ERASMUS Teaching (2008), Technische Universität Berlin

Digital Signal Processing

Assoc. Prof. Dr. Pelin Gundes Bakir
gundesbakir@yahoo.com

References

- Mitra S.K., ‘Digital Signal Processing’, *A Computer based approach, Mc-Graw Hill, 3rd Edition, 2006.*
- Heylen W., Lammens S. And Sas P., ‘Modal Analysis Theory and Testing’, Katholieke Universiteit Leuven, 1997.
- Keith Worden ‘Signal Processing and Instrumentation’, Lecture Notes, http://www.dynamics.group.shef.ac.uk/people/keith/mec409.htm
- "The Scientist and Engineer's Guide to Digital Signal Processing, copyright ©1997-1998 by Steven W. Smith. For more information visit the book's website at: www.DSPguide.com"
- *Boore, D. M. and J. J. Bommer (2005). Processing of strong-motion accelerograms: Needs, options and consequences, Soil Dynamics and Earthquake Engineering* ***25****,93-115*
- *Boore, D. M. (2005). On pads and filters: Processing strong-motion data, Bull. Seism. Soc. Am.* ***95****,745-750.*

Signals

Definition

- A **signal** is a function of independent variables such as time, distance, position, temperature and pressure.
- A signal carries information, and the objective of **signal processing** is to extract useful information carried by the signal.
- **Signal processing** is concerned with the mathematical representation of the signal and the algorithmic operation carried out on it to extract the information present.

Definition

- For most purposes of description and analysis, a signal can be defined simply as a mathematical function,

$$y = f(x)$$

where x is the independent variable which specifies the domain of the signal e.g.:

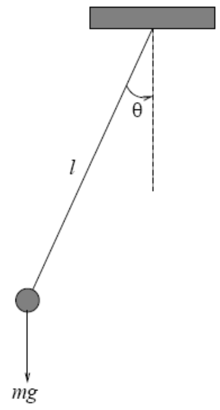
- $y=sin(\omega t)$ is a function of a variable in the **time domain** and is thus a time signal;

- $X(\omega)=1/(-m\omega^2+ic\omega+k)$ is a **frequency domain signal**;

- An image $I(x,y)$ is in the spatial domain.

Signal types

- For a simple pendulum as shown, basic definition is:



$$\theta(t) = \theta_m \sin(\omega t)$$

where θ_m is the peak amplitude of the motion and $\omega=\sqrt{l/g}$ with l the length of the pendulum and g the acceleration due to gravity.

- As the system has a constant amplitude (we assume no damping for now), a constant frequency (dictated by physics) and an initial condition (θ=0 when t=0), we know the value of $\theta(t)$ for all time.

Signal types

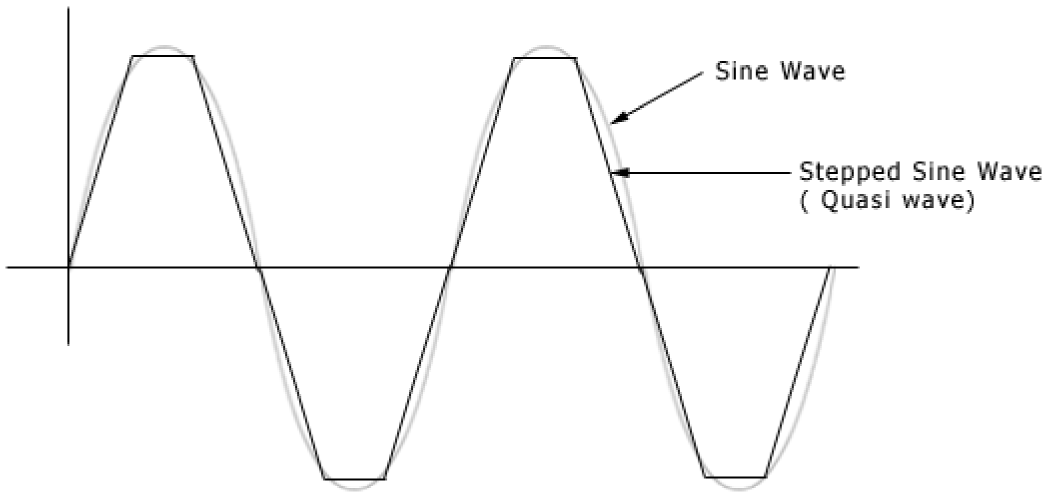
- Also, two identical pendula released from $\theta = \theta_0$ at t=0, will have the same motions at all time. There is no place for uncertainty here.

- If we can uniquely specify the value of θ for all time, *i.e.,* we know the underlying functional relationship between t and θ, the motion is **deterministic** or predictable. In other words, a signal that can be uniquely determined by a well defined process such as a mathematical expression or rule is called a **deterministic signal**.

- The opposite situation occurs if we know all the physics there is to know, but still cannot say what the signal will be at the next time instant-then the signal is **random** or **probabilistic.** In other words, a signal that is generated in a random fashion and can not be predicted ahead of time is called a **random signal**.

