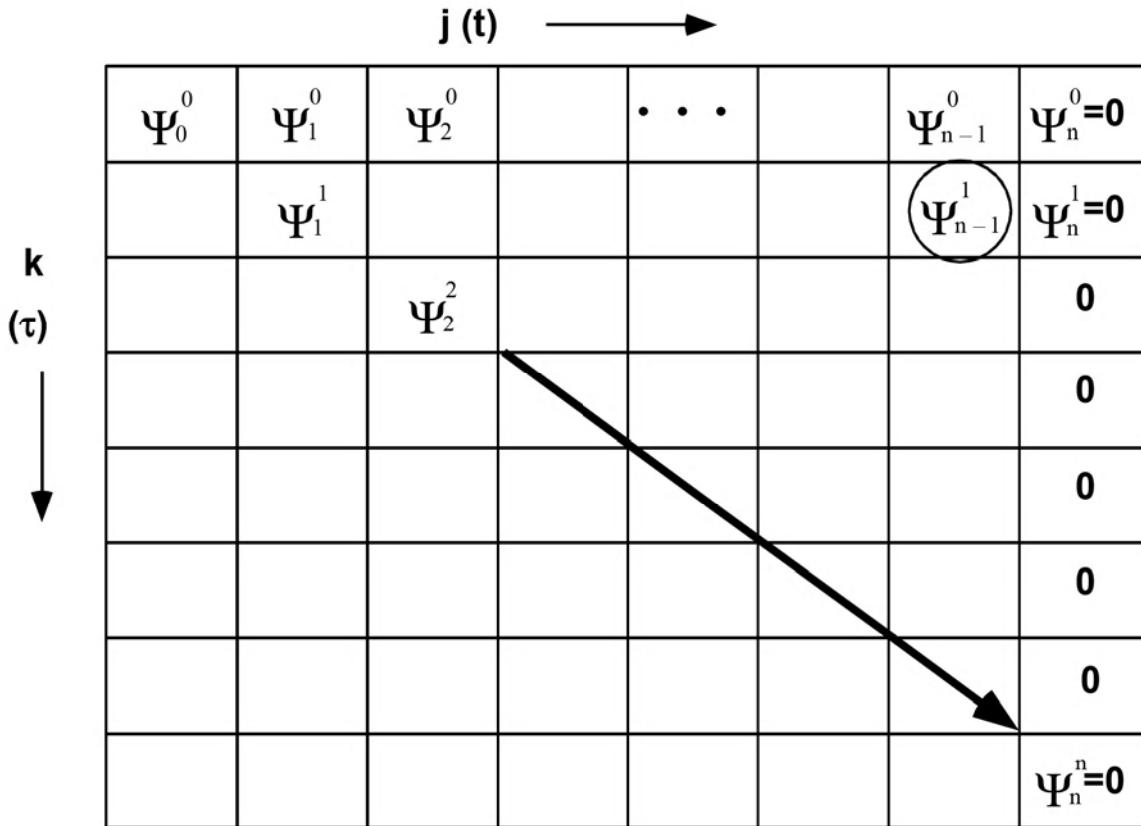
This can be rewritten as:

$$(\mathbf{I} - a \mathbf{T}) \Psi_j^{k+1} = (\mathbf{I} + a \mathbf{T}) (\Psi_{j+1}^{k+1} + \Psi_j^k) - (\mathbf{I} - a \mathbf{T}) \Psi_{j+1}^k \quad (11)$$

where $a = \frac{\Delta t \Delta \tau v^2}{8 \Delta x^2}$

Equation (11) is solved numerically for the unknown Ψ_j^{k+1} where it is assumed that all of the quantities on the right hand side are known. As stated, it is an *implicit* solution because it requires the numerical inversion of the matrix operator $(\mathbf{I} + a \mathbf{T})$.

Finite Difference Migration



The differencing procedure can be viewed on a (j, k) grid where the first row is specified as the cmp stack and the last column is set to zero (this is similar to a zero pad around the data in a Fourier migration). The process begins by computing

Ψ_{n-1}^1 given the values of Ψ_{n-1}^0 , Ψ_n^0 and Ψ_n^1 using equation (11). Essentially, knowing the three upper-right elements in a 2×2 cell allows the fourth (lower left) to be computed. Each row (a $\Delta\tau$ step) is completely computed before the next row is begun. The final migrated section lies on the $t=\tau$ diagonal which is the imaging condition for a time migration algorithm.

Finite Difference Migration

From this example, many of the well known characteristics of finite difference migration algorithms are easily deduced such as:

- They are recursive extrapolation schemes.
- They use an approximate form of the exact dispersion relation for scalar waves. This means that they are all dip or scattering angle limited.
- They are very flexible in handling velocity variations since the velocity may simply be varied with each grid cell.
- The use of finite difference derivatives means that even the approximate dispersion relation is not exactly realized.
- Finite difference derivatives are not unique and the accuracy of a method depends strongly on the sophistication of the approximations used. Generally, finite difference methods need many more samples per wavelength than Fourier methods (6-10 versus 2-3).
- They can be either time or depth migrations.
- They can be posed in the space-time or space-frequency domain.

Methods of Seismic Data Processing

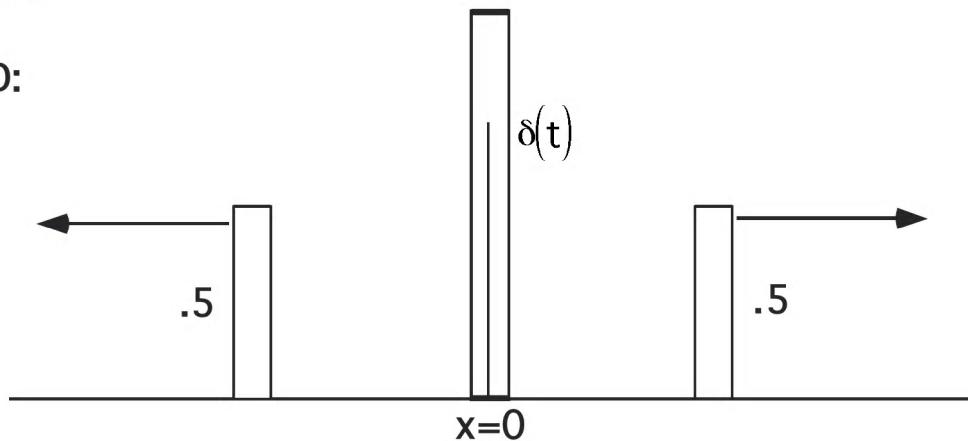
**Lecture Notes
Geophysics 557**

**Chapter 9
The Third Dimension**

Impulse Responses

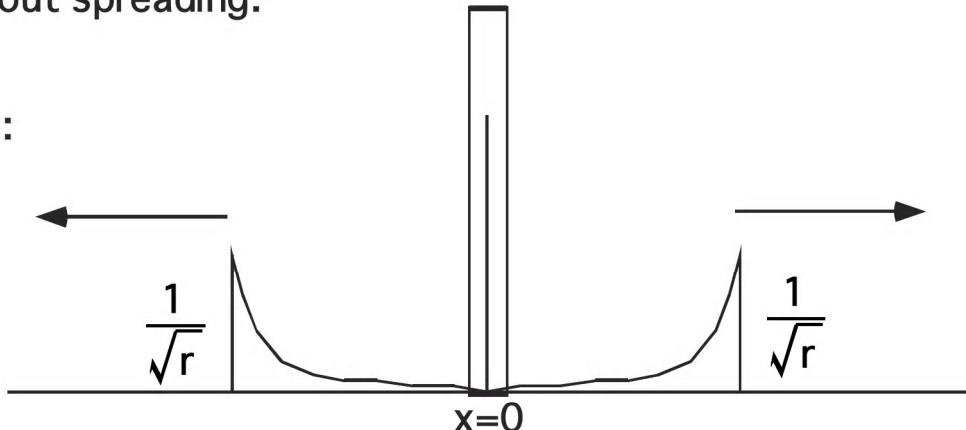
The scalar wave equation is the simplest mathematical model of wave motion. More realism is obtained through elastic or viscoelastic wave methods but the scalar wave equation delivers much insight. Of first order importance are the Green's function solutions, or impulse responses, which are known exactly for constant velocity in 1, 2, and 3-D. Suppose that a unit height impulse is applied at $t=0$ in each of these situations. Then:

1-D:



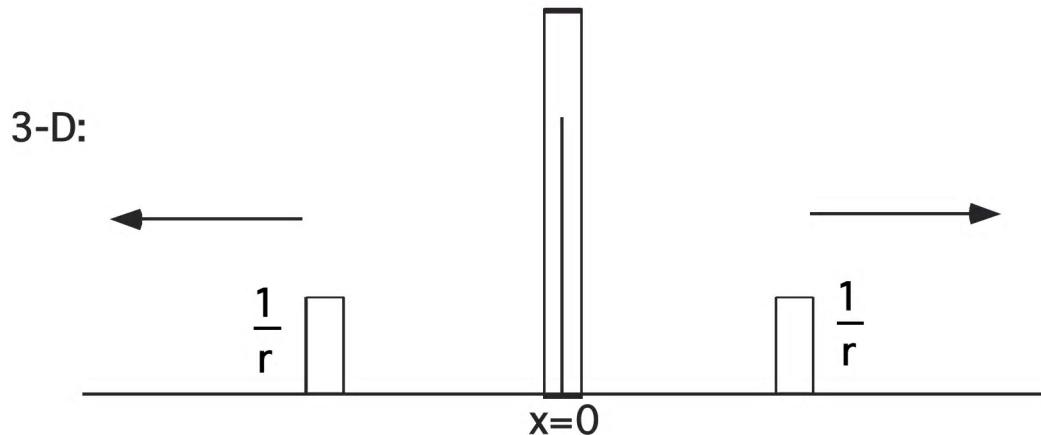
The 1-D solution propagates two half-height impulses along a line without spreading.

2-D:



The 2-D solution propagates a circular impulse with amplitude decay inversely proportional to the square root of r . Note the presence of the wake following the impulsive wave front.

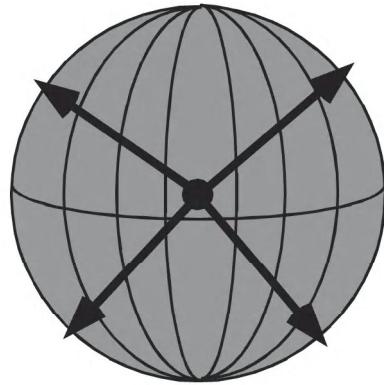
Impulse Responses



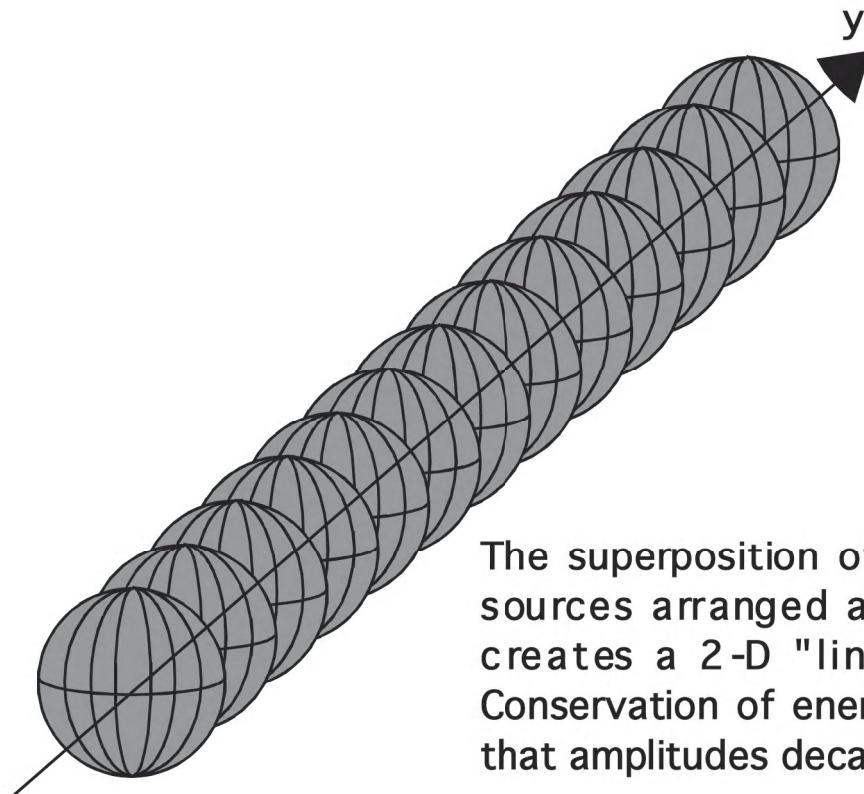
The 3-D solution propagates a spherical-shell impulse with amplitude decay of inversely proportional to r . Note the absence of any wake following the wavefront.

- The 2-D solution can be obtained from the 3-D wave equation as then response of a line source. Mathematically, this requires an integration in the y direction.
- The 1-D solution can be obtained from the 3-D wave equation as then response of a plane source. Mathematically, this requires an integration in the x and y directions.

Impulse Responses

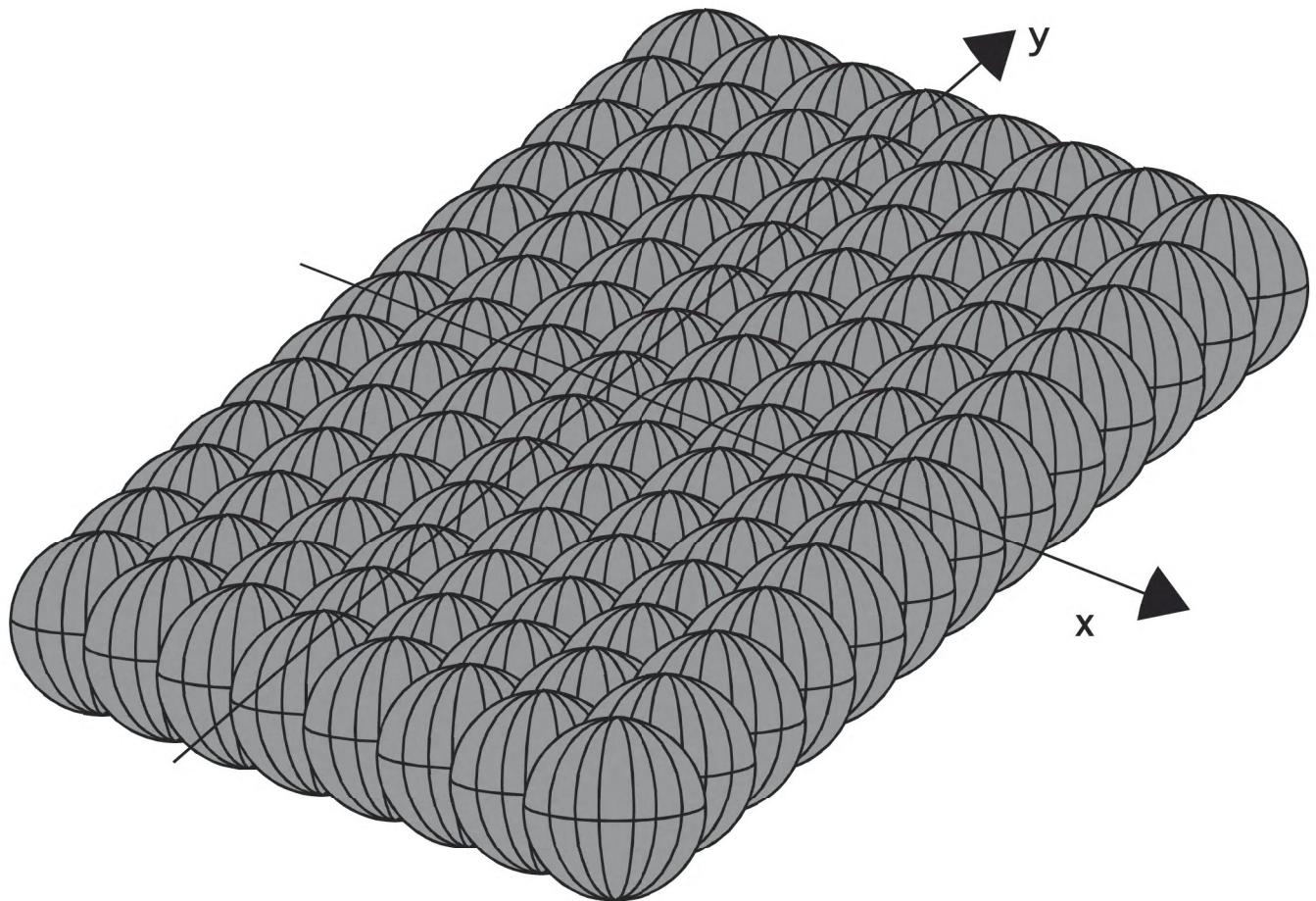


A 3-D wavefront expands as a spherical shell. Conservation of energy requires that amplitudes decay as $1/r$.