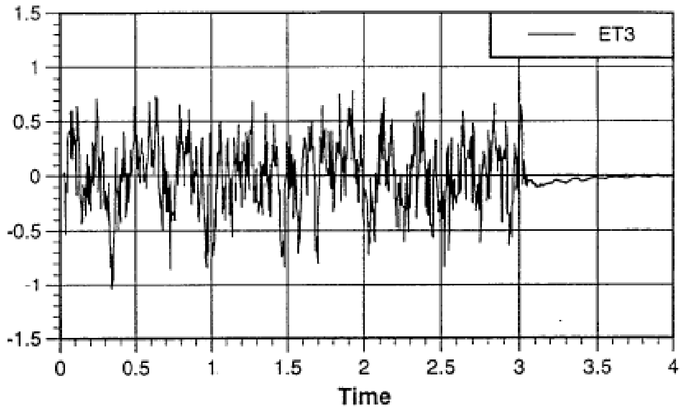
Signal types

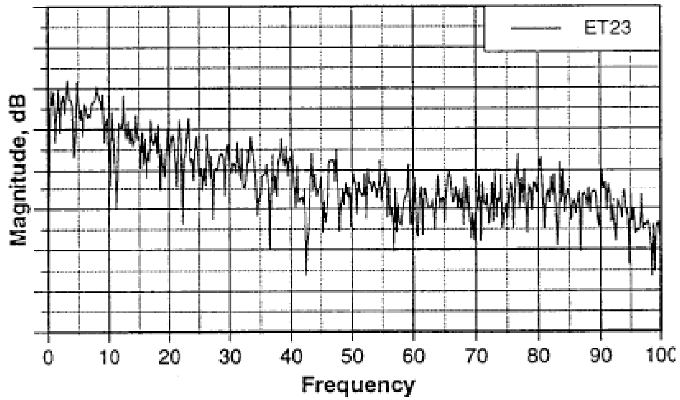
- Typical examples to deterministic signals are sine chirp and digital stepped sine.



Signal types

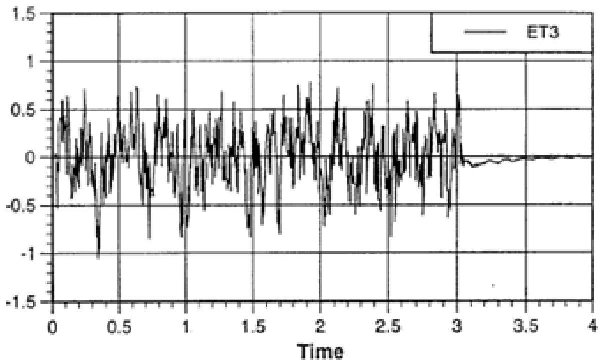
- Typical examples to random signals are random and burst random.

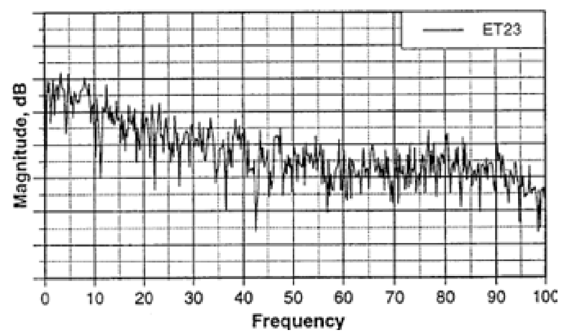




Signal types

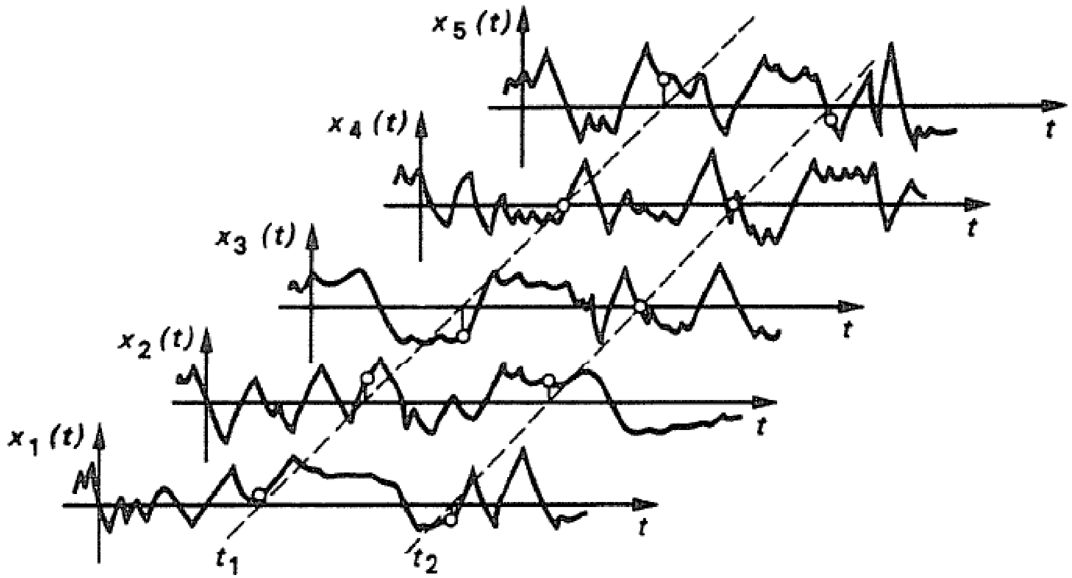
- **Random signals** are characterized by having many frequency components present over a wide range of frequencies.
- The amplitude versus time appears to vary rapidly and unsteadily with time.
- The 'shhhh' sound is a good example that is rather easy to observe using a microphone and oscillloscope. If the sound intensity is constant with time, the random signal is stationary, while if the sound intensity varies with time the signal is nonstationary. One can easily see and hear this variation while making the 'shhhh' sound.





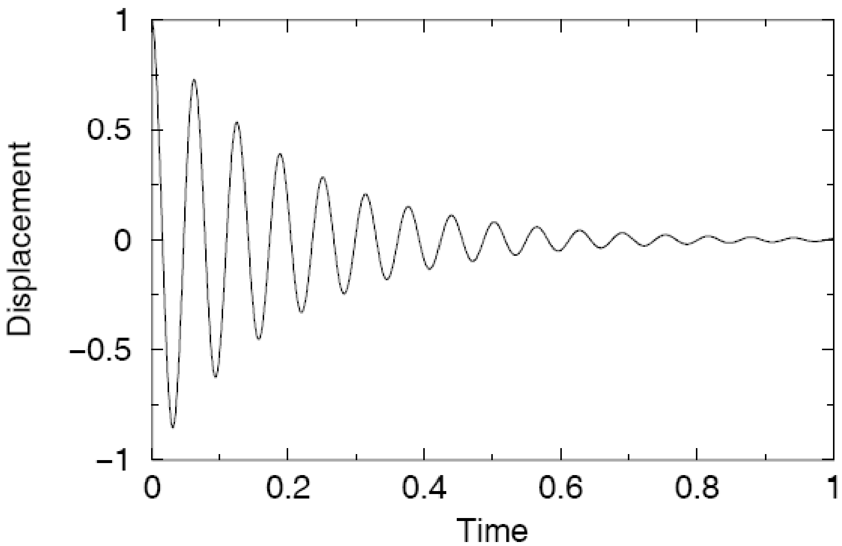
Signal types

- **Random signals** are characterized by analyzing the statistical characteristics across an ensemble of records. Then, if the process is ergodic, the time (temporal) statistical characteristics are the same as the ensemble statistical characteristics. The word temporal means that a time average definition is used in place of an ensemble statistical definition.



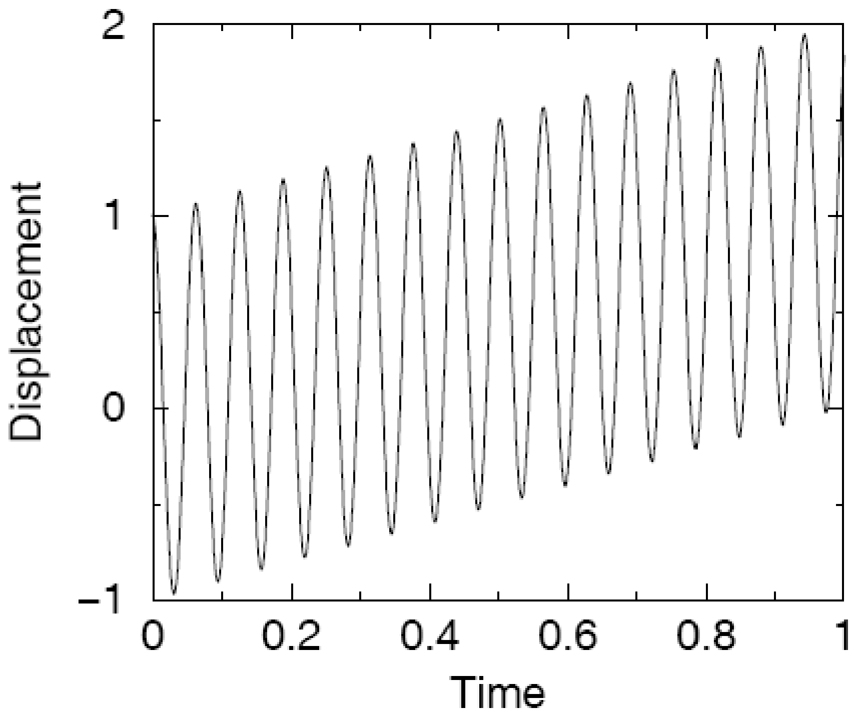
Signal types

- **Transient signals** may be defined as signals that exist for a finite range of time as shown in the figure. Typical examples are hammer excitation of systems, explosion and shock loading etc.



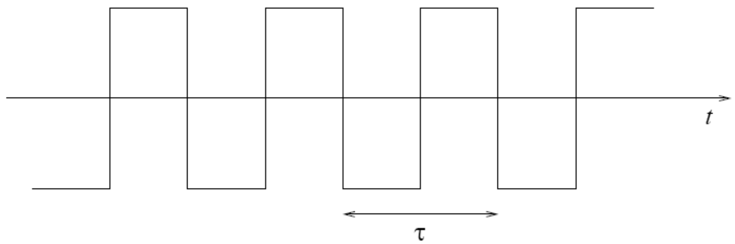
Signal types

- A signal with a time varying mean is an **aperiodic** signal.



Signal types

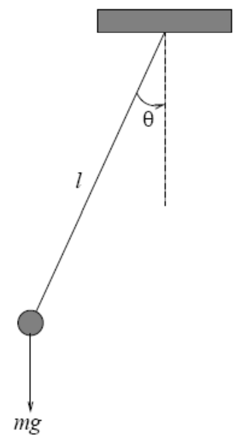
- It should be noted that periodicity does not necessarily mean a sinusoidal signal as shown in the figure.



$$S(t) = \sum_{i=1}^{\infty} s_i \sin\left(\frac{2\pi i}{\tau}.t\right)$$

- For a simple pendulum as shown, if we define the period τ by

$\tau = 2\pi/\omega = 2\pi\sqrt{g/l}$, then for the pendulum,



$$\theta(t) = \theta(t + \tau)$$

and such signals are defined as periodic.

Signal types

- A periodic signal is one that repeats itself in time and is a reasonable model for many real processes, especially those associated with constant speed machinery.

- **Stationary signals** are those whose average properties do not change with time. **Stationary signals** have constant parameters to describe their behaviour.

- **Nonstationary signals** have time dependent parameters. In an engine excited vibration where the engines speed varies with time; the fundamental period changes with time as well as with the corresponding dynamic loads that cause vibration.

Deterministic vs Random signals

- The signals can be further classified as **monofrequency** (sinusoidal) signals and **multifrequency** signals such as the square wave which has a functional form made up of an infinite superposition of different sine waves with periods $\tau, \tau/2, \tau/3, \ldots$

- **1 D signals** are a function of a single independent variable. The speech signal is an example of a 1 D signal where the independent variable is time.

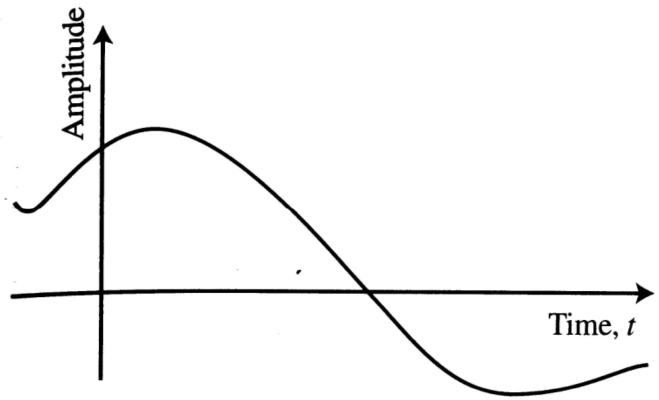
- **2D signals** are a function of two independent variables. An image signal such as a photograph is an example of a 2D signal where the two independent variables are the two spatial variables.