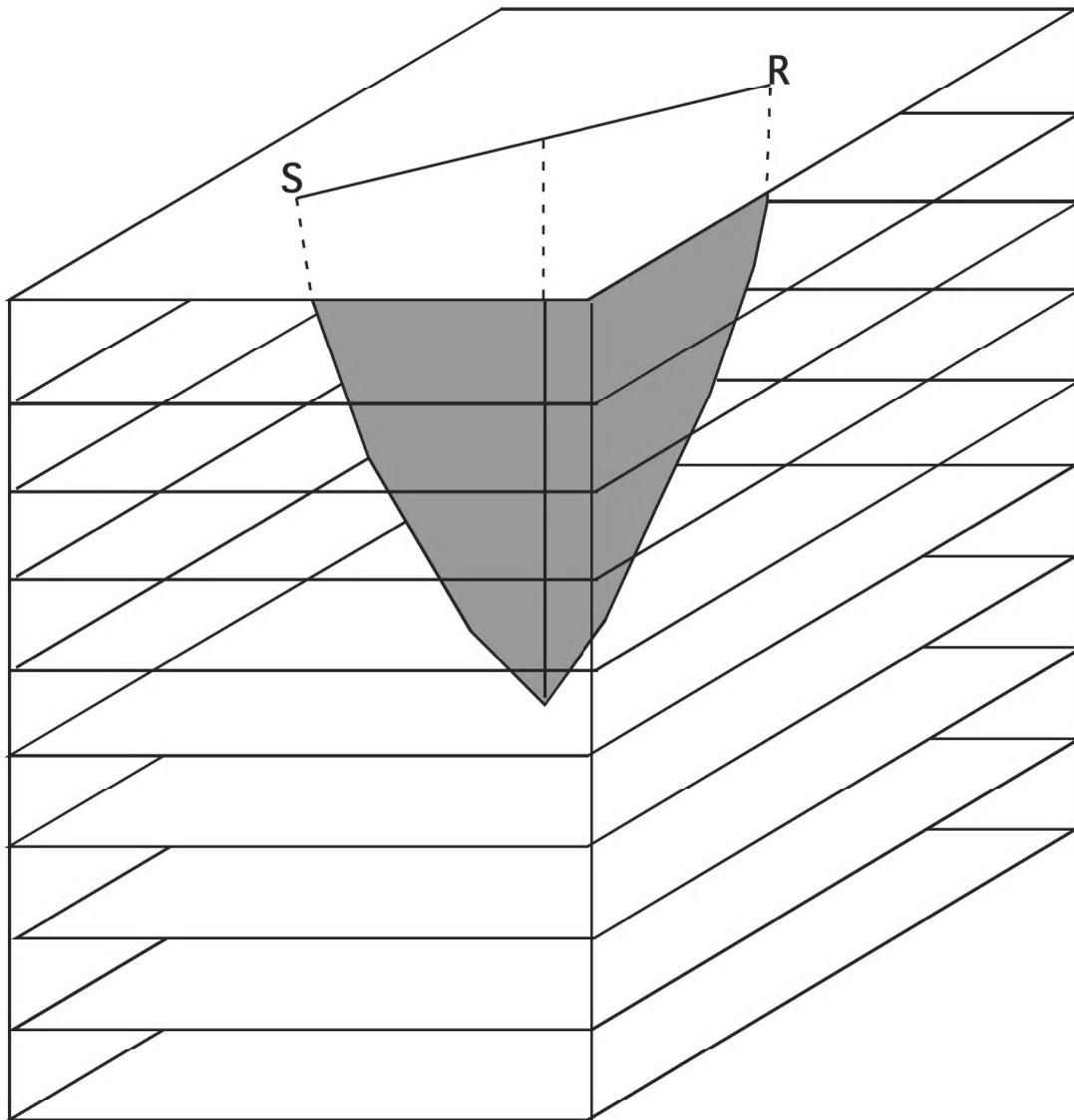
The superposition of many 3-D sources arranged along a line creates a 2-D "line source". Conservation of energy requires that amplitudes decay as $r^{-1/2}$

Impulse Responses



The superposition of many 3-D sources arranged along a plane creates a 1-D "plane source". Conservation of energy requires that amplitudes do not decay.

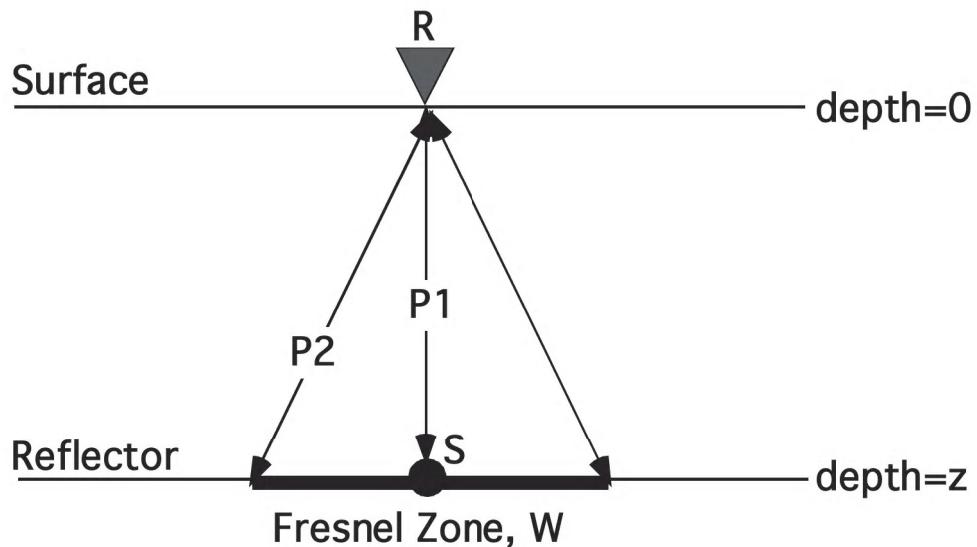
Wave Propagation



In a stratified (i.e. no lateral variations) 3-D earth, it can be shown that the raypath from source to receiver lies in a single vertical plane. In the elastic case, a p-wave propagating in this plane can excite only one mode of s-wave. Called an Sv wave, it has particle motion confined to the same vertical plane. Sh waves have particle motion orthogonal to this plane. A reasonable inference is that, in a typical sedimentary basin, these effects are nearly true.

Fresnel Zones

Consider the zero offset scattering of waves from a point in the subsurface. High frequency raytracing (e.g. Snell's Law) predicts that the receiver, R, will only record information from the infinitesimal point, S. However, this is only a very gross approximation to the behavior of real waves. If a vertically travelling, monochromatic wave, of frequency f , strikes the reflector, it is actually "backscattered" in all directions and from all points.



The receiver at R will respond to the total backscattered wavefield whose amplitude will be the sum of all individually scattered wavefields. Since waves interfere both constructively and destructively, we seek a measure of the width, W , for energy arriving at R is "in phase". This is usually called the Fresnel Zone and is defined by the requirement that the difference in lengths of the paths P1 and P2 be one-quarter wavelength (assuming constant velocity). Thus:

$$\sqrt{z^2 + \left(\frac{W}{2}\right)^2} - z = \frac{\lambda}{4}$$

or, solved for W ,

$$W = 2\sqrt{\left(\frac{\lambda}{4} + z\right)^2 - z^2}$$

Fresnel Zones

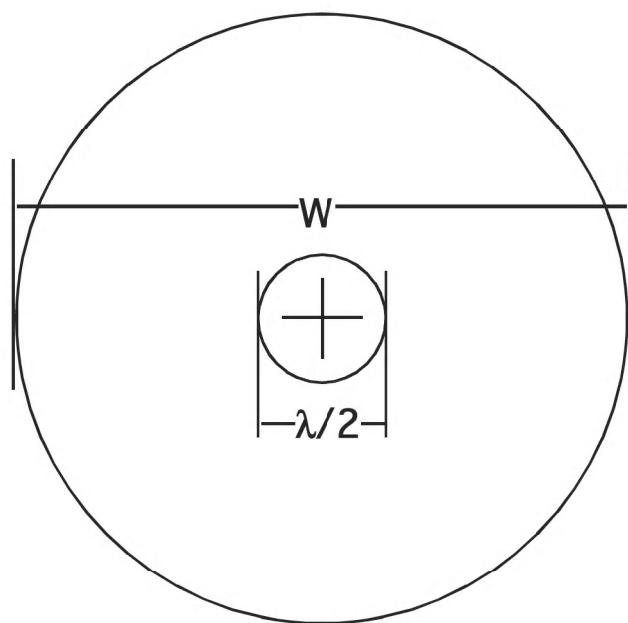
Other, alternate expressions for W are:

$$W = \sqrt{\frac{\lambda^2}{4} + 2z\lambda} = \sqrt{\frac{v^2}{4f^2} + 2z\frac{v}{f}}$$

$$W \approx \sqrt{2z\lambda} = \sqrt{2z\frac{v}{f}} \quad \text{for } z \gg \lambda$$

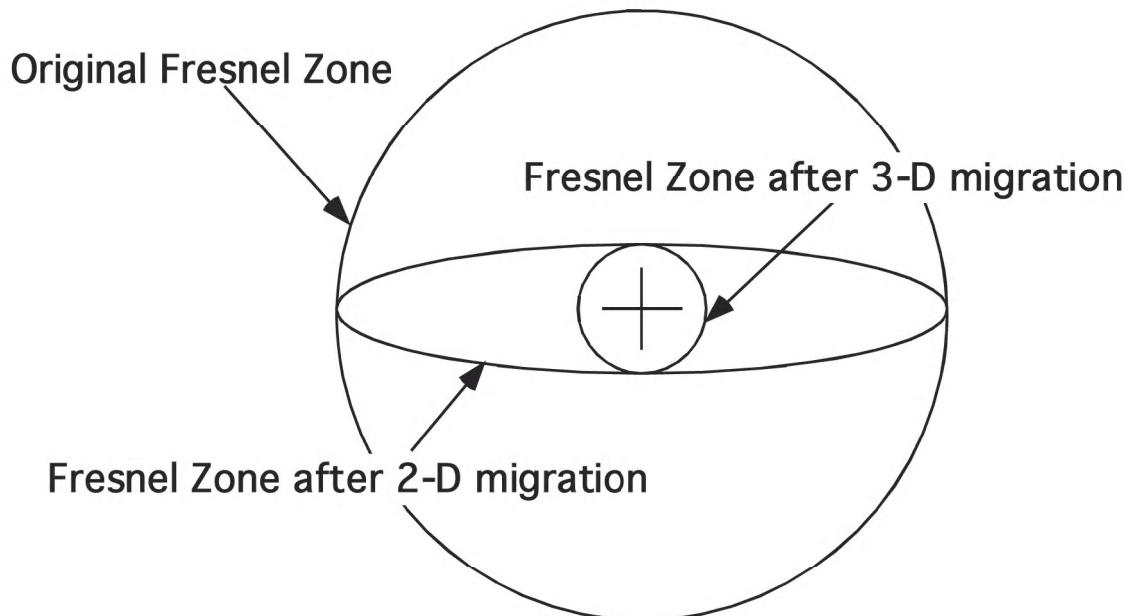
Migration, which can be conceptualized as lowering the receiver to the point S , can be said to shrink the Fresnel zone to its theoretical minimum of $\lambda/2$. (Note that it does not shrink to zero as is sometimes erroneously stated when the approximate expression for W is used.)

In 3-D, the Fresnel zone is a disk which defines the inherently unfocused nature of stacked data.

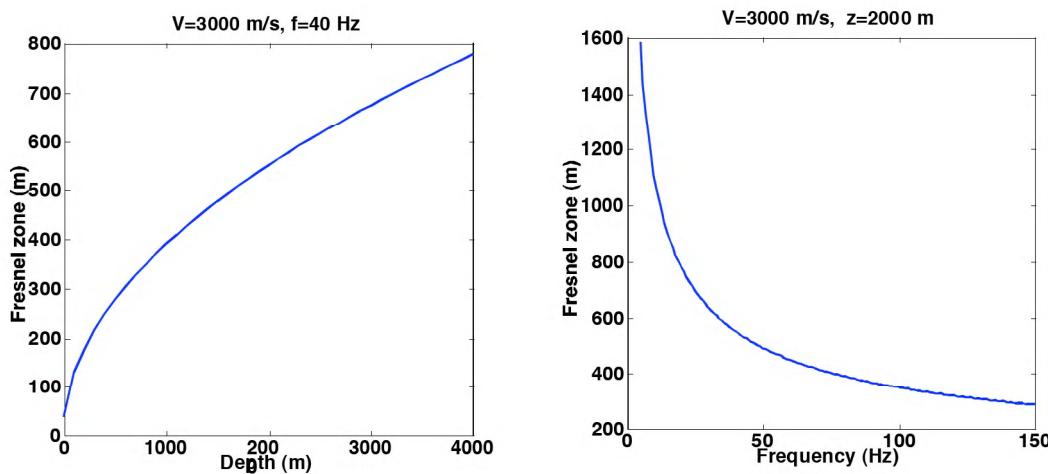


Fresnel Zones

Real earth seismic data is always 3-D regardless of whether the acquisition geometry is 2-D or 3-D. Thus, when 2-D migration is run, the Fresnel zone is only collapsed in the inline direction.

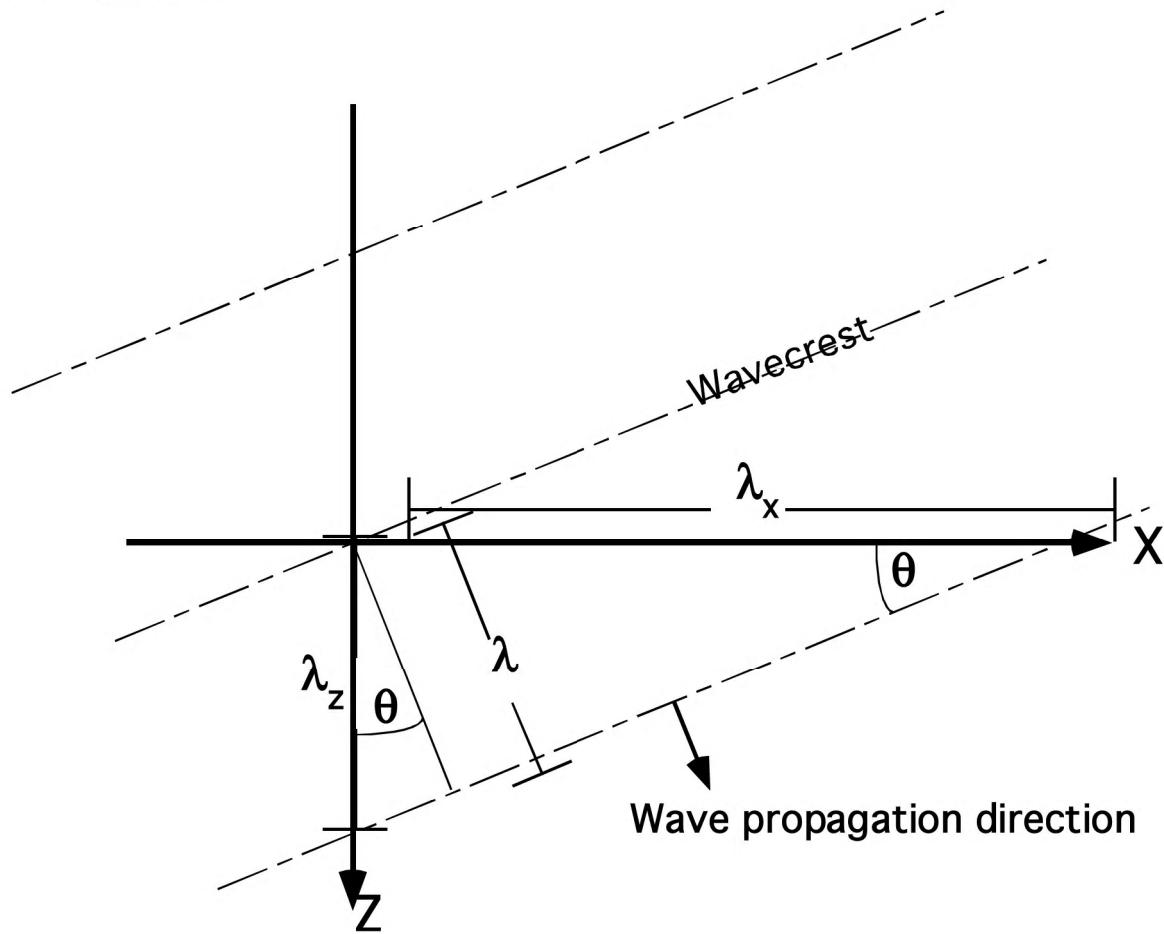


This is one of the most compelling justifications for 3-D imaging techniques, even in the case of sedimentary basins with flat-lying structures. A 2-D migration of a seismic line must be considered as giving the average reflectivity over a Fresnel zone width in the cross line direction.



Wavelength Components

Consider a series of planar wavefronts propagating as shown below.



The distance between wavefronts, measured perpendicular to them, is defined as the wavelength, λ . We can also speak of the wavelength "components" in the various coordinate directions. For example, the horizontal wavelength, λ_x , is the distance between wavefronts measured in the x coordinate direction. Thus:

$$\lambda_x = \frac{\lambda}{\sin(\theta)} \quad \text{and} \quad \lambda_z = \frac{\lambda}{\cos(\theta)}$$

Wavelength Components

We see that the components of wavelength are never less than the wavelength itself. In fact, for a vertically traveling wave, λ_x is infinite. The components add as inverse squares:

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_x^2} + \frac{1}{\lambda_z^2}$$

It is often convenient to deal with vector quantities so we define the wavenumber, k , and its components as the inverse of the wavelength and its components.

$$k = \lambda^{-1} \quad k_x = \lambda_x^{-1} \quad k_z = \lambda_z^{-1}$$
$$k^2 = k_x^2 + k_z^2$$

k is the magnitude of a vector, \underline{k} , which points in the direction of wave propagation and whose components are the inverse wavelengths.

In 3-D, we have planar wavefronts instead of linear but a simple extension of this result still holds:

$$k_x^2 + k_y^2 + k_z^2 = k^2 \text{ The dispersion relation for scalar waves.}$$

Where:

$$k_y = \lambda_y^{-1}$$

Wavelength Components

This geometric relation between components of the wavenumber vector is fundamental to the study of wave propagation. It can be considered as the Fourier domain equivalent of the scalar wave equation. A fundamental result from theory is that the extrapolation of surface recorded data into the subsurface (z direction) requires knowledge of k_z . On the surface, we can measure k_x , k_y , and f , and since $f=vk$, this allows k_z to be calculated from the dispersion relation:

$$k_z = \sqrt{\frac{f^2}{v^2} - k_x^2 - k_y^2}$$

Since k_z must be a real number (in order to be interpreted as an inverse wavelength) we see from this equation another fundamental result. Not all values of (k_x, k_y, f) can be considered as wavelike. In fact, we must have

$$\frac{f^2}{v^2} \geq k_x^2 + k_y^2$$

in order for a triplet of (k_x, k_y, f) to be a propagating wave.