Classification of signals

- The value of a signal at a specific value of the independent variable is called its **amplitude**.
- The variation of the amplitude as a function of the independent variable is called its **waveform**.
- For a 1 D signal, the independent variable is usually labelled as time. If the independent variable is continuous, the signal is called a **continuous-time signal**. A continuous time signal is defined at every instant of time.
- If the independent variable is discrete, the signal is called a **discrete-time signal**. A discrete time signal takes certain numerical values at specified discrete instants of time, and between these specified instants of time, the signal is not defined. Hence, a discrete time signal is basically a sequence of numbers.

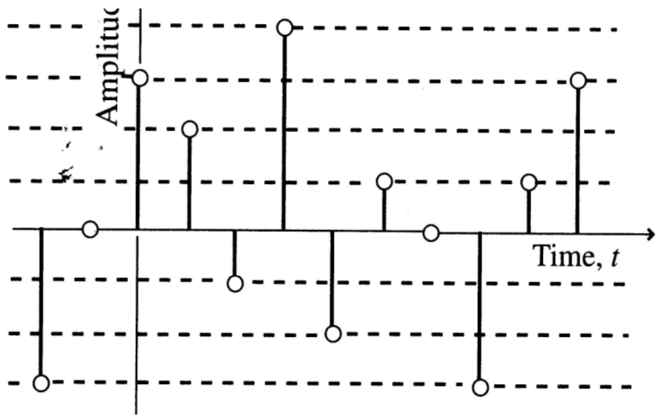
Classification of signals

- A continuous-time signal with a continuous amplitude is usually called an **analog signal**. A speech signal is an example of an analog signal.



(a)

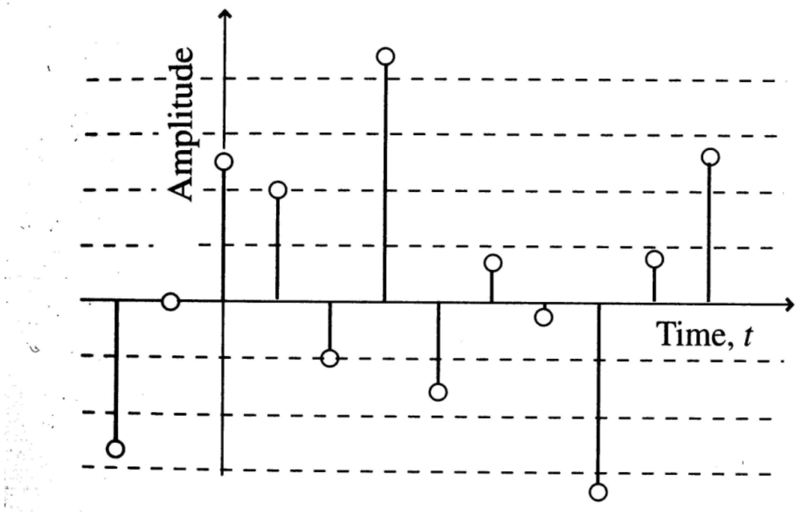
- A discrete time signal with discrete valued amplitudes represented by a finite number of digits is referred to as a **digital signal.**



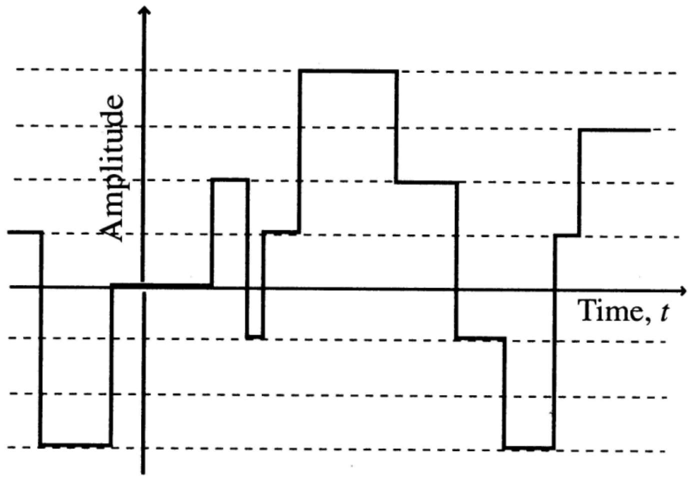
(b)

Classification of signals

- A discrete time signal with continuous valued amplitudes is called a **sampled-data signal**. A digital signal is thus a quantized sampled-data signal.