Apparent Velocity (or phase velocity)

The wavelength components and the corresponding wavenumbers are closely related to the wave velocity and its components which are called apparent velocities. Recalling the basic relation, $\lambda f = v$, we see that the addition formula for wavelength components:

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_x^2} + \frac{1}{\lambda_z^2}$$

leads directly to:

$$\frac{1}{v^2} = \frac{1}{v_x^2} + \frac{1}{v_z^2}$$

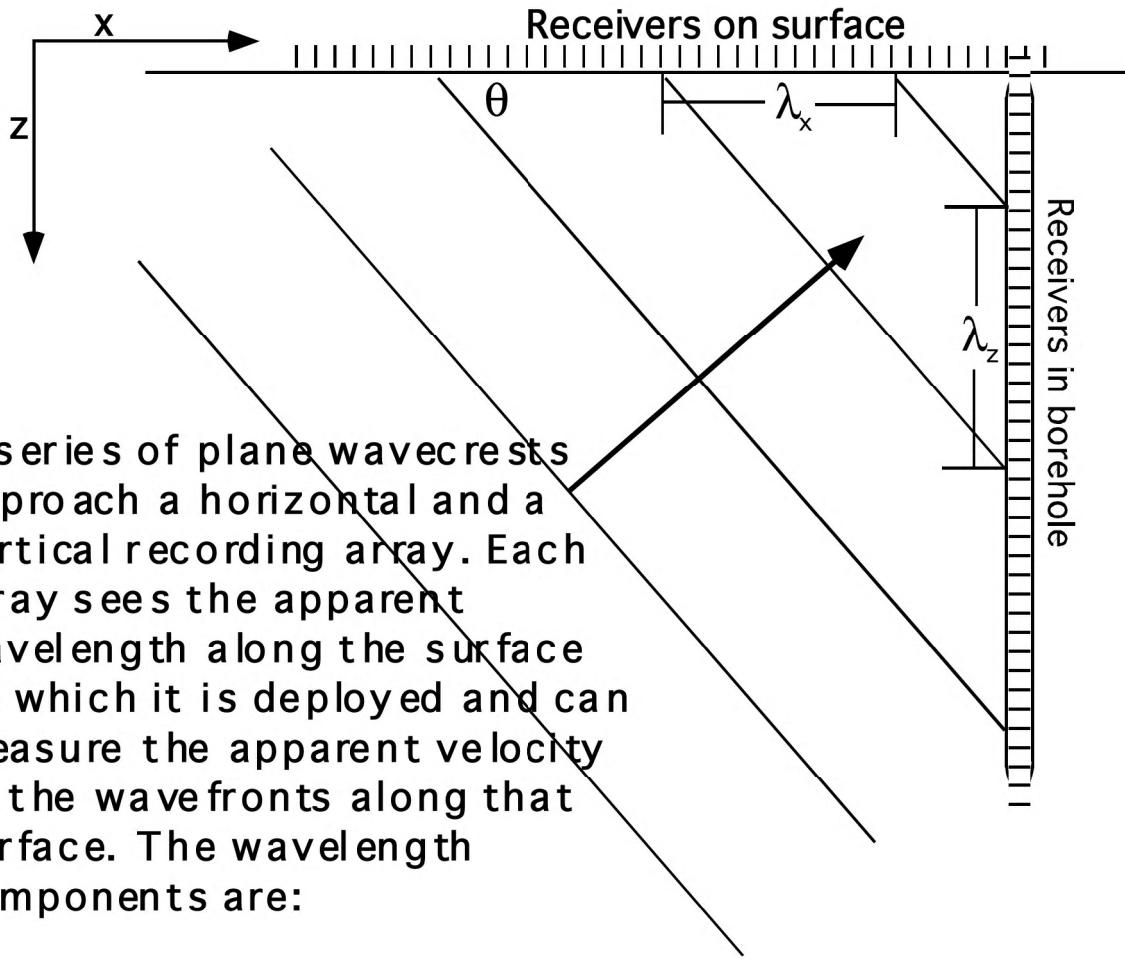
$$\text{where } v_x = f\lambda_x \quad \text{and} \quad v_z = f\lambda_z$$

If we use wavenumber components, we have:

$$v = \frac{f}{k} \quad v_x = \frac{f}{k_x} \quad v_z = \frac{f}{k_z}$$

Nothing physical actually propagates at any of the apparent velocities. Rather, they are simply related to the arbitrary choice of coordinate directions and can be visualized as the wavelength along a coordinate direction divided by the time between wavecrests (i.e. the period of the waves.)

Apparent Velocity (or phase velocity)



$$\lambda_x = \frac{\lambda}{\sin(\theta)} \quad \text{and} \quad \lambda_z = \frac{\lambda}{\cos(\theta)}$$

And the apparent velocities are:

$$v_x = \frac{\Delta x}{\Delta t} = f\lambda_x = \frac{f\lambda}{\sin(\theta)} = \frac{v}{\sin(\theta)}$$

similarly $v_z = \frac{\Delta z}{\Delta t} = f\lambda_z = \frac{f\lambda}{\cos(\theta)} = \frac{v}{\cos(\theta)}$

The F-K Transform

The f-k transform is a fundamental tool which essentially allows the direct computation of wavenumber components and frequency for a multidimensional wavefield. In 2-D, it can be written:

$$\varphi(k_x, f) = \iint_{-\infty}^{\infty} \psi(x, t) e^{2\pi i (k_x x - f t)} dx dt$$

and in 3-D

$$\varphi(k_x, k_y, f) = \iiint_{-\infty}^{\infty} \psi(x, y, t) e^{2\pi i (k_x x + k_y y - f t)} dx dy dt$$

The inverse transforms are mathematically similar:

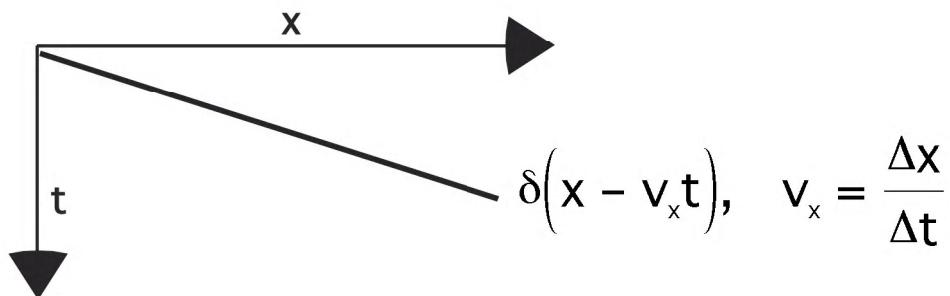
$$\psi(x, t) = \iint_{-\infty}^{\infty} \varphi(k_x, f) e^{-2\pi i (k_x x - f t)} dk_x df \quad \text{2-D}$$

$$\psi(x, y, t) = \iiint_{-\infty}^{\infty} \varphi(k_x, k_y, f) e^{2\pi i (k_x x + k_y y - f t)} dk_x dk_y df \quad \text{3-D}$$

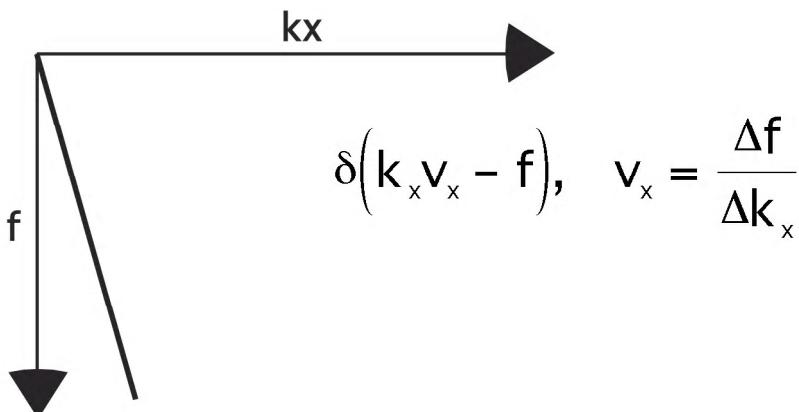
These inverse transforms have the physical interpretation of presenting a wavefield as a superposition of individual Fourier components or "plane waves".

The F-K Transform

A few f-k transforms are known analytically. Perhaps the most important is the transform of a single linear event. Using the mathematics of Dirac delta functions, a seismic wavefield consisting of an isolated linear event can be written:



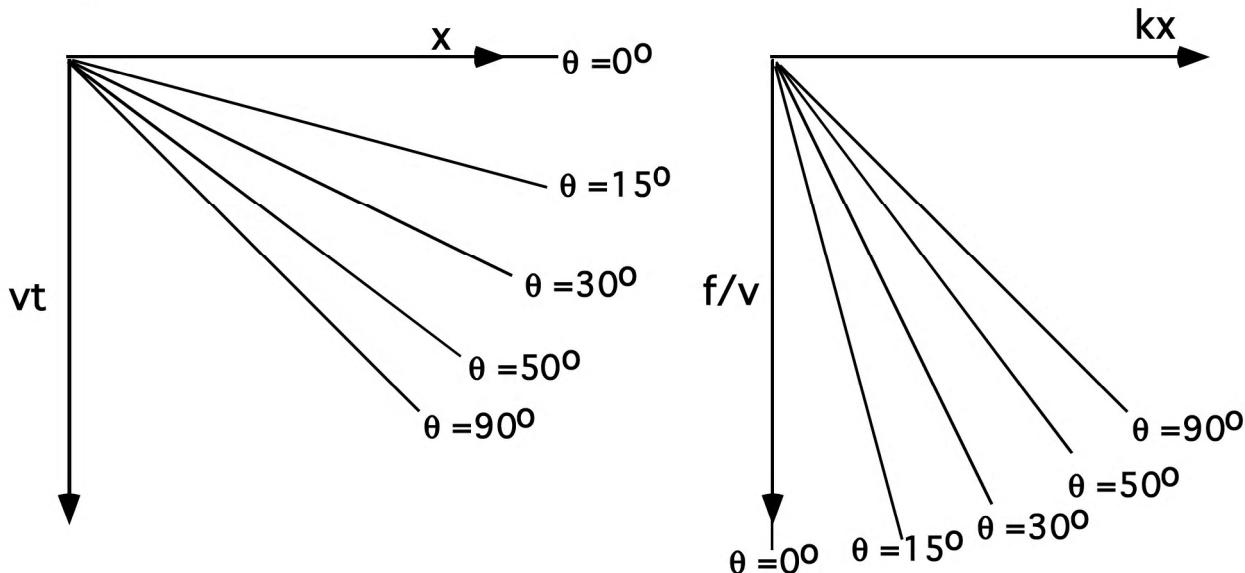
$$\begin{aligned}\varphi(k_x, f) &= \iint_{-\infty}^{\infty} \delta(x - v_x t) e^{2\pi i (k_x x - f t)} dx dt \\ &= \int_{-\infty}^{\infty} e^{2\pi i (k_x v_x - f)t} dt = \delta(k_x v_x - f)\end{aligned}$$



- Horizontal events in (x,t) are vertical in (kx,f) and vice-versa.
- All events in (x,t) with the same apparent velocity, v_x , are collected into a single linear event in (kx,f). The different events are distinguished by their phase spectra but have differing phase spectra.

The F-K Transform

If we consider all possible linear events characterized by $v_x = v/\sin(\theta)$, then we have:

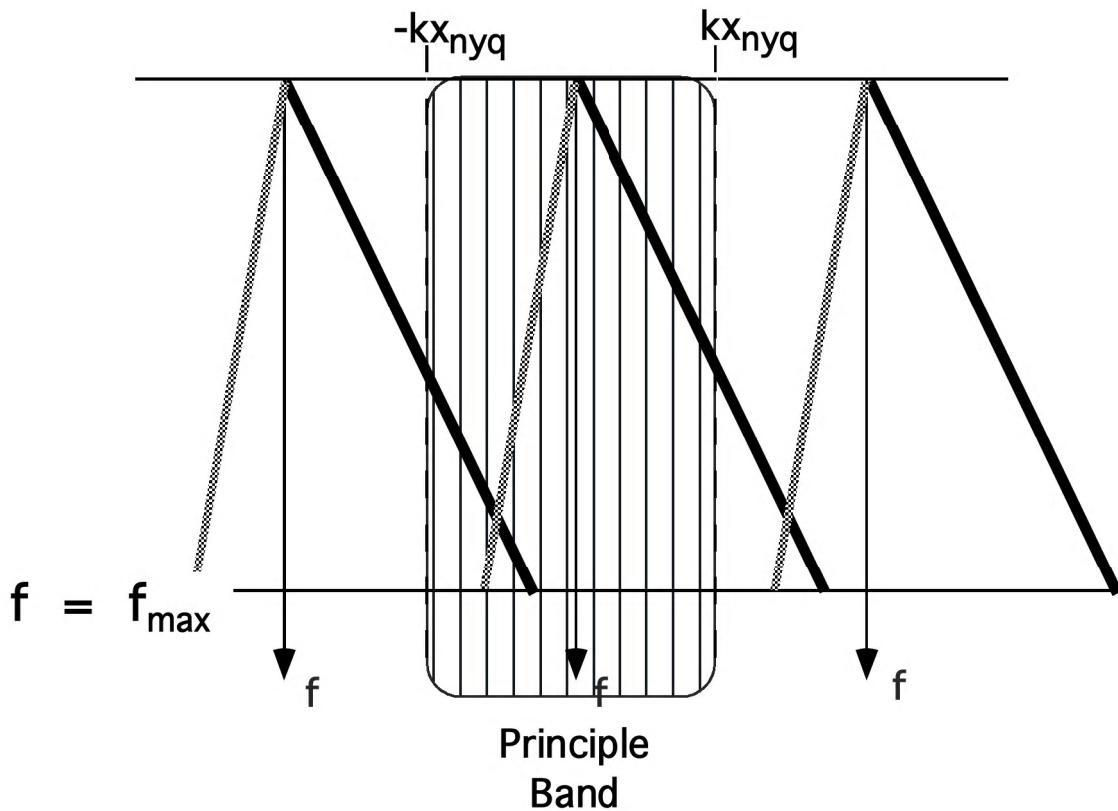


The previously encountered fact that $f/v > kx$ is reexpressed in this analysis through the fact that a portion of (kx, f) space is not populated. As a general rule of thumb, we see that large kx values are only wavelike at high frequencies. This fact will turn out to be fundamental in describing the ability of seismic images to resolve small features. Small features require large kx values which, in turn, require a large temporal frequency bandwidth.

- In proceeding from analytic to discrete f-k transforms, it turns out that the implementations of the Fourier transform integrals are approximate but the forward and backward DFKT are exact inverses of each other. This fact is a great convenience in data processing and is not generally true of other transforms such as the Radon transform.

The F-K Transform

When we proceed from the continuous F-K transform to the discrete, a situation directly analogous to the 1-D case occurs. That is, the act of spatial sampling induces a periodicity in the (k_x, f) domain.

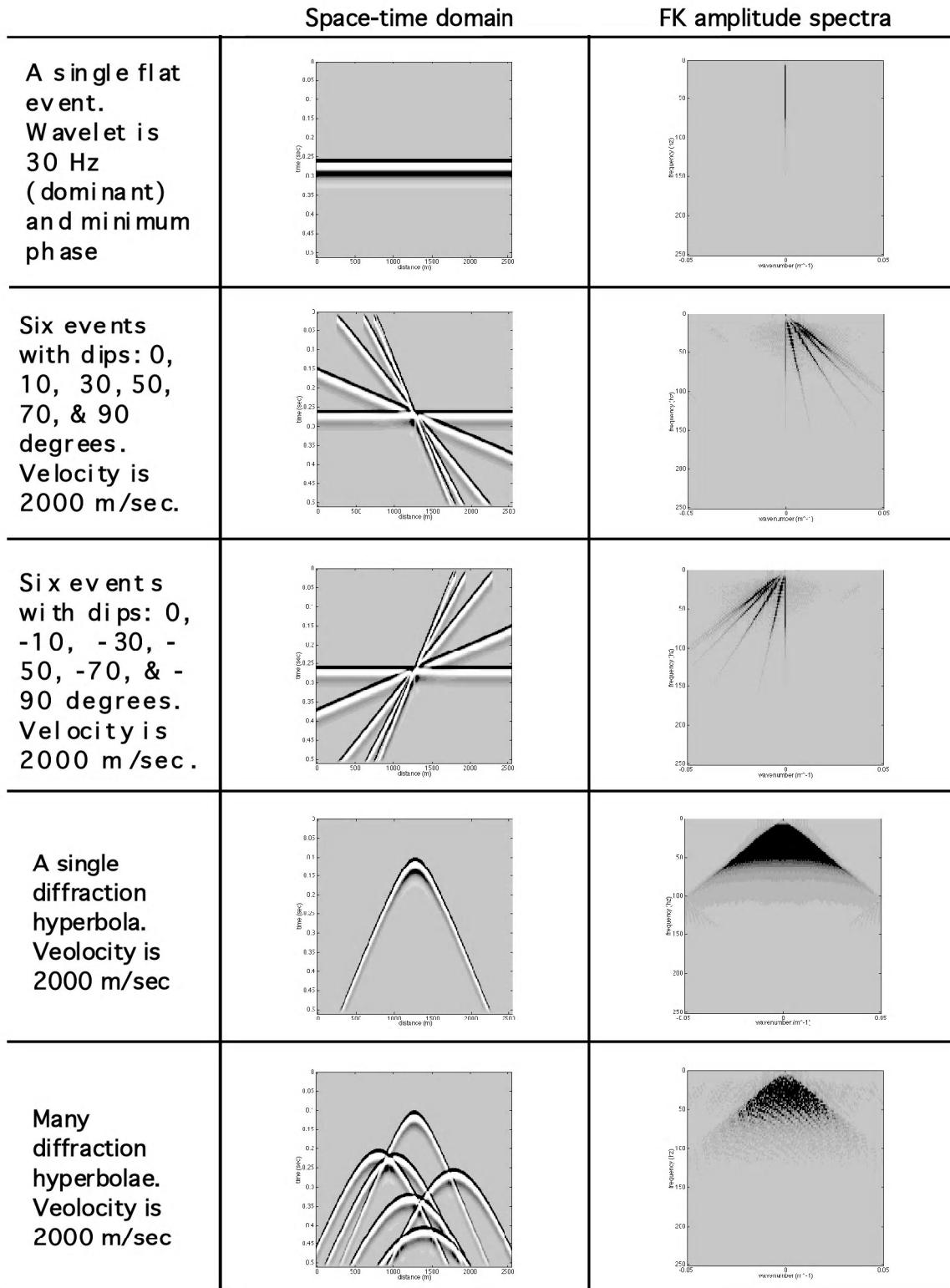


Here we see one event showing spatial aliasing and another which does not. Given a spatial sample rate of Δx and an apparent velocity v_a then all temporal frequencies higher than:

$$f_{\text{crit}} = 2\pi\omega_{\text{crit}} = v_a k_{\text{nyquist}} = \frac{v_a}{2\Delta x}$$

will be spatially aliased. For excellent illustrations of spatial aliasing see Hatton et al. pp43-45 and Yilmaz pp62-69

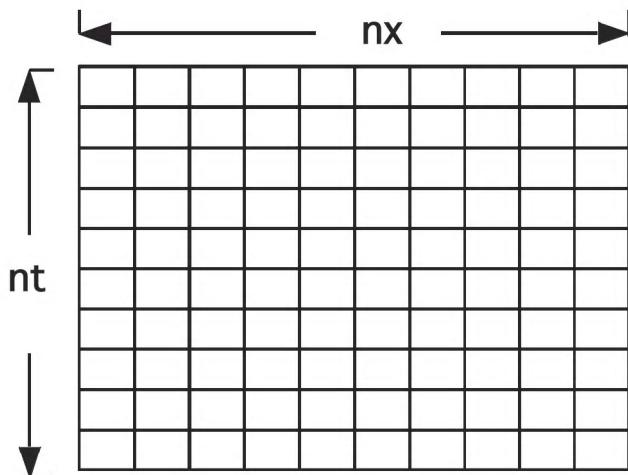
FK Transform Pairs



FK Transform Computation

The computation of an f-k transform is a fundamental one in seismic data processing and can be used as a conceptual device to appreciate the relative costs of 1-D, 2-D, and 3-D algorithms.