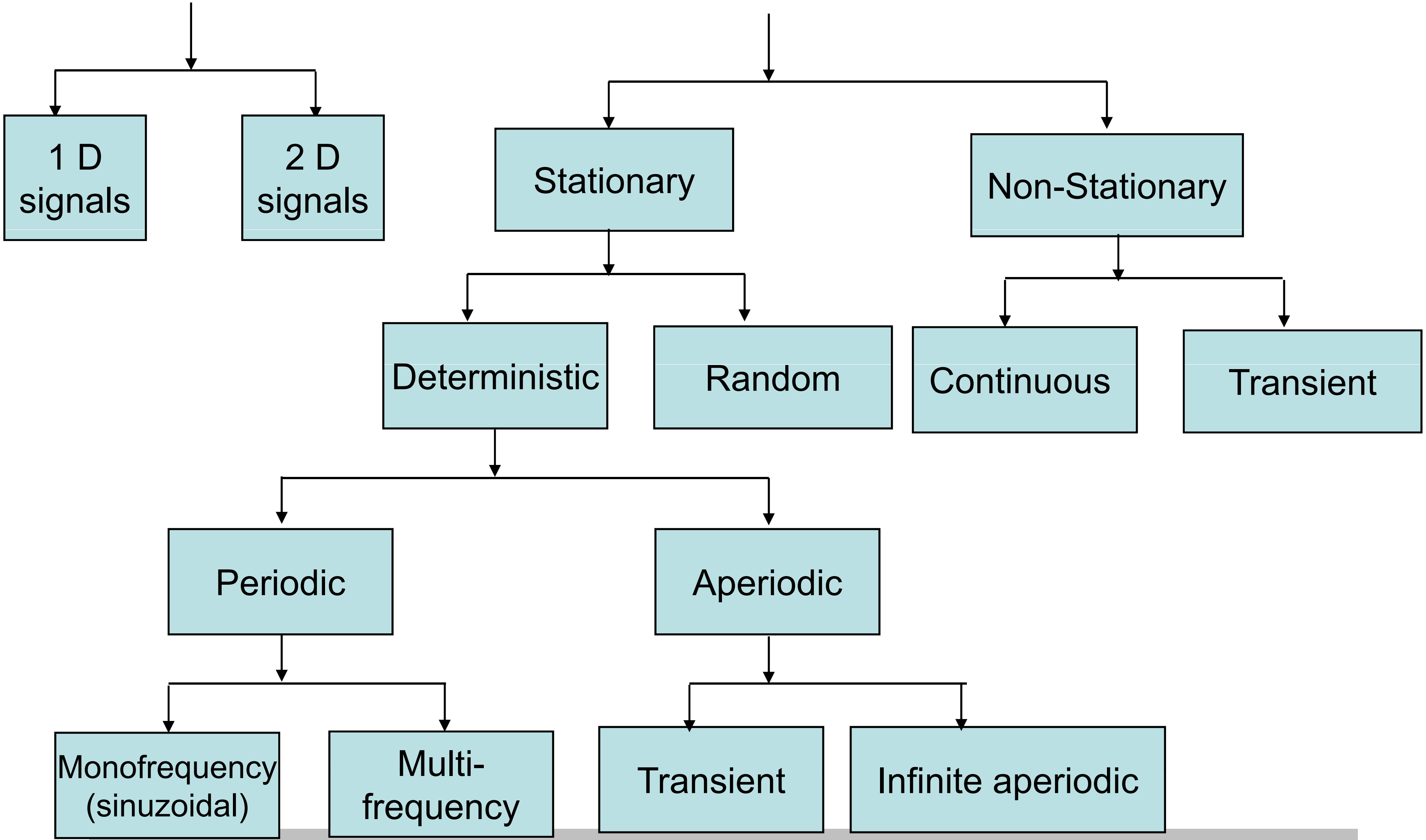


- A continuous-time signal with discrete valued amplitudes has been referred to as a **quantized boxcar signal**. This type of signal occurs in digital electronic circuits where the signal is kept at fixed level (usually one of two values) between two instants of clocking.



(d)

CLASSIFICATIONS OF SIGNALS



Typical Signal Processing Operations

Typical signal processing operations

- In the case of **analog signals**, most signal processing operations are usually carried out in the **time domain**.

- In the case of **discrete time signals**, **both time domain and frequency domain** applications are employed.

- In either case, the desired operations are implemented by a combination of some **elementary operations** such as:
  - Simple time domain operations
  - Filtering
  - Amplitude modulation

Simple Time Domain Operations

The three most basic time-domain signal operations are:

- **Scaling**
- **Delay**
- **Addition**

➢ **Scaling** is simply the multiplication of a signal by a positive or a negative constant. In the case of analog signals, this operation is usually called **amplification** if the magnitude of the multiplying constant, called **gain**, is greater than one. If the magnitude of the multiplying constant is less than one, the operation is called **attenuation.** Thus, if $x(t)$ is an analog signal, the scaling operation generates a signal $y(t)=\alpha x(t)$, where α is the multiplying constant.

Simple Time Domain Operations

The three most basic time-domain signal operations are:

- **Scaling**
- **Delay**
- **Addition**

➢ **Delay** operation generates a signal that is delayed replica of the original signal. For an analog signal $x(t)$, $y(t)=x(t-t_0)$ is the signal obtained by delaying $x(t)$ by the amount t_0, which is assumed to be a positive number. If t_0 is negative, then it is an **advance** operation.

Simple Time Domain Operations

The three most basic time-domain signal operations are:

- **Scaling**
- **Delay**
- **Addition**

➢ **Addition** operation generates a new signal by the addition of signals. For instance, $y(t)=x_1(t)+x_2(t)-x_3(t)$ is the signal generated by the addition of the three analog signals $x_1(t)$, $x_2(t)$ and $x_3(t)$.

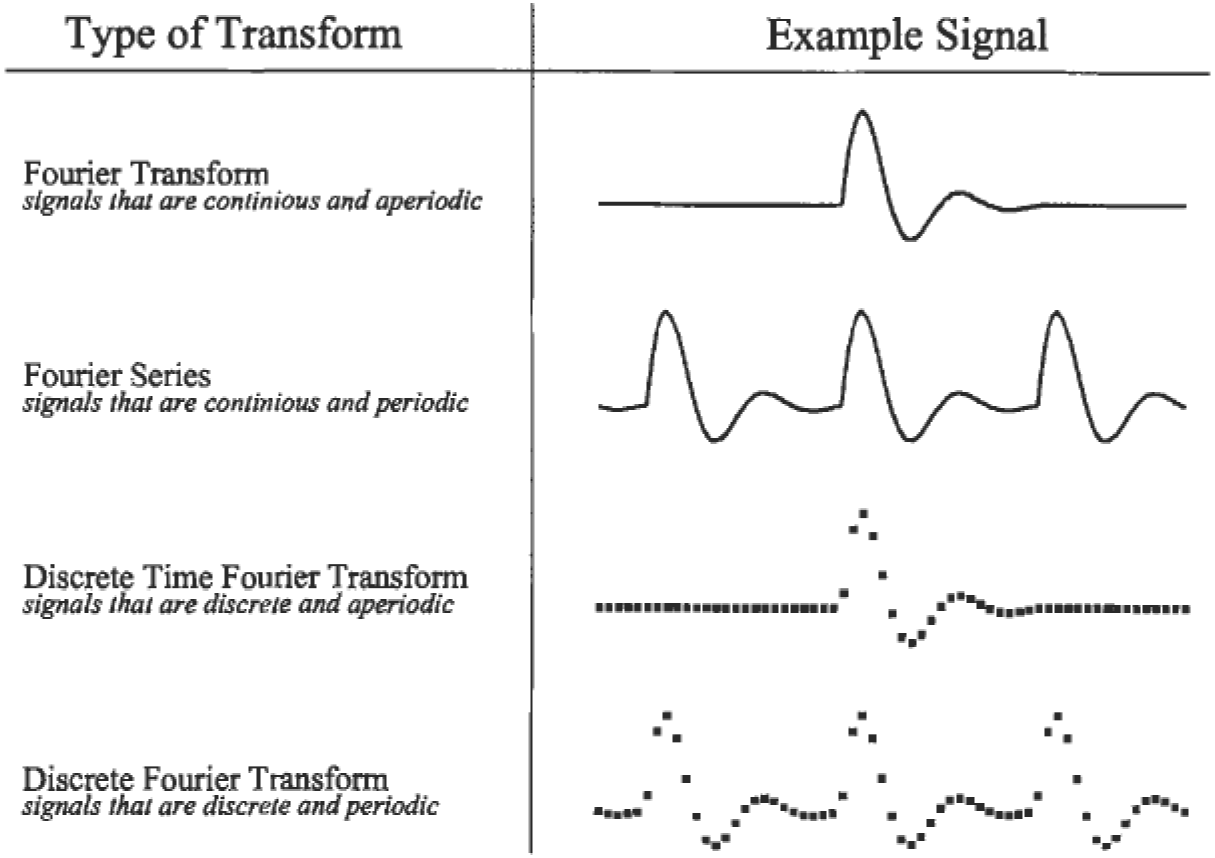
Fourier transforms

This chapter focuses on Fourier-series expansion, the discrete Fourier transform, properties of Fourier Transforms and Fast Fourier Transform

Fourier transforms

- Fourier analysis is a family of mathematical techniques, all based on decomposing signals into sinusoids.
- The discrete Fourier transform (DFT) is the family member used with ***digitized*** signals.
- Why are sinusoids used? A sinusoidal input to a system is guaranteed to produce a sinusoidal output. Only the amplitude and phase of the signal can change; the frequency and wave shape must remain the same. Sinusoids are the only waveform that have this useful property.
- The general term ***Fourier transform*** can be broken into four categories, resulting from the four basic types of signals that can be encountered.

Categories of Fourier Transforms



Fourier transforms

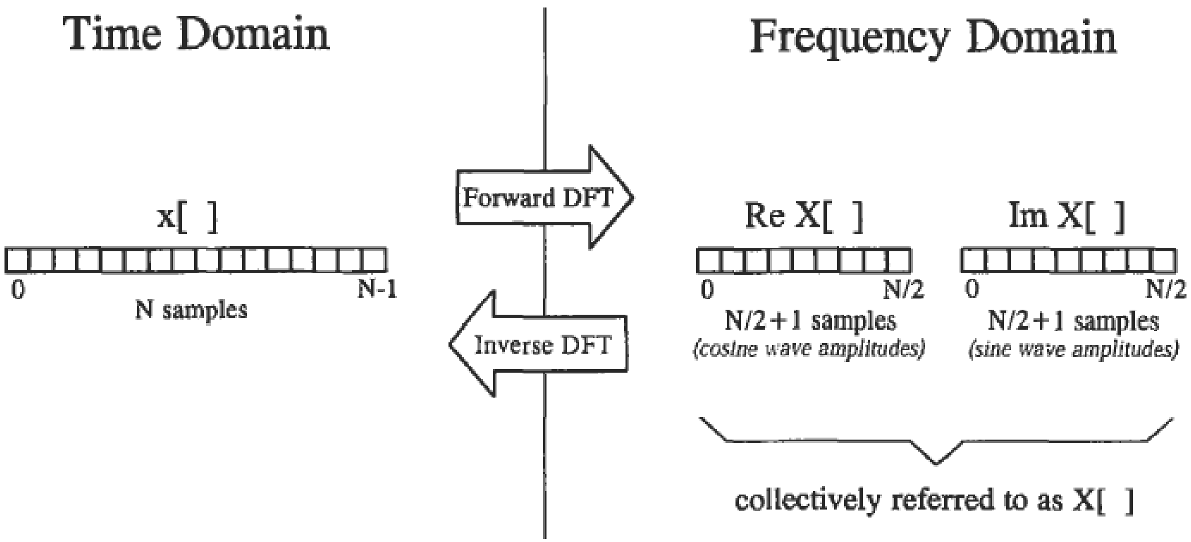
- These four classes of signals all extend to positive and negative ***infinity.*** What if you only have a finite number of samples stored in your computer, say a signal formed from 1024 points?

- There isn't a version of the Fourier transform that uses finite length signals. Sine and cosine waves are ***defined*** as extending from negative infinity to positive infinity. You cannot use a group of infinitely long signals to synthesize something finite in length. The way around this dilemma is to make the finite data look like an infinite length signal. This is done by imagining that the signal has an infinite number of samples on the left and right of the actual points. If all these "imagined" samples have a value of zero, the signal looks ***discrete*** and ***aperiodic,*** and the discrete time Fourier transform applies.

- As an alternative, the imagined samples can be a duplication of the actual 1024 points. In this case, the signal looks discrete and periodic, with a period of 1024 samples. This calls for the discrete Fourier transform to be used.

Time and frequency domains

- As shown in the figure, the **Discrete Fourier transform** changes an *N* point input signal into two *N/2* + 1 point output signals. The input signal contains the signal being decomposed, while the two output signals contain the ***amplitudes*** of the component sine and cosine waves. The input signal is said to be in the **time domain.** This is because the most common type of signal entering the DFT is composed of samples taken at regular intervals of ***time.*** The term “time domain” in Fourier analysis, may actually refer to samples taken over time. The term **frequency domain** is used to describe the amplitudes of the sine and cosine waves.