A 2-D f-k transform is generally computed by a cascade of 1-D transforms. The seismic wavefield can be considered as a matrix of nt rows (samples per trace) by nx columns (traces).



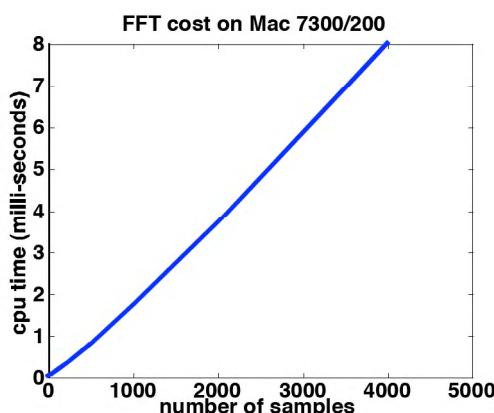
Then a 2-D f-k transform can be described as:

2-D f-k transform =

$$nx * \text{fft}(nt) + nt * \text{fft}(nx)$$

In this expression, $\text{fft}(n)$ refers to a 1-D fft (fast Fourier transform) of length n.

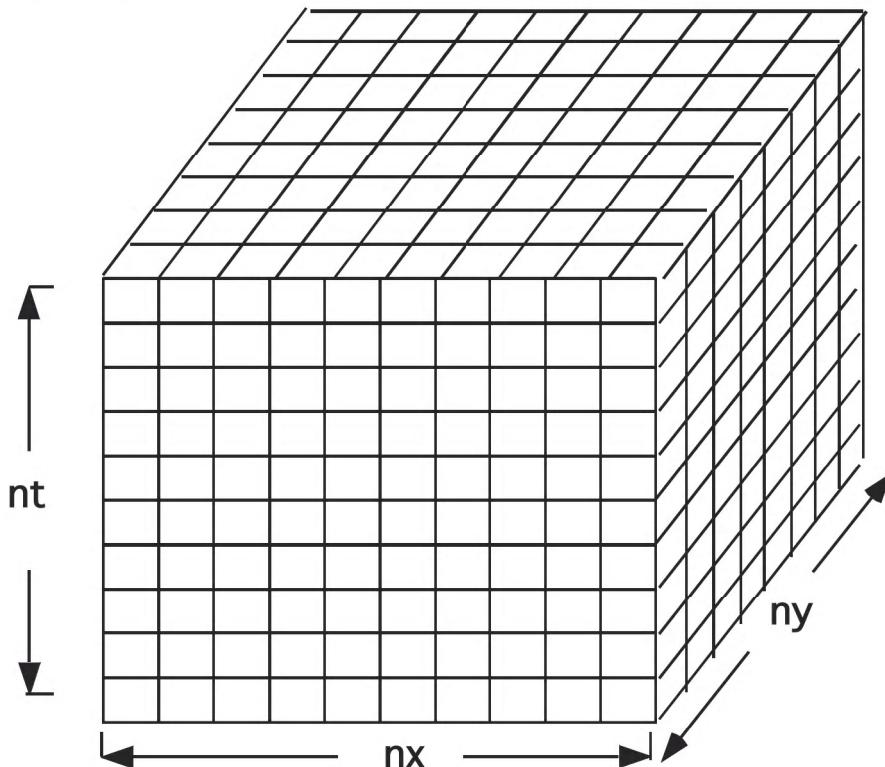
On a typical computer, an fft algorithm achieves a run time proportional to $n * \log(n)$ and specific performance on the author's Macintosh is given below.



From this chart, we see that a 1024 length fft takes about .002 seconds of cpu time. Thus a 2-D f-k transform of 1024 samples by 1024 traces would take about $1024 * (.002) + 1024 * (.002)$ seconds or about 4 seconds.

FK Transform Computation

A 3-D f-k transform is similarly computed by a cascade of 1-D transforms. Now let the seismic wavefield be a 3-D matrix of nt rows (samples per trace) by nx by ny.



Then a 3-D f-k transform can be described as:

$$\begin{aligned} \text{3-D f-k transform} &= nx * ny * \text{fft}(nt) + nt * ny * \text{fft}(nx) + \\ &\quad nt * nx * \text{fft}(ny) \\ &= ny * (nx * \text{fft}(nt) + nt * \text{fft}(nx)) + nt * nx * \text{fft}(ny) \\ &= ny * (\text{2-D f-k transform}) + nt * nx * \text{fft}(ny) \end{aligned}$$

So, assuming 1024x1024x1024 as a dataset size, an estimate for the 3-D f-k transform is $(1024 * 4 + 1024 * 1024 * (.002))$ seconds which works out to about 6200 seconds which is about 1.7 hours. This is a vast increase over the 2-D case.

FK Transform Computation

Computer memory requirements are similarly huge for 3-D. Assuming that each seismic sample is stored in single precision (4 bytes) the 2-D data matrix requires storage of $4*1024*1024$ bytes which is about 4 megabytes (mb). The 3-D case requires about $4*1024^3$ bytes or about 4000 mb. Clearly the 3-D transform cannot be done in-core on most computers.

Some obvious conclusions from this analysis are:

- 3-D methods require vastly greater resources than 2-D.
- There is much to be gained from controlling the size of 3-D surveys.
- It is understandable that many approximate 3-D schemes have been developed.
- Do not attempt a 3-D migration on a Macintosh.

Though your migration algorithm of choice may not be phase shift, it is still reasonable to expect performance that is proportional to these calculations.

Finally, a common migration algorithm is the phase-shift method of Gazdag (1978, Geophysics). This method requires about 1 f-k transform per depth step and, assuming the number of depth steps will also be around 1024, we can estimate run times for 2-D and 3-D phase shift as:

2-D phase shift: $1024 * 4$ seconds ~ 4000 seconds

3-D phase shift: $1024 * 1.7$ hours ~ 72 days

Exercises:

Name_____

A) Suppose your seismic survey is recorded at 2ms, with a 4 second record length and covers 64 kmsq. If a 20x20m bin size is used for stacking, estimate the run time for a post-stack 3-D phase-shift migration on the author's Macintosh. (Assume 500 depth steps are required).

B) What would be the run time for a double 2-D migration?. In this scheme, each inline is migrated in 2-D and then each crossline. Thus $nx+ny$ 2-D migrations are done. What might be the major reason that double 2-D was preferred by seismic contractors during the early development of 3-D software?

3-D Migration by Double 2-D

3-D migration became computationally feasible when Stolt published his f-k migration technique (Geophysics 1978). However, as the exercise on the previous page demonstrates, 3-D migrations can require a lot of resources. A common approximate scheme to attempt to save computer costs is the double 2-D approach:

- Migrate all inlines in 2-D and express the results in migrated time.
- Sort data into crosslines
- Migrate all crosslines in 2-D.

Perhaps surprisingly, this scheme has been shown to be exactly equivalent to 3-D migration for constant velocity. While not a proof, the following argument shows why this works:

For a 3-D migration, the k_z wavenumber is given by:

$$k_z = \sqrt{\frac{f^2}{v^2} - k_x^2 - k_y^2}$$

Now, let's see what results from the double 2-D approach.

Inline migration: These 2-D migrations result in a set of inlines with a vertical wavenumber:

$$k_z' = \sqrt{\frac{f^2}{v^2} - k_x^2}$$

Or, if these lines are output with a vertical time scale:

$$F' = v k_z' = \sqrt{f^2 - v^2 k_x^2}$$

3-D Migration by Double 2-D

Crossline migration: After crossline migration, the final vertical wavenumber is:

$$k_{z-2Dx2D} = \sqrt{\frac{F'^2}{v^2} - k_y^2}$$

When the expression for F' is substituted into this, we obtain:

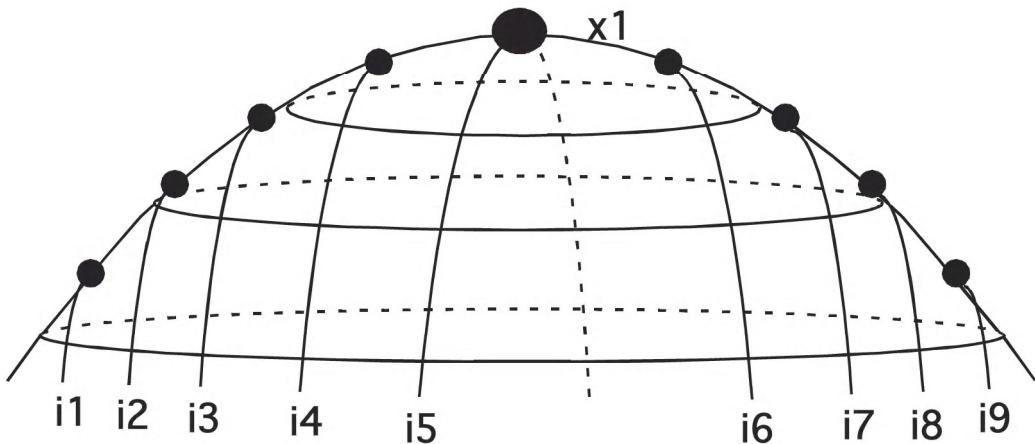
$$k_{z-2Dx2D} = \sqrt{\frac{1}{v^2} (f^2 - v^2 k_x^2) - k_y^2}$$

or

$$k_{z-2Dx2D} = \sqrt{\frac{f^2}{v^2} - k_x^2 - k_y^2}$$

Which is the correct 3-D result.

3-D Migration by Double 2-D

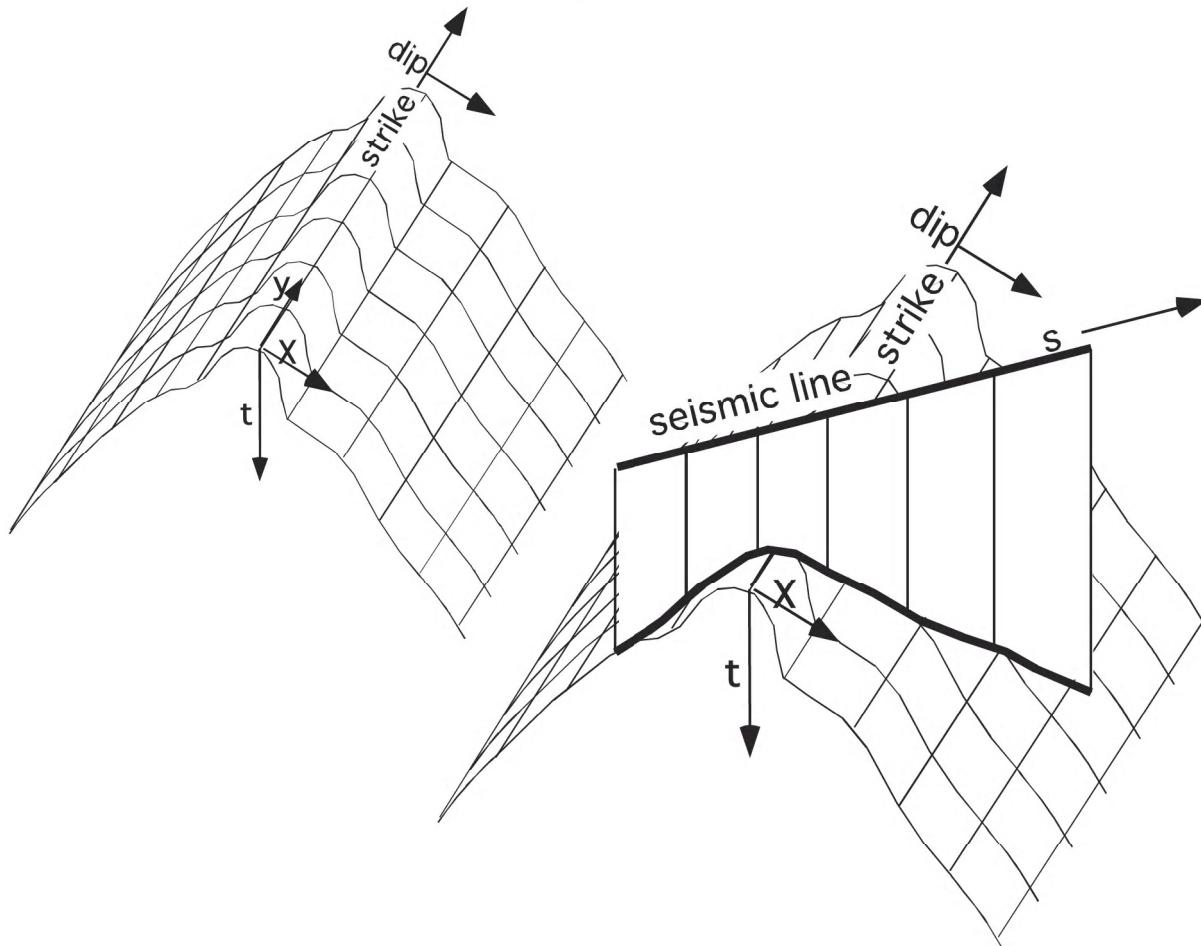


This diagram shows a 3-D diffraction surface for zero-offset and constant velocity. It is a hyperboloid which is a 2-D hyperbola rotated about a vertical axis through its apex. It can be shown that the intersection of this hyperboloid with any vertical plane is a 2-D hyperbola. The curves labeled i_1 through i_9 all represent where inline sections of a 3-D survey intersect the hyperboloid. Since constant velocity is presumed, each of these inlines will contain a hyperbola which will focus at the black dots with a constant velocity 2-D migration. All of these focal points will lie in a single crossline and will focus at the big dot in the crossline migration.

Though exact for constant velocity, the process breaks down rapidly in realistic settings. Even with only vertical velocity variations, $v=v(z)$, the process is worse than a 3-D time migration with a V_{rms} assumption because a systematic velocity error is made. In the true 3-D case, the entire hyperboloid is migrated with $V_{rms}(t_{o3d})$ where t_{o3d} refers to the zero offset time to the apex of the hyperboloid. In the double 2-D case, each inline is migrated with $V_{rms}(t_{oj})$ where t_{oj} refers to the apex of each 2-D hyperbola.