Time and frequency domains

- The **frequency domain** contains exactly the same information as the **time domain**, just in a different form. If you know one domain, you can calculate the other.

- Given the time domain signal, the process of calculating the frequency domain is called **decomposition, analysis,** the **forward DFT,** or simply, the **DFT.**

- If you know the frequency domain, calculation of the time domain is called **synthesis,** or the **inverse DFT.** Both synthesis and analysis can be represented in equation form and computer algorithms.

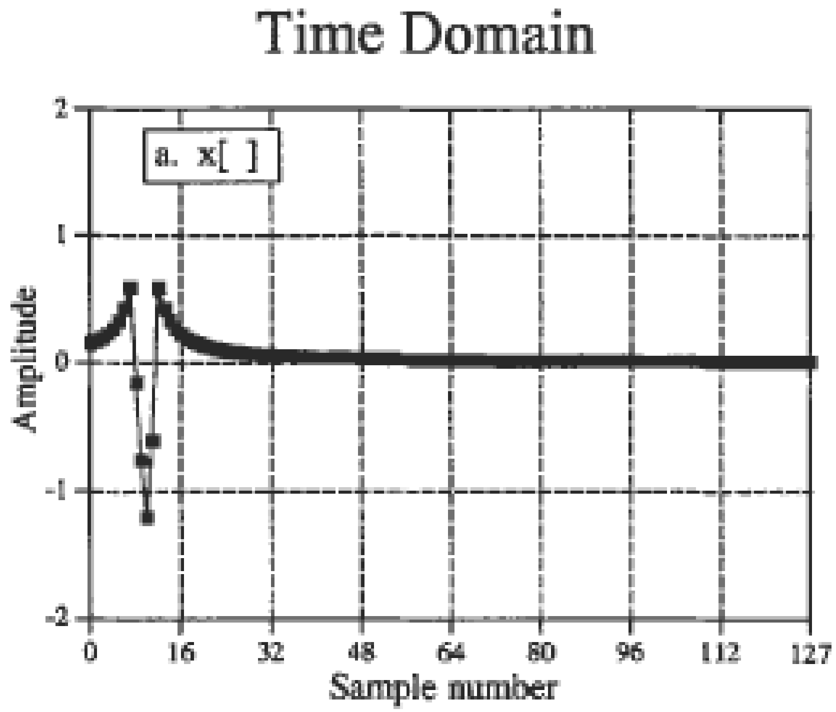
- The number of samples in the time domain is usually represented by the variable N. While N can be any positive integer, a power of two is usually chosen, i.e., 128, 256, 512, 1024, etc. There are two reasons for this. First, digital data storage uses binary addressing, making powers of two a natural signal length. Second, the most efficient algorithm for calculating the DFT, the Fast Fourier Transform (FFT), usually operates with N that is a power of two. Typically, N is selected between 32 and 4096. In most cases, the samples run from 0 to N-1 , rather than 1 to N.

Time and frequency domains

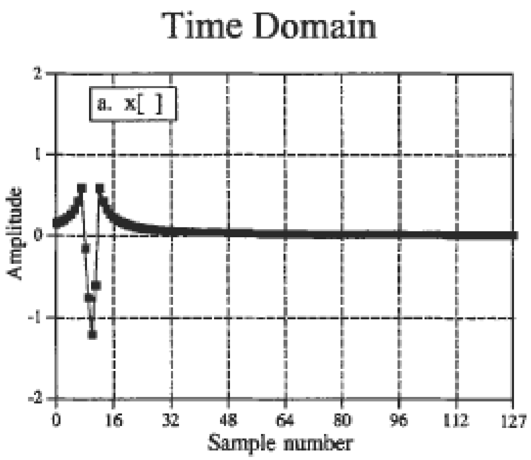
- Lower case letters represent time domain signals and upper case letters represent frequency domain signals.

- The figure shows an example DFT with *N = 128.* The time domain signa is contained in the array: ***x* [0]** to ***x* [127].** Notice that 128 points in the time domain corresponds to *65* points in each of the frequency domain signals, with the frequency indexes running from 0 to 64.

- That is, *N* points in the time domain corresponds to ***N/2* + *1*** points in the frequency domain (not *N/2* points). Forgetting about this extra point is a common bug in DFT programs.



Time and frequency domains



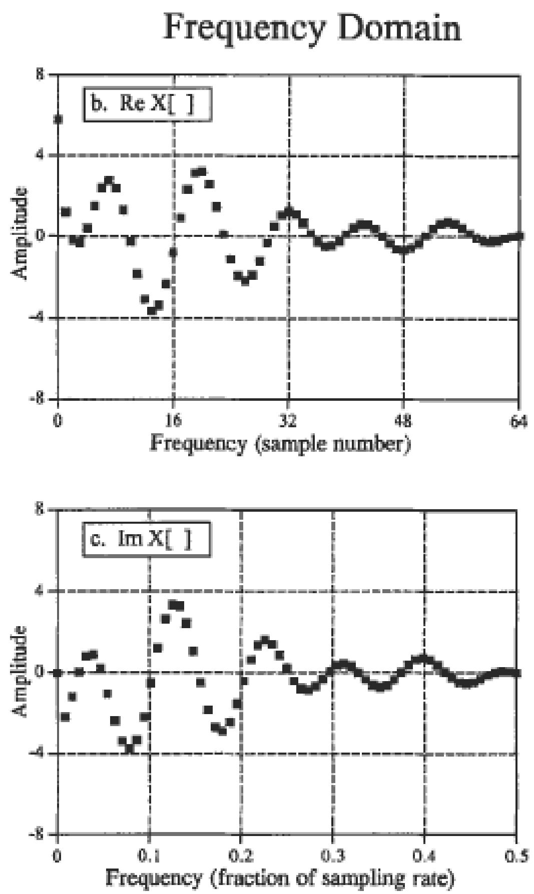


FIGURE 8-4
Example of the DFT. The DFT converts the time domain signal, $x[\ ]$, into the frequency domain signals, $ReX[\ ]$ and $ImX[\ ]$. The horizontal axis of the frequency domain can be labeled in one of three ways: (1) as an array index that runs between 0 and $N/2$, (2) as a fraction of the sampling frequency, running between 0 and 0.5, (3) as a natural frequency, running between 0 and π. In the example shown here, (b) uses the first method, while (c) use the second method.