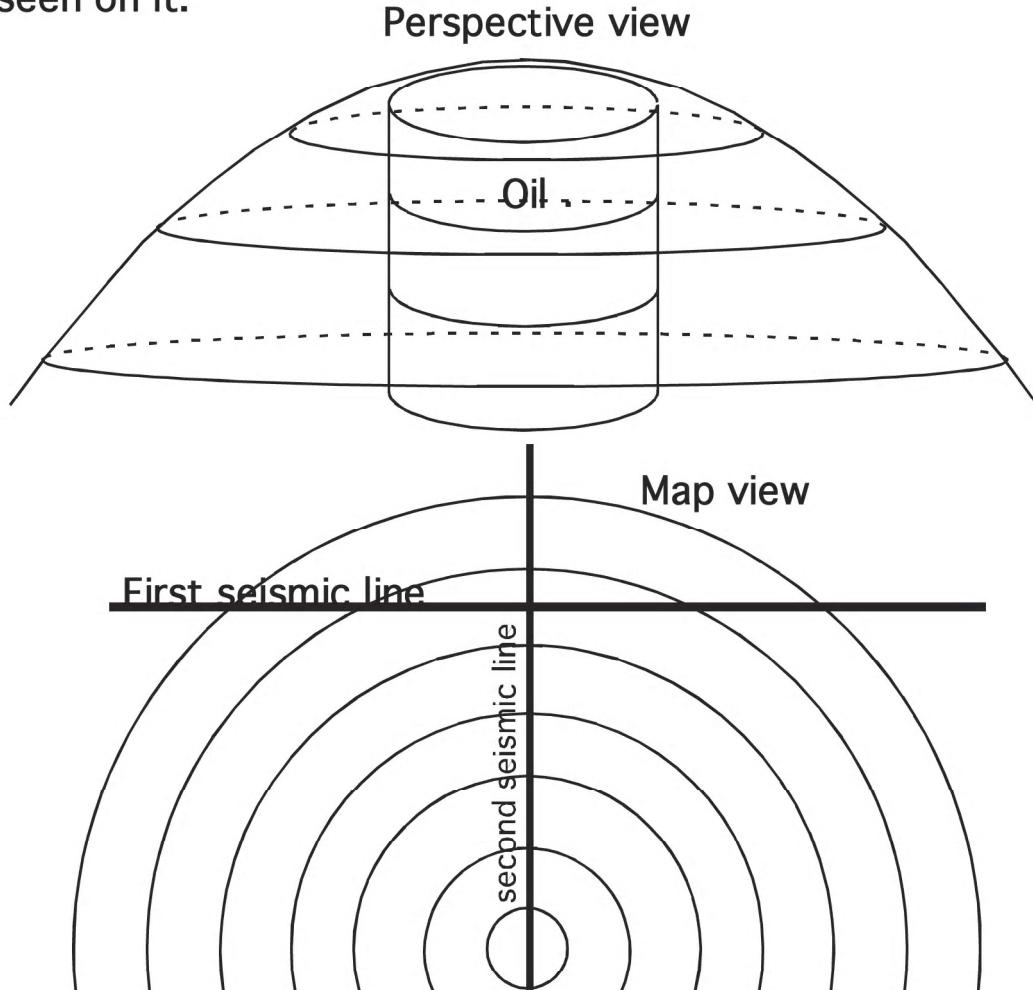
Exploitable Symmetries

Sometimes data is collected over geologic structures which show a high degree of symmetry. A traveltime surface showing invariance in the y or strike direction is shown below. While a general traveltime surface is described by $t(x,y)$, this one is fully modeled by $t(x)$. Along a mountain front, for example, we might expect rough approximations to this symmetry. Logically, this structure is 2-D and seismic line shot in the dip direction will image correctly with a 2-D migration. This is true even if the velocity structure is sufficiently complex to require depth migration provided that there is no strike variation. It follows that a seismic line shot at an angle α to the dip direction can be correctly migrated by simply "projecting" the line into the dip direction by $x=s \cos(\alpha)$ and using a dip cross section as a velocity model.



Exploitable Symmetries

Some structures show approximate radial symmetry such as pinnacle reefs or salt domes. If such a structure is seen on a 2-D line, then its correct position in 3-D can be determined from a second line, perpendicular to the first and passing through the traveltime high seen on it.



Questions:

- Should strike lines be migrated?
- Suppose an anomaly is found on a projected line, is there any confusion about where to drill?
- Suppose an anticlinal anomaly is found on a projected line in rugged terrain. Should a strike line be shot before drilling? How can it be used to refine the target location?

Mapping Strategies

The number of different strategies used to convert 2-D data into 3-D horizon maps is probably proportional to the total number of interpreters times the number of prospects. However, most strategies fall into one of two categories:

- i) Migrate the 2-D data and map the migrated sections.
- ii) Map the cmp stacks and migrate the map.

	Migrating then mapping	Mapping then migrating
P r o	<ul style="list-style-type: none">• Easier to see migrated structures.• Wavefield migrations are stable and accurate.• Strike symmetry can be exploited.	<ul style="list-style-type: none">• Line ties are better.• A 3-D migration is possible.
C o n	<ul style="list-style-type: none">• Line ties are usually worsened by 2-D migration.• Out-of-plane effects cannot be handled.	<ul style="list-style-type: none">• Map migration is a raytrace migration and can be unstable.• All velocity interfaces must be mapped.• Multivalued time surfaces are difficult to handle.• Map migration algorithms are often simplistic.

If migrating-then-mapping is preferred, care should be taken to shoot most lines along dip or project them into the dip direction before migration. A few unmigrated strike lines can aid in tying the map together. With careful planning, even wave equation depth migration can be done if reasonable strike symmetry exists.

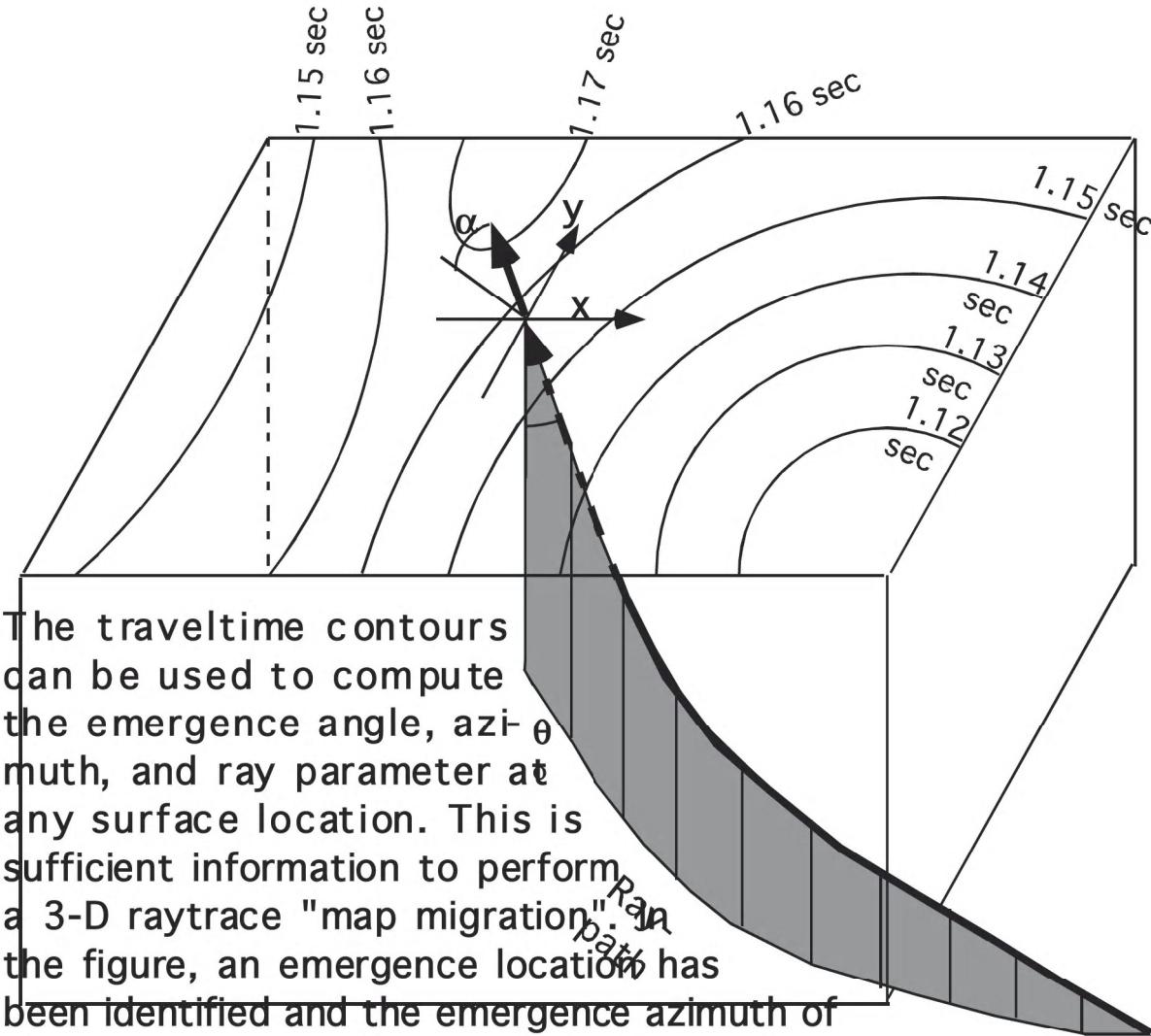
Mapping Strategies

If no symmetry assumptions can be made, then generally map migration is the only alternative.

This decomposition of a travelt ime contour map into a suite of 3-D ray parameter information is the basis for map migration. For a depth migration of these rays, a 3-D velocity model must be constructed which means that, not only must the target horizon be mapped, but every significant velocity interface between target and surface must also be mapped. Then, proceeding iteratively from the surface, each map is migrated using a velocity model consisting of the migrated maps above it. This is a tedious and complex procedure and is often unstable. Instabilities can result from: incorrect mapping, erroneous interval velocities, insufficient smoothing of maps, and insufficient maps. A map can be insufficient to define a structural horizon if that horizon has a double valued travelt ime surface. (A buried syncline is an example.) Such structural problems can be dealt with by making multiple maps and merging them after migration.

A simple, reasonable alternative to the process above is to perform a time migration with a single $V_{rms}(t)$ function. A time migration of a 3-D map has some interesting properties. First, since any ray-path in a $v(z)$ medium is confined to a single vertical plane, and since the ray slowness vector is given by the gradient of the travelt ime map, the plane containing the ray must be orthogonal to the travelt ime contours. Since migration is basically a process of tracking each ray into the subsurface, then the vertical ray planes must also be orthogonal to the migrated map. In general, contours move updip along trajectories orthogonal to them and steepen (contour interval becomes smaller).

Time migration of traveltime maps



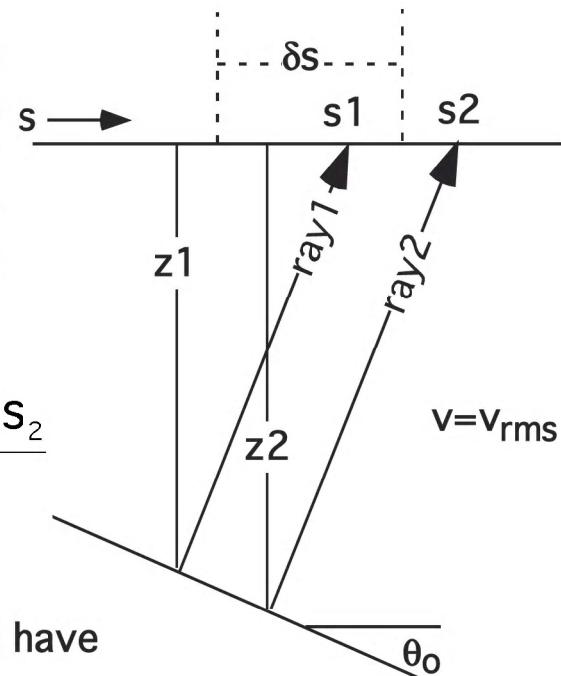
$$\tan(\alpha) = \frac{\Delta t_y / \Delta y}{\Delta t_x / \Delta x} \quad \sin(\theta_0) = \frac{v_0 |\nabla t|}{2}$$

$$|\nabla t| = \sqrt{\left(\frac{\Delta t_x}{\Delta x}\right)^2 + \left(\frac{\Delta t_y}{\Delta y}\right)^2}$$

Time migration of traveltime maps

In this diagram, s is a surface coordinate orthogonal to the traveltime contours. Two rays emerge at s_1 and s_2 at contour times t_1 and t_2 . Thus a ray "pick" is made at:

$$t_{\text{pick}} = \frac{t_1 + t_2}{2} \quad s_{\text{pick}} = \frac{s_1 + s_2}{2}$$



Now, from the ray geometry, we have

$$\sin(\theta_0) = \frac{1}{2} \frac{\Delta t}{\Delta s} v_{\text{rms}}$$

$$\text{migrated time} = \tau = t_{\text{pick}} \cos(\theta_0)$$

$$\text{updip movement} = \delta s = t_{\text{pick}} \sin(\theta_0)$$

Then, the contour interval on the migrated map at coordinate $s_{\text{pick}} + \delta s$ is related to that on the unmigrated map at s_{pick} by

$$\Delta\tau(s_{\text{pick}} + \delta s) = \Delta t(s_{\text{pick}}) \cos(\theta_0)$$

This page was unintentionally left blank

Methods of Seismic Data Processing

**Lecture Notes
Geophysics 557**

**Chapter 10
Seismic Resolution Limits**

Resolution Concepts

The resolution of small earth features with seismic data is a topic of major importance, both from a scientific and an economic perspective. Several rules of thumb seem obvious enough such as:

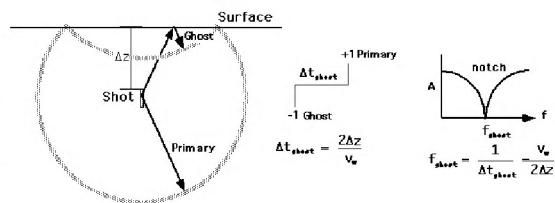
- The more source power the better the resolution
- The more receivers the better the resolution
- The higher the frequency the better the resolution
- Digital sample rates and acquisition geometry play a role
- The earth's velocity structure is also a factor

On some of these points we can be more specific while on others we can do little better than intuitive arguments. The issue of source power is an example. While it is obvious that more dynamite or more vibrators means more source power, it is rarely the case that resolution increases in proportion to the source effort. For one thing, a great deal of source energy is usually trapped in the near surface and is never available to illuminate the subsurface. The effectiveness of different source types and configurations is also strongly arealy dependent. A successful effort will often fail to meet objectives when repeated only a few kilometers away. So, although we can say a few things, there is no substitute for local knowledge gained from experience.

Resolution Concepts

Considering dynamite sources, several points should be noted:

- Shots drilled deep enough to be placed in competent layering give superior results.
- The large impedance contrast at the base of the weathering traps much energy in the weathering layer, so it is better to drill shot holes below the weathering.
- The surface ghost (or base of weathering ghost) can cause a spectral notch which often falls in the desired signal band.

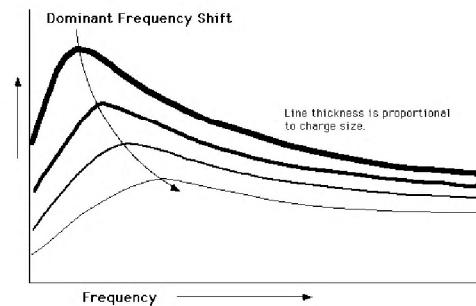


Sources which are too deep or occur in low velocity pockets have this problem

- Often an array of smaller charges in shallow holes will produce better results than the same charge weight in a single hole. This is attributed to array effects, possible ghost notches, and the relationship between charge size and dominant frequency.

With decreasing charge size¹:

- Dominant frequency increases
- Power at any frequency decreases
- Total power decreases roughly with the square root of charge size

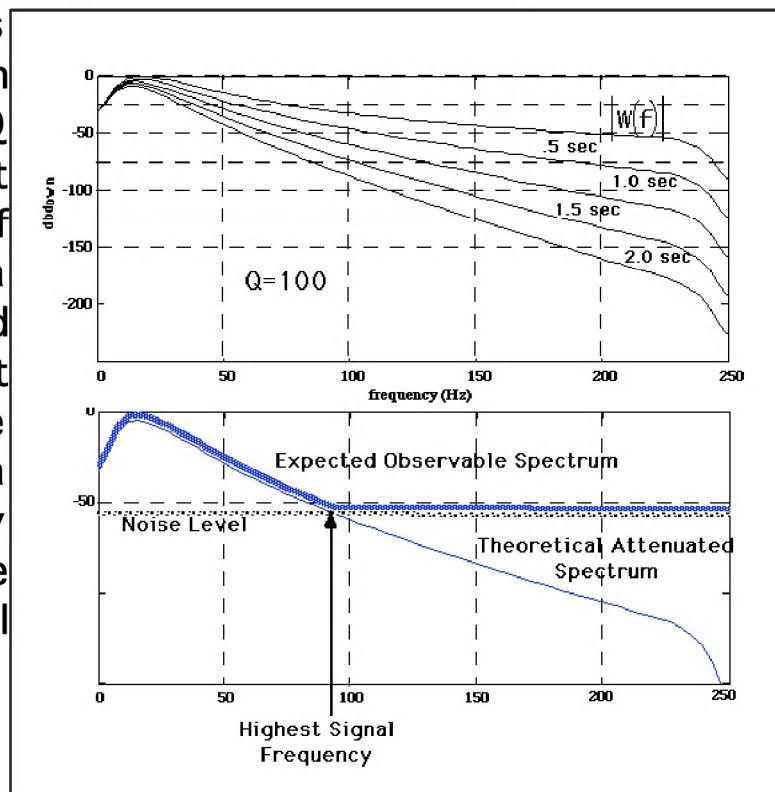


¹Sources:

Goodway (1991 CSEG Abstract)
Chevron Canada folklore

Resolution Concepts

- Whatever frequencies the source generates, the highest signal frequency received at the geophone is guaranteed to be less due to absorption (Q) effects. Simple Q models predict exponential decay of frequency after a given traveltime, and with a constant background noise level, this implies a "corner" frequency which predicts the maximum signal frequency.

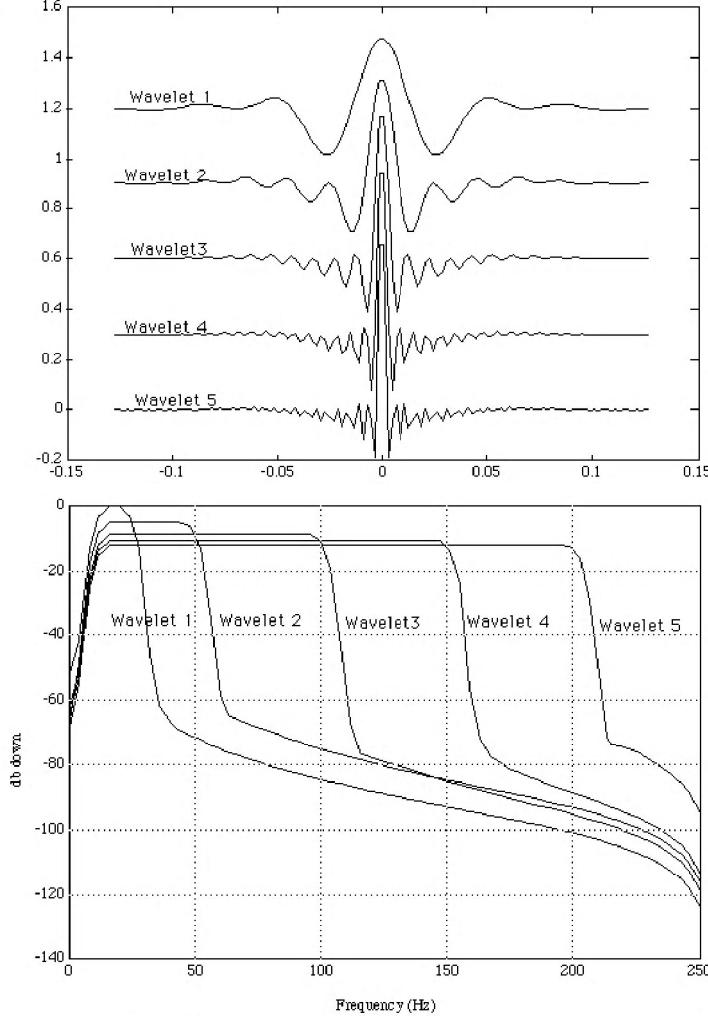


Some observations with Vibroseis sources:

- The signal band can be no broader than the swept band (duh...)
- Source power at any frequency is proportional to the sweep time at that frequency.
- Arrays are typically large because of the truck size.
- Penetration below the weathering is always a problem.
- A stack of several short sweeps usually beats one long one.

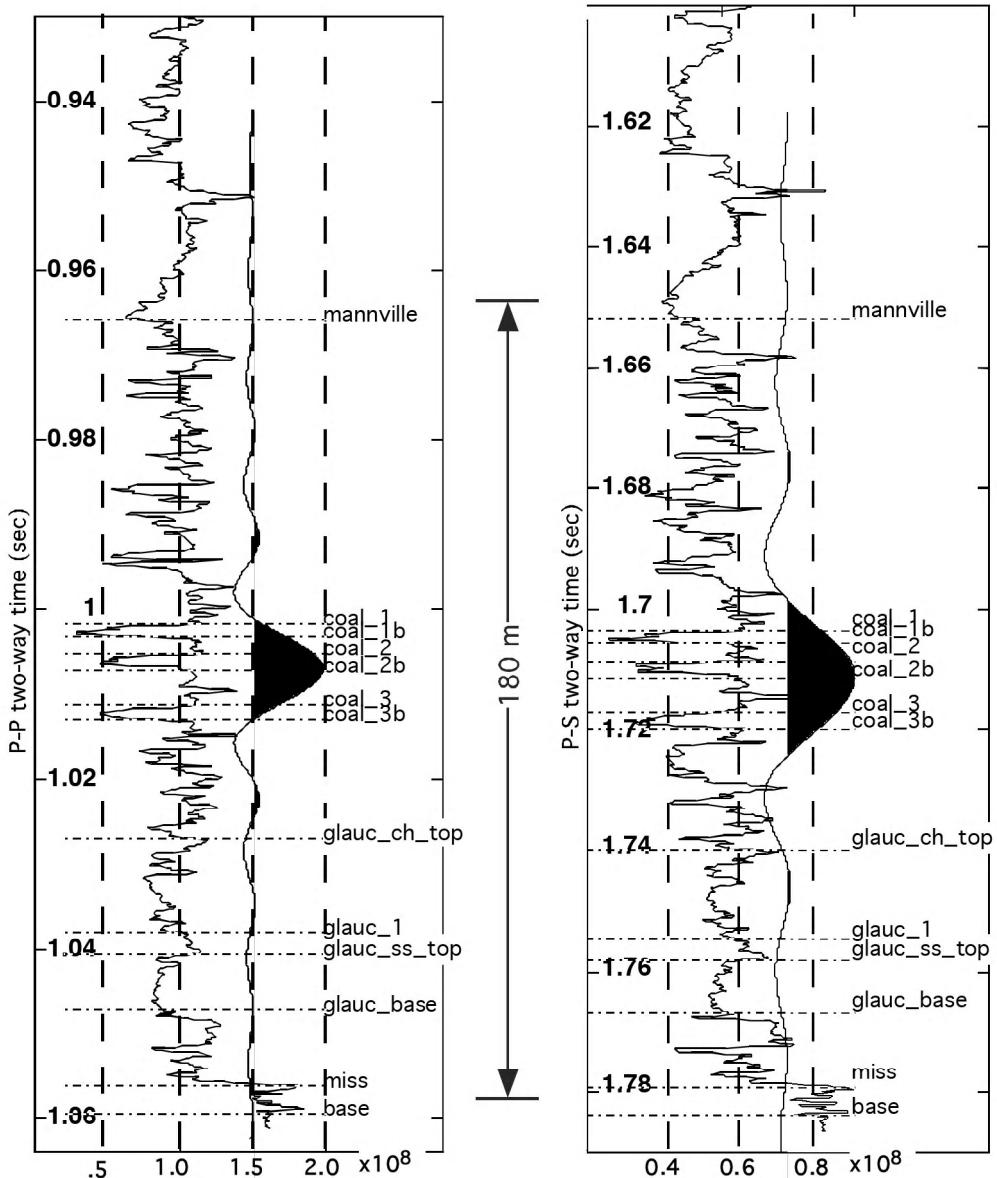
Resolution Concepts

Temporal (or vertical) resolution is controlled directly by the signal bandwidth. Greater bandwidth is greater resolution:



However, vertical resolution in depth is directly dependent upon the formation velocity. Since $\lambda = v/f$, lower velocities have lower wavelengths and hence higher resolution results. This turns out to be a major reason to explore with shear waves since V_s is often around half of V_p . We see an example on the next page.

Resolution Concepts

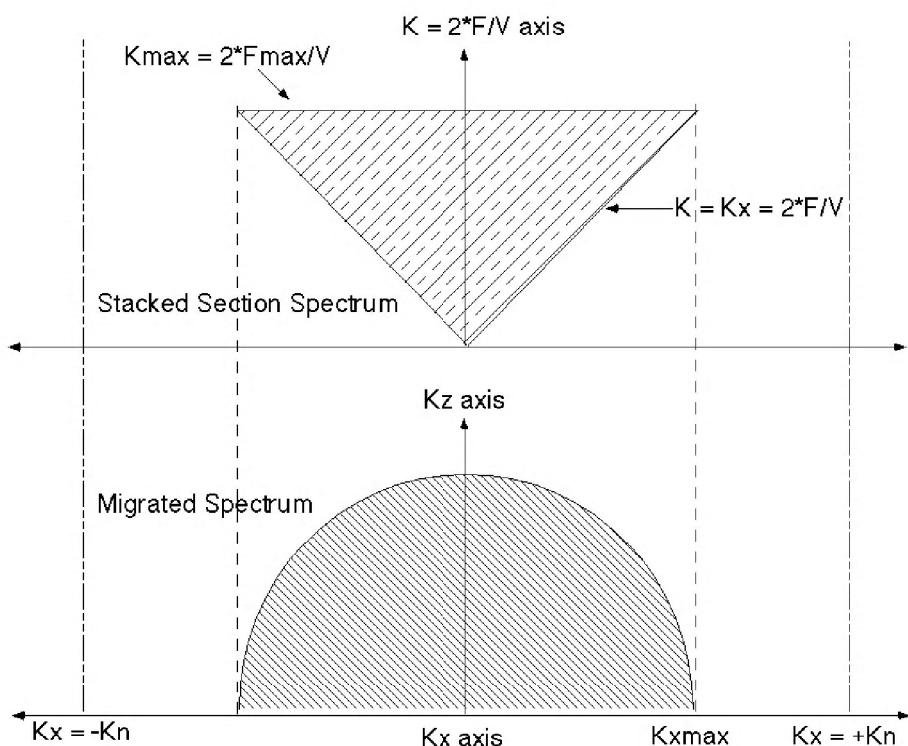


On the left is P-wave impedance from the Blackfoot 08-08 well plotted in P-P time together with a 2-5-75-85 Ormsby wavelet. On the right is S-wave impedance from the same well plotted in P-S time with a 2-5-33-38 Ormsby wavelet. The impedance calculations used P and S velocity logs from a dipole sonic and a density log.

Resolution Concepts

The principles of lateral resolution are less commonly known but are still well developed from migration theory. The fundamental observation that resolution depends on bandwidth still holds except that now we are concerned with the k_x bandwidth.

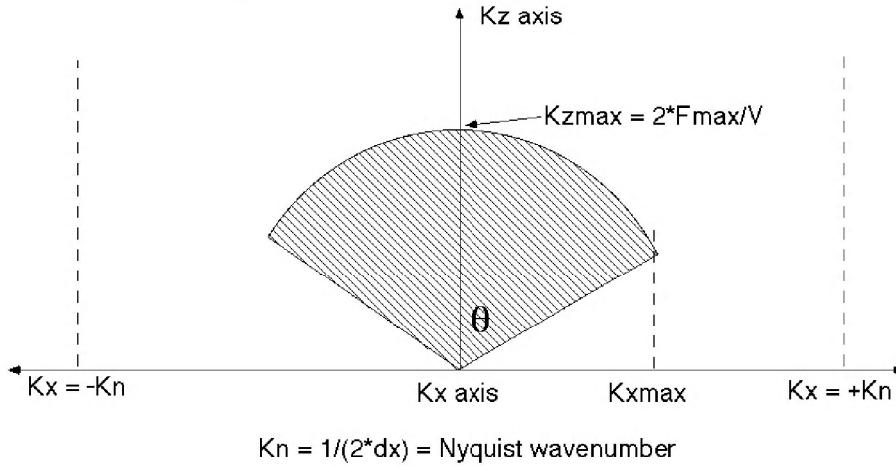
Migration maps the triangular "body wave" spectrum into the semi-circular region at constant k_x . The mapping is determined by: $K^2 = K_x^2 + K_z^2$



This result is from F-K migration theory and is a direct consequence of basic physics derived from the scalar wave equation. The mapping occurs at constant k_x and maps a line of constant frequency in the triangle to a circle in the migrated spectrum. Thus the maximum k_x value is directly predictable from the maximum frequency and the velocity.

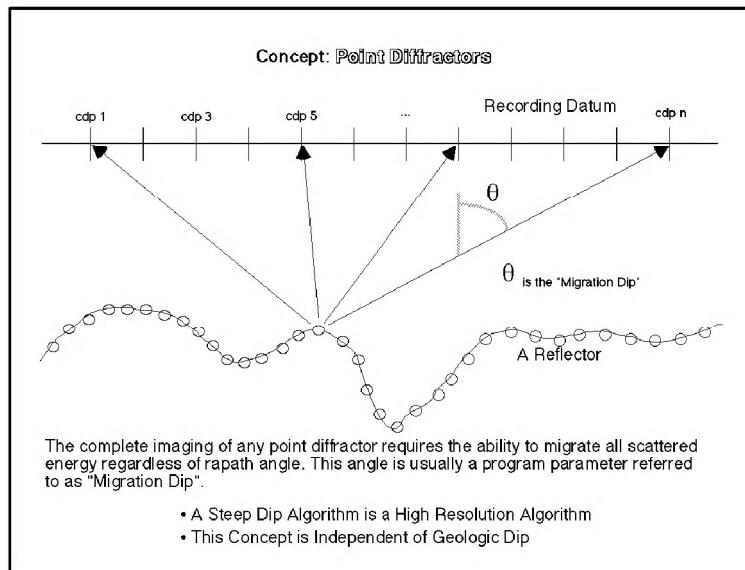
Resolution Concepts

Constant Velocity post stack migration gives a simple direct relation between the factors controlling resolution.



The resolution of small horizontal features improves with increasing $K_{x\text{max}}$.

The angle θ in the $K_{x\text{max}}$ bandwidth equation is commonly called dip but is more properly thought of a scattering angle.

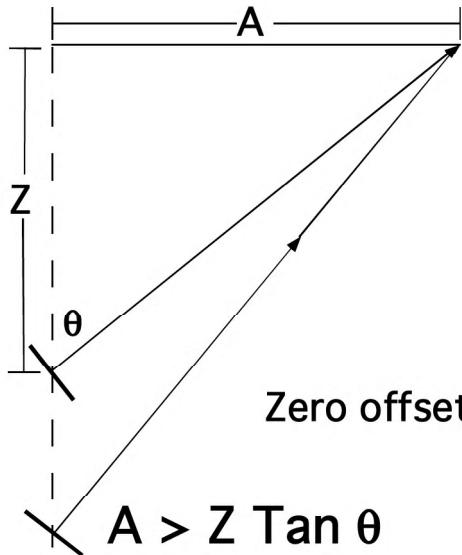


Resolution Concepts

We have derived a fundamental result which predicts k_x bandwidth and hence horizontal resolution. We repeat it here in larger symbols so you'll be sure to notice:

$$k_{x_{\max}} = \frac{2 F_{\max} \sin(\theta)}{V}$$

Here we must interpret θ as the largest scattering angle captured by the seismic survey at the location of interest. A little thought suffices to realize that horizontal resolution is therefore variable throughout a seismic survey. There are three major limits on θ : aperture, record length, and spatial aliasing.



Zero offset scattering angle is limited by:

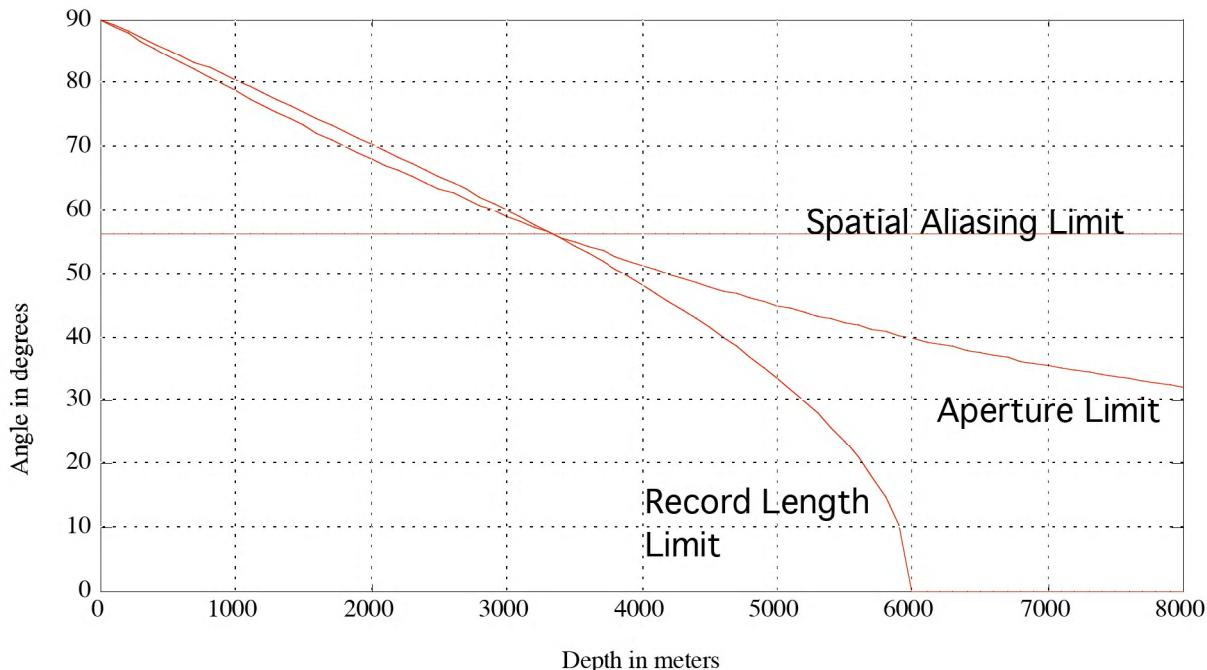
- | | |
|------------------------------------|--------------------------|
| $A > Z \tan \theta$ | : Aperture Limit |
| $T > 2Z / (V \cos \theta)$ | : Record Length Limit |
| $\Delta X < V / (4 F \sin \theta)$ | : Spatial Aliasing Limit |

Resolution Concepts

Example: Zero Offset Constant Velocity Scattering Angle Limits
for the case of:

Velocity = $v = 4000 \text{ m/sec}$
Record Length = 3 sec
Frequency = $f = 60 \text{ Hz}$

Aperture = $A = 5000 \text{ m}$
CDP interval = $dx = 20 \text{ m}$



To achieve a resolution angle of θ at depth z , a seismic survey over a constant velocity earth must satisfy :

$$A \geq z \cdot \tan(\theta) : \text{Aperture Limit}$$

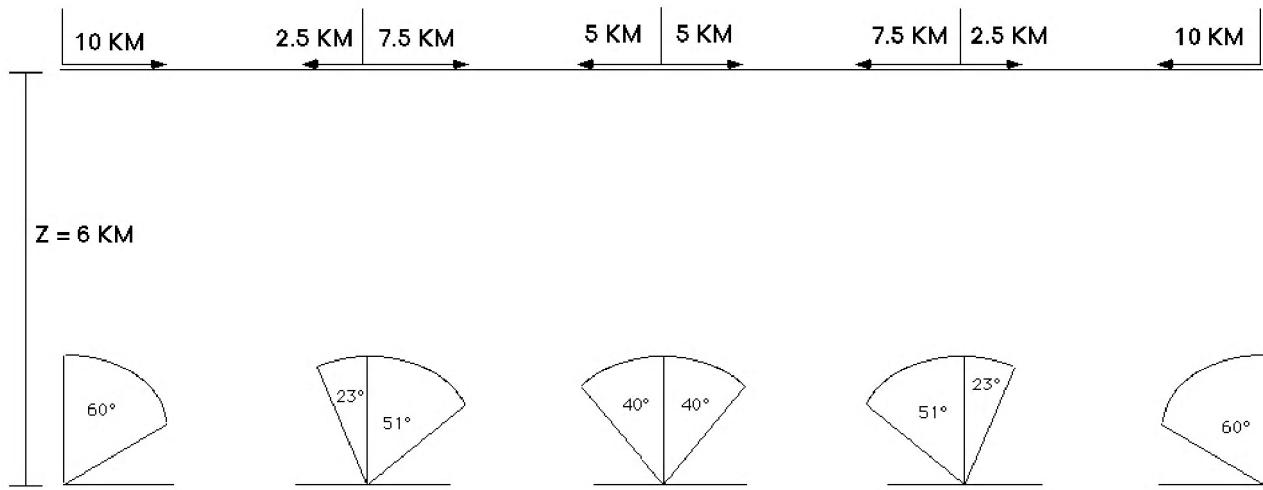
$$T \geq 2 \cdot z / (v \cdot \cos(\theta)) : \text{Record Length Limit}$$

$$dx \leq v / (4 \cdot f \cdot \sin(\theta)) : \text{Spatial Aliasing Limit}$$

Resolution Concepts

The effect of aperture can be especially dramatic as it varies along a seismic line or across a 3-D survey.

Finite Line Length and the Lateral Variation of Resolution



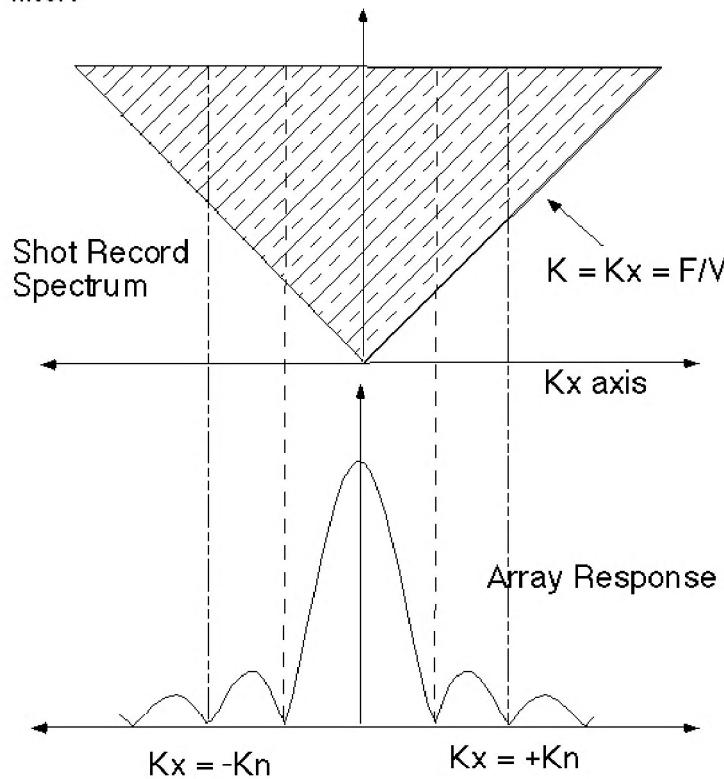
Effective Aperture varies laterally along a seismic line which means that the captured zero offset scattering angles also vary along the line.

For any analysis site, the scattering angle to use in the bandwidth equation is the most restrictive limit imposed by any of these three effects.

Resolution Concepts

We should also mention the effect of arrays (and any digital seismic process which directly limits wavenumber) is to reduce k_x and therefore to reduce resolution.

A recording array (source or receiver) is equivalent to multiplication of the recorded spectrum by a K_x filter.



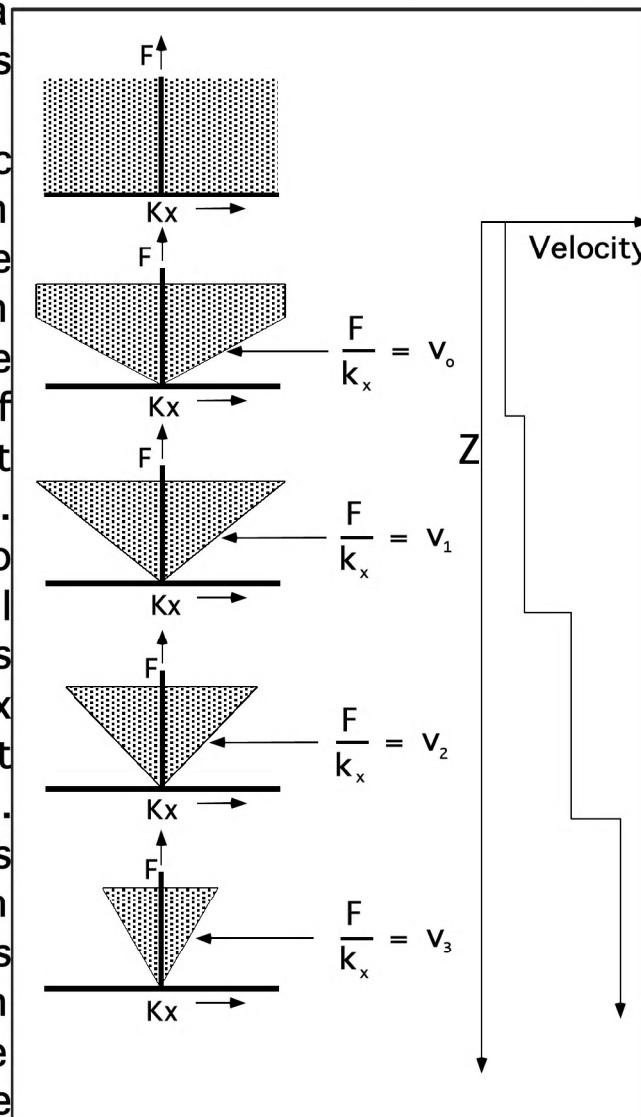
We oversimplify here since the connection between the prestack wavenumbers and the final k_x of the migrated section is a bit complicated. However, the basic concept that arrays limit lateral resolution is correct.

Resolution Concepts

It is a fundamental result that lateral resolution is controlled by the maximum vertical frequency, the maximum scattering angle and the velocity. The spatial sample rate is only a factor if the data is aliased.

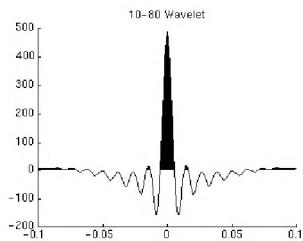
As band limited seismic waves propagate down into a $V(z)$ medium, the wavenumber bandwidth continually shrinks due to the refraction of evanescent energy at each velocity interface. Thus, the ability to resolve small horizontal features (which is directly dependent on K_x bandwidth) must decrease with depth. This effect is independent of (and in addition to) the effects of inelastic attenuation and finite survey size which also cause resolution to decrease with depth.

When seismic data is migrated, the migration algorithm must apply an evanescent filter to bandlimit the data in accordance with this principle.

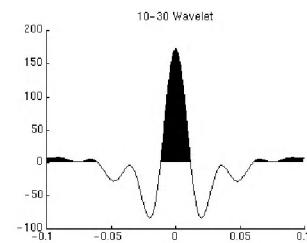


Resolution Concepts

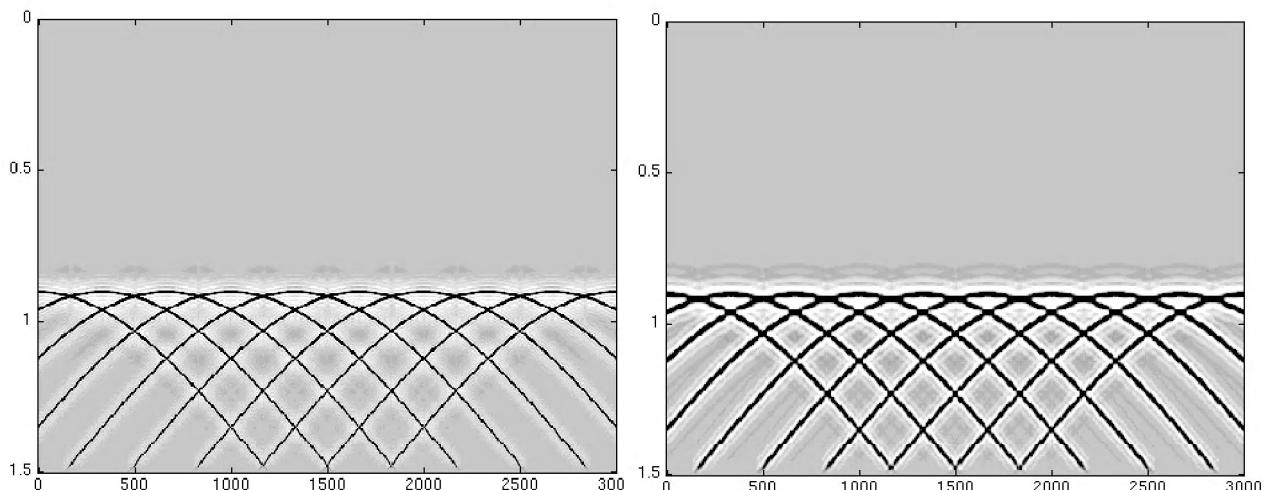
An Example:



A 10-80 zero
phase wavelet



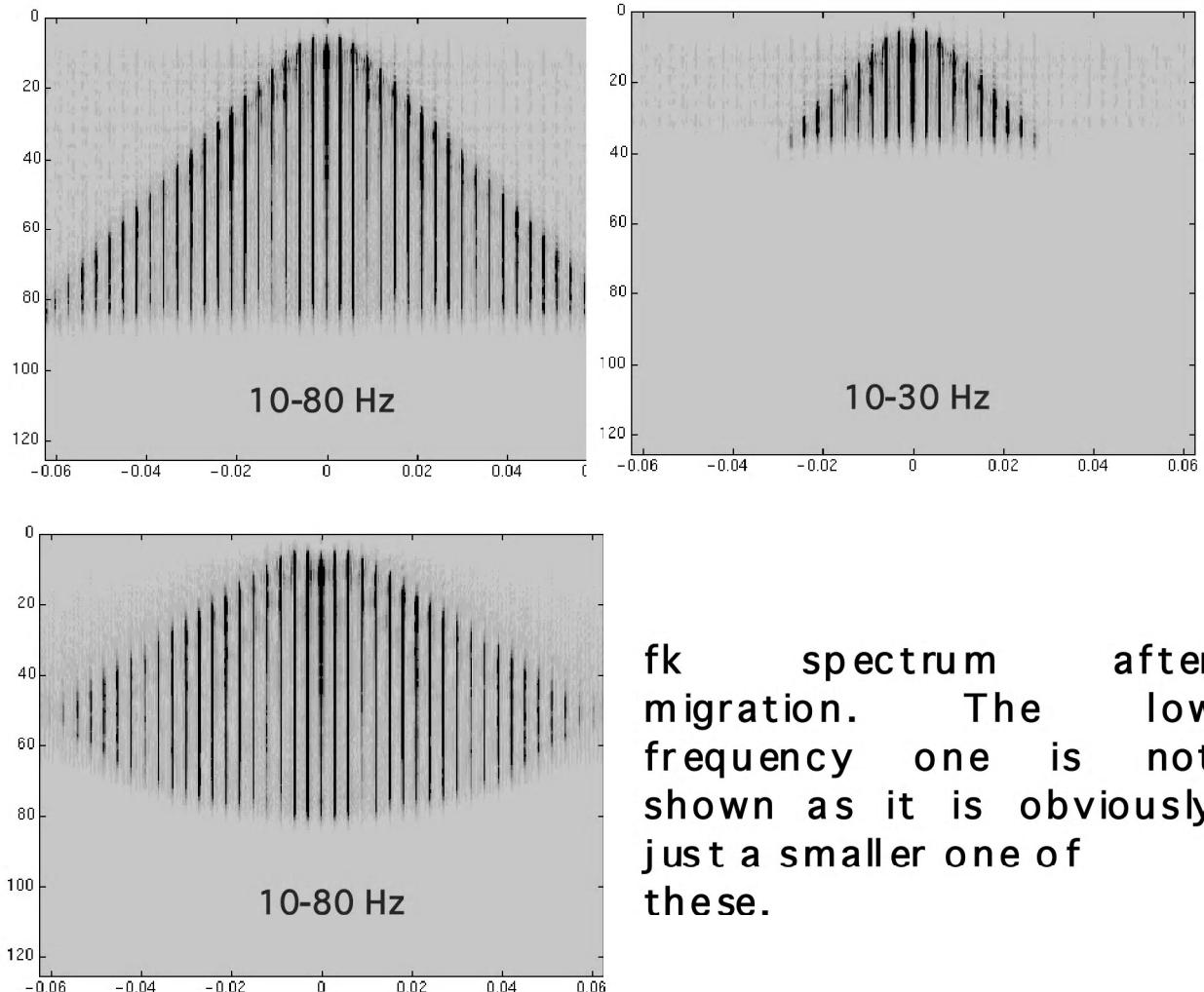
A 10-30 zero
phase wavelet



Here are two sections of point scatterers at a constant z and illuminated with a high frequency and a low frequency source. Parameters are $\Delta x=8\text{m}$, $\Delta t=.004\text{ sec}$, and $v=2000\text{ m/sec}$

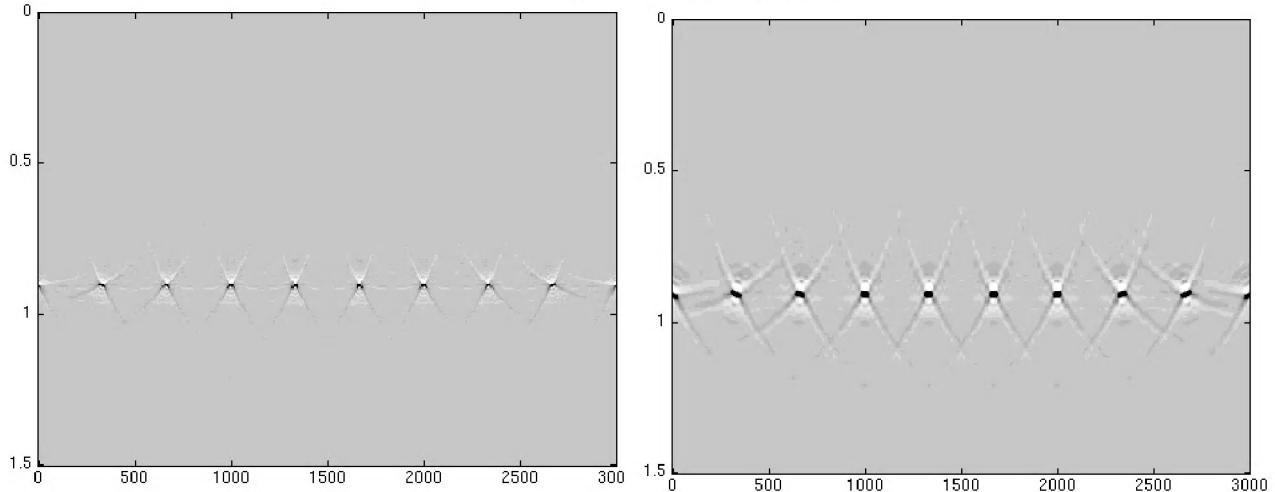
Resolution Concepts

fk spectra before migration

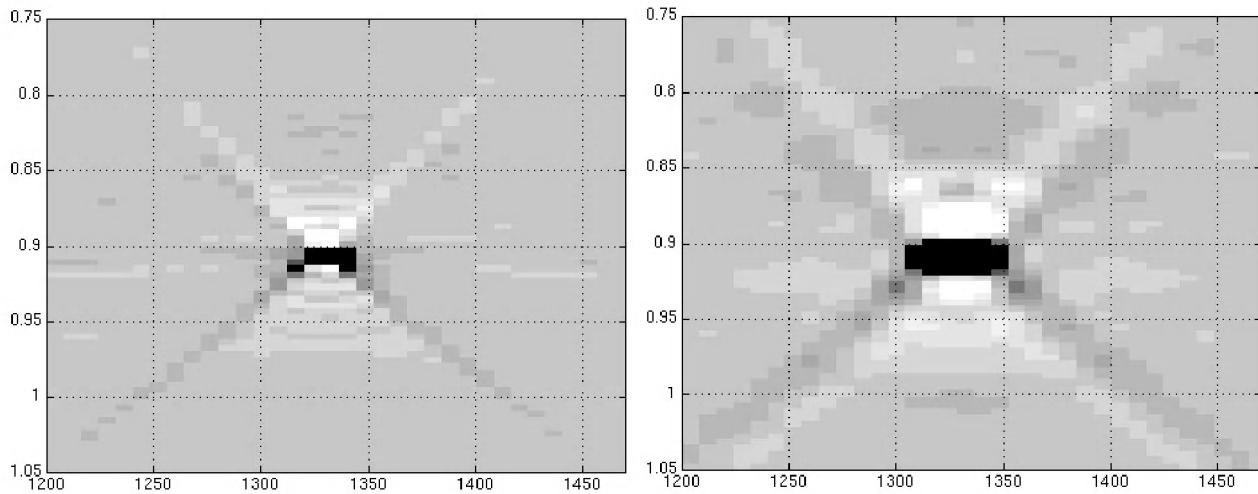


Resolution Concepts

The migrated sections



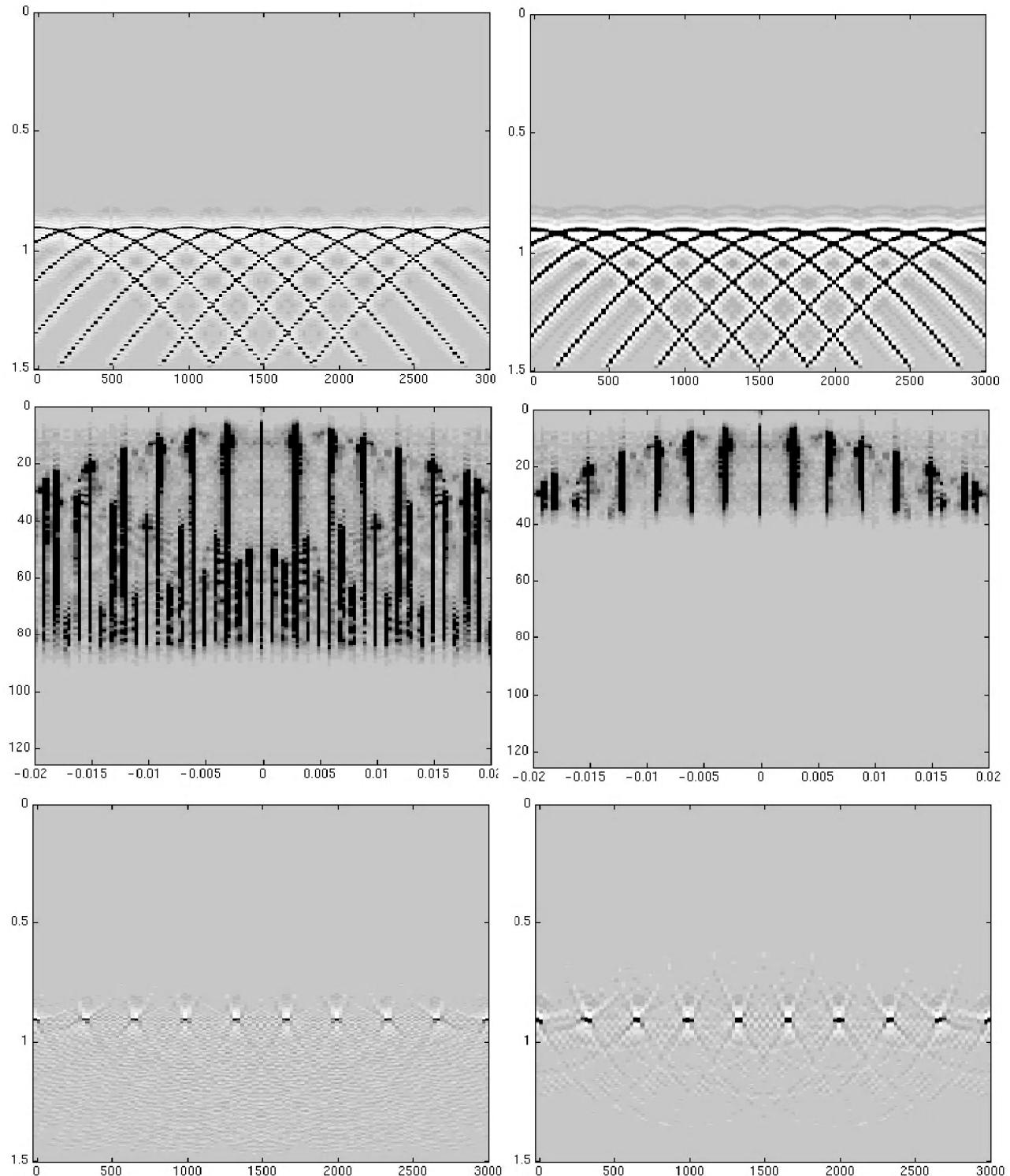
Zooms of the above



In the upper panels, note the variation in shape of the images across the section. This is a direct consequence of the aperture limit discussed previously.

In the lower panels, we see that both the vertical and horizontal resolution are dramatically improved with the increased bandwidth. The ratio of the horizontal dimensions of the two images is roughly in inverse proportion to the ratios of the maximum frequencies.

Resolution Concepts



Similar to the previous example except that spatial sampling is coarse ($\Delta x = 25m$) and aliasing occurs.

Geophysics 557 Final Exam Study Guide

What are the expressions for P and S wave velocities in terms of the Lame constants?

How does the Vp/Vs ratio depend on poison's ratio?

How is the normal incidence reflection coefficient (for P waves) related to impedance?

What is impedance?

How are normal incidence reflection and transmission coefficients related?

What is meant by the term "impulse response"?

What physical effects are modeled in construction of a normal incidence seismogram a discussed in lecture?

Under what conditions can the earth response to a real source be modeled as a convolution of a source waveform with an impulse response?

What is the major use for 1-D synthetic seismograms?

Why are multiples and transmission losses not typically included in such models?

What information is typically input to the 1-D synthetic seismogram computation?

Describe convolution by replacement.

Describe convolution as a weighted sum.

Use the integral form of convolution to prove that convolution is linear. That is, the convolution of c with a+b is equal to the convolution of c with a plus the convolution of c with b.

What happens when any function is convolved with a complex sinusoid?

What is technical meaning of a phrase " a 90 degree wavelet" ? 45 degree? any degree?

What is the definition of the 1-D Fourier transform? The inverse transform?

How is the Fourier phase spectrum defined? The amplitude spectrum?

How are the "time width" and "frequency width" of a function related?

What is a Dirac Delta function? What is its Fourier transform?

When g and f are convolved in time, what happens to their Fourier spectra? What happens to their amplitude spectra? Their phase spectra?

How is the Nyquist frequency related to the temporal sample rate?

What is the frequency sample rate of a time series of length T?

To avoid aliasing a signal of frequency Fmax, what sample rate must be used?

From a practical viewpoint, the signal frequencies in your data should be less than a constant, c, times Fnyquist. What is a good value for c?

What sample rate should you use for 80 Hz signal? For 130 Hz? For 40 Hz?

What is an anti aliasing filter? When should it be used?

What are the two main purposes of a zero pad when doing a discrete Fourier transform (DFT)?

What is circular convolution?

What is the relationship between the DFT and the fast Fourier transform (FFT)?

How is the Z transform related to the DFT?

A filter which has a pole close to the unit circle at some frequency does what to that frequency? How about a zero close to the unit circle?

Where are the zeros of a minimum phase filter located in the z plane?

What is the definition of minimum phase in terms of causality and stability?

How is an inverse filter defined in the z plane?

What is meant by a stable inverse?

What is meant by the statement that "This seismic data is minimum phase"? (Note that the statement is technically always false but it has a practical, definite meaning.)

In order to assert that seismic data is "minimum phase" at some stage of the processing what conditions must be met?

How is cross correlation defined? What does it mean? What does the cross correlation lag mean?

What is an autocorrelation? What is the expected autocorrelation of a random sequence?

What are the two central problems of spectral estimation?

What is the role of the "window" in spectral estimation?

What kind of spectrum is well modeled by the Burg spectrum?

What is the 2-D Fourier transform of a linear event with apparent velocity v ? Draw a sketch showing (x,t) space and $(f-k)$ space for a range of different apparent velocities.

What is the most likely apparent velocity and where is it found in (f,k) space?

What is spatial aliasing? For a given apparent velocity and spatial sample rate, what is the critical frequency at which spatial aliasing begins?

How can convolution be expressed as a matrix operation? Draw a diagram showing the Toeplitz matrix symmetry.

Describe the six basic modes of seismic attenuation.

Geometric spreading corresponds to what conservation law?

Under geometric spreading, amplitude decreases proportional to what?

How is Q defined?

What is the formula for amplitude loss in a constant Q theory?

What is the formula for transmission losses in a layered medium?

What phenomenon is responsible for trapping large amounts of seismic energy in the near surface?

What is true amplitude processing?

What do constant Q models predict about the signal bandwidth of seismic data?

The phase effects associated with Q attenuation are known as what?

What assumptions are required to derive these phase effects?

What is unique about a minimum phase wavelet?

What is the most important property of minimum phase wavelets from the viewpoint of deconvolution theory?

What is meant by velocity dispersion?

What is the convolutional model? Write a mathematical expression for the model as it is applied in deconvolution theory. Define each term and state the assumptions which constrain each.

Can all types of multiples be included in the convolutional model? Why or why not?

The convolutional model expects the seismic trace to be stationary, What is meant by this? Is it a reasonable expectation?

What are the essential steps of frequency domain spiking deconvolution?

How is the seismic wavelet estimated in frequency domain spiking deconvolution?

If a wavelet is known to be minimum phase, then its inverse can be found by solving a set of matrix equations whose left hand side involves not the wavelet itself but a statistical measure of it. What is this measure? Explain intuitively how this result is equivalent to computing a wavelet's phase spectrum from its amplitude spectrum under the minimum phase assumption.

Wiener spiking deconvolution assumes that the autocorrelation of the seismogram is similar to the autocorrelation of the wavelet. Justify this assumption by argument from the convolutional model.

What is meant by the 'stab' factor or 'white noise' factor in deconvolution?

It is customary in seismic data processing to follow deconvolution by a bandpass filter. Is this a sensible practice? Either justify it or refute it by argument from deconvolution theory.

What is the relationship between prediction error filters and spiking deconvolution?

Explain why deconvolution keeps the prediction error and rejects the predictable part of a seismic trace.

How is gapped predictive deconvolution implemented? What is a typical example of a type of multiple which it is designed to attenuate? How can the prediction gap be chosen?

How are midpoint and offset defined in terms of source and receiver coordinate?

Illustrate with a diagram.

The near surface is generally assumed to cause effects which are a function of what coordinates?

The subsurface effects are often assumed to be a function of what coordinates?

How is a static delay defined? What physical effect is often used to justify the assumption of static delays in the near surface?

What is the definition of source static? Receiver static?

What is meant by the term "datum"? How does its choice effect the statics application?

What datum is most appropriate for pre stack processing?

What are surface consistent methods? Why are they useful? List some examples of common surface consistent applications?

What is the definition of vertical travelttime?

How can instantaneous velocity be computed as a function of vertical travelttime?

What is a time-depth curve and what is it used for?

How is average velocity defined? In what sense is it an average?

What is the mean velocity and how is it defined?

Define Vrms in terms of instantaneous velocity and as is relates to the mean and average velocities.

Which is always greater the average or the rms velocity?

What is interval velocity? Define at least two type of interval velocities.

What is the expression for the addition of two interval velocities?

What is the expression for the "Dix" interval velocity calculated from two closely positioned rms velocity measurements?

Under what conditions can an interval velocity be said to approximate a local wave propagation velocity?

How can imaginary interval velocities result from a Dix interval velocity calculation?

Derive the travelttime equation for normal moveout. Use a diagram to show the meaning of all quantities. What is the shape of the travelttime curve?

What is the shape of the wavefront in the nmo experiment as it approaches the receivers?

How must the nmo equation be modified if the reflector is dipping?

What is the stacking velocity for a dipping reflector beneath a constant velocity overburden?

How must the nmo equation for a dipping reflector be modified to take the azimuth of the seismic line into account?

What is the definition of stacking velocity? Explain why stacking velocity is always a function of offset.

Using a diagram, derive the geometric relation between wavelength components and wavelength for a periodic planar wavefront.

What are wavenumbers?

What is apparent velocity? What are the mathematical limits (upper and lower) of apparent velocity?

How does apparent velocity relate to wavelength components and wavenumbers?

What is Snell's law? Snell's law can be considered as the conservation of what quantity?

What is a $v(z)$ medium?

When raytracing in a $v(z)$ medium, what quantity is conserved and how is it defined?

Derive the distance traveled and traveltime integrals for raytracing in a $v(z)$ medium.

How can the ray parameter be measured?

When velocity increases linearly with depth, what shape are the raypaths? The wavefronts? (Exact equations not necessary.)

For the nmo experiment in a $v(z)$ medium, explain how the result that stacking velocities may be approximated by rms velocities arises. What assumptions are required? In practice, when can we expect it to be roughly valid?

The Dix equation moveout can be interpreted as allowing the replacement of the real $v(z)$ medium by a constant one with properly chosen parameters. Explain this.

Explain why interpolation of trace sample values is needed in nmo removal. What is moveout stretch? Why does it arise?

What are residual statics? How are they computed? What is their purpose? What processes should be run on seismic data prior to attempting a residual statics solution?

What is velocity analysis? How is it performed? What processes should be run on seismic data prior to attempting a velocity analysis?

Do statics and moveout removal commute? That is, do you get the same result regardless of the order of the processes? If not what is the preferred order?

In the extension of nmo and dip to $v(z)$, what quantity must be measured in addition to stacking velocity in order to allow the computation of apparent dip and the "dip correction" of stacking velocities?

What can be said about the stacking velocities of multiples? Where will they be found on a stacking velocity analysis chart?

After stacking, the power of random noise can be expected to be reduced by what factor? Considered as an "f-k" process, stacking can be said to pass what portion of the offset wavenumber spectrum?

Are "f-k" filters applied to cmp gathers likely to improve a stack? What if they are applied to shot or receiver gathers?

What is a zero offset section? How does it serve as a model for a stack?

What is the relation between traveltime gradient measured on a stack and the normal incidence ray parameter?

What information is needed for the raytrace migration of a normal incidence seismogram?

What are the algorithmic steps in normal incidence raytrace migration?

Time migration processes are biased towards what class of rays? How are these rays handled? When is time migration a valid process? When is depth migration a valid process?

What is the migrator's formula? How can it be used?

Explain post stack migration by replacement of each point with a wavefront. How are the wavefronts defined?

What is meant by "migration dip"?

Should data from an area where all geologic dips are less than 5 degrees be migrated?

Why? Should a steep dip algorithm or a low dip algorithm be used?

What is Huygen's principle?

What is the traveltime curve of a point diffractor for post stack migration? For pre stack migration?

What is a diffraction chart? Explain how diffraction curves can be used to construct a zos image from a geologic model?

What is the exploding reflector model? Why is it useful?

In the exploding reflector model, what is the mathematical expression for the migrated section? For the zos image?

Using the exploding reflector model to explain wavefield extrapolation. What is the mathematical expression for an extrapolated section? What is the relationship between any extrapolated section and the migrated depth section?

What is the dispersion relation? Use the dispersion relation to derive the mapping which defines f-k migration. Use a diagram to illustrate the mapping of the f-kx spectrum to the kx-kz spectrum. What is the meaning of the evanescent boundary? What determines the maximum kx wavenumber after migration? What determines it before migration? Draw a flow chart for f-k migration.

What is f-k wavefield extrapolation? Derive the expression for the f-k phase shift required to shift the datum by Δz . Explain how recursive f-k phase shifting can be used to create a v(z) migration algorithm. Draw a flow chart.

What is the geometric shape of the wavefield extrapolation operator? What are the two distinct components of the operator? How can it be applied in the (x,t) domain?

What is the major distinction between time and depth migration? Can time migration produce a depth section and depth migration a time section? Explain.

What is Kirchoff migration? What is the shape of the Kirchoff migration operator (constant velocity) when applied post stack? Pre stack?

Describe a general method to determine the shape of the Kirchoff migration operator. Your method must be valid for any (x,y,z) location, any velocity, and pre or post stack. What is pre stack time migration? When is it a valid process? Should it be inferior, the same, or superior to stacking and post stack time migration?

What is DMO? Describe a flow using DMO that should give similar results to pre stack time migration. What are the strengths and weaknesses of the DMO approach? Under what circumstance is DMO->stack->migration exactly the same as pre stack migration?

What kind of velocities should be input to the NMO removal step in a flow involving DMO? How can these velocities be obtained?

Describe, without equations, the essential steps in CSP migration? How does CSP analysis effect velocity resolution?

What is the central (most difficult and most important) problem in the application of depth migration to the thrust belt? Describe at least one approach to solving this problem.

What is wavelet processing? What are the essential steps in wavelet processing? When should it be done in a processing flow? When is it necessary? What are two common methods of wavelet estimation?

What is impedance inversion? When should it be run in a processing sequence? How can the convolutional model (from deconvolution theory) be used to justify impedance inversion? What is the major computation involved in impedance inversion? Describe at least two common problems with impedance inversions that are difficult to solve.

What is the expected behavior of the amplitude spectrum of the radiated waves from a dynamite sources as a function of charge size?

Explain how Q effects necessarily lead to a time variant (i.e. non-stationary) signal bandwidth. What is the relationship between spectral width and wavelet width?

What is a "corner frequency"? When can it be observed? What does it mean?

For a constant velocity earth, what are the equations which express the limits of observable scattering angle due to aperture, record length, and spatial aliasing? Make a sketch of their basic form.

Starting from the theory of f-k migration, derive an expression for the maximum k_x after migration as a function of frequency, velocity, and scattering angle. Explain the relevance of this to the problem of resolving small horizontal features? What steps can be taken in recording or processing to increase horizontal resolution?

Exam sampler. There will be between 30-35 multiple choice questions and 4-8 short answer questions.

PLEASE ANSWER ALL OF THE FOLLOWING QUESTIONS.

There are a total of 100 marks (points) for the examination. You have about 100 minutes for the exam.

Write all work directly on the examination sheets. If you need more room, you may attach a work sheet with your name and the question number on it. PLEASE HAND IN THE EXAMINATION SHEET AND ALL WORK SHEETS WITH YOUR ANSWERS.

Multiple Choice Questions (2 points each)

INSTRUCTIONS: For each question, there are two best (most correct) responses. Choosing both correct responses and no incorrect ones is worth two points. One correct and one incorrect is worth one point, and any other result (including more than two selections) is worth zero. Write your answers in the space provided below each question.

- 1) The 1-D synthetic seismogram, as discussed in lecture:
- a) can be made to contain all possible multiples.
 - b) is useful for modeling AVO and converted waves.
 - c) applies the source waveform of a band limited source by correlation.
 - d) is an excellent model of a trace on a stacked and migrated section, provided that all possible multiples are included in the solution.
 - e) is based on ray theory and normal incidence reflection and transmission coefficients.

answer _____

- 13) Average velocity:
- a) characterizes the shape of diffraction curves on a cmp stack.
 - b) is depth divided by the vertical travelttime to that depth.
 - c) is a mathematical average over depth.
 - d) can be measured on an f-k plot.
 - e) is a mathematical average over travelttime.

answer _____

- 20) Post-stack F-K migration:
- a) easily handles variable velocity.
 - b) is useful to explain the transformation of the data spectrum under migration.
 - c) shows that a constant frequency, f , maps to a hyperbola in (k_x, k_z) .
 - d) is a steep dip algorithm.

e) works by applying a phase shift to the f-k transform.

answer _____

31) Minimum phase:

- a) is the state of all raw seismic data.
- b) is the desired state of final migrated sections.
- c) means that the phase is the Hilbert transform of the log of the amplitude spectrum.
- d) is the smallest possible phase.
- e) is only possible for time series which have inverses.

answer _____

Short answer problems (10 points each)

Please work directly on the examination sheets. Show all work.

1) Deconvolution

Suppose a dynamite dataset has mistakenly had a zero phase bandpass filter applied before deconvolution. Assuming that the original data was noise free, with a minimum phase wavelet and a white reflectivity:

a) Write a time domain convolutional expression for a single trace after the bandpass filter was applied but before deconvolution. Then rewrite this expression in the frequency domain.

b) By working through the steps in frequency domain deconvolution, derive a frequency domain expression for the embedded wavelet remaining after standard minimum phase deconvolution. Show each step in the frequency domain deconvolution process and indicate any smoothing but you may assume that stab factors (white noise) are not needed.

c) Derive an expression for a correction filter which can be applied after deconvolution to give the desired result from the deconvolution process.