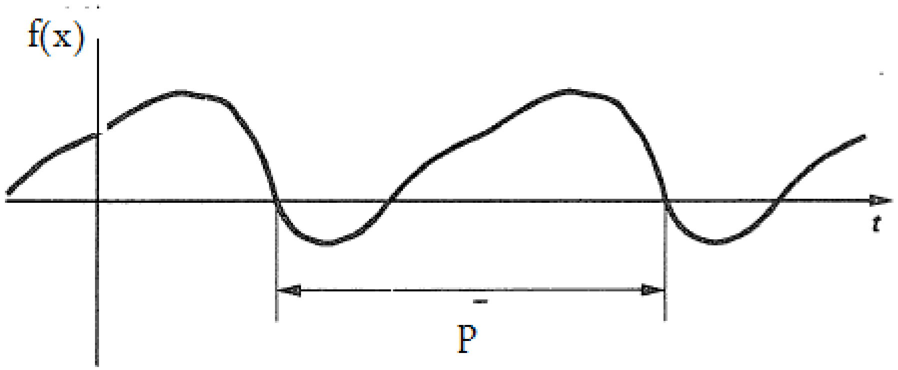
Fourier series expansion

- Fourier series are infinite series designed to represent general periodic functions in terms of simple ones, namely cosines and sines.

- A function f(x) is called a periodic function if f(x) is defined for all real x and if there is a positive number p, called a period of f(x), such that

$$f(x+p)=f(x)$$

- The graph of such a function is obtained by periodic repetition of its graph in any interval of length p.



Fourier series expansion

- Familiar periodic functions are the cosine and sine functions. Examples of functions that are not periodic are:

$$x, x^2, x^3, e^x, \cosh x, \ln x$$

- If f(x) has period p, it also has the period 2p because the equation

$$f(x+p) = f(x)$$

implies that

$$f(x+2p) = f([x+p]+p) = f(x+p) = f(x)$$

thus for any integer n=1,2,3,...

$$f(x+np) = f(x)$$

Fourier series expansion

- Furthermore if f(x) and g(x) have period p, then af(x)+bg(x) with any constants a and b also has the period p.Our problem in the first few slides will be the representation of various functions f(x) of period 2π in terms of the simple functions

$$1, \quad \cos x, \quad \sin x, \quad \cos 2x, \quad \sin 2x, \cdots, \quad \cos nx, \quad \sin nx$$

- All these functions have the period 2π. They form the so called trigonometric system. The figure shows these functions all have period 2π except for the constant 1, which is periodic with any period.

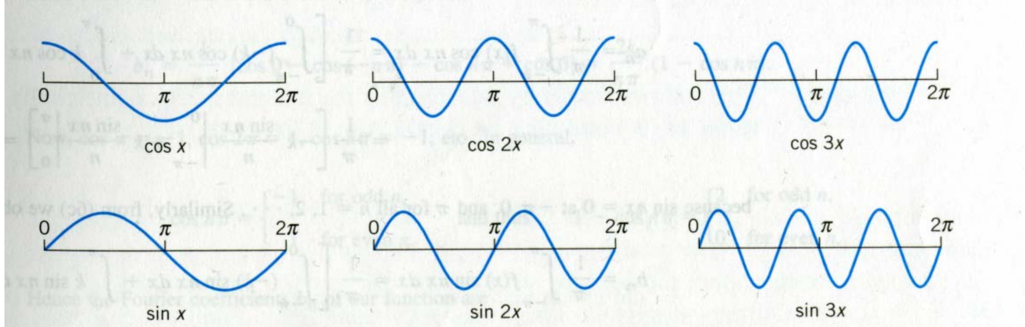


Fig. 256. Cosine and sine functions having the period 2π

Fourier series expansion

- The series to be obtained will be a trigonometric series, that is a series of the form:

$$a_o + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots = a_0 + \sum_{n=}^{\infty} \left(a_n \cos nx + b_n \sin nx \right)$$

- Here $a_o, a_1, b_1, a_2, b_2, \cdots$ are constants called the coefficients of the series. We see that each term has the period 2π. Hence if the coefficients are such that the series converges, its sum will be a function of period 2 π.
- Now suppose that f(x) is a given function of period 2 π and is such that it can be represented by a series as above which converges and moreover has the sum f(x). Then using the equality sign, we write:

$$f(x) = a_0 + \sum_{n=}^{\infty} \left(a_n \cos nx + b_n \sin nx \right)$$

Fourier series expansion

- The equation

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx \right)$$

is called the Fourier series of f(x). We shall prove that in this case, the coefficicents of the above equation are the so called Fourier coefficients of f(x) given by the **Euler formulas**.

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx \quad n = 1,2,\cdots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx \quad n = 1,2,\cdots$$

Fourier series expansion

- Let f(x) be periodic with period 2π and piecewise continuous in the interval $-\pi \le x \le \pi$. Furthermore, let f(x) have a left hand derivative and a right hand derivative at each point of that interval. Then the Fourier series of

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx \right)$$

with coefficients

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx \quad n = 1,2,\cdots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx \quad n = 1,2,\cdots$$

converges. Its sum is f(x) except at points xo where f(x) is discontinuous. There the sum of the series is the average of the left and right limits of f(x) at xo.

Fourier series expansion

- The left hand limit of f(x) at xo is defined as the limit of f(x) as x approaches xo from the left and is commonly denoted by f(xo-h). Thus,

$$f(x_o - h) = \lim_{h \to 0} f(x_o - h) \text{ as } h \to 0 \text{ through positive values.}$$

- The right hand limit of f(x) at xo is defined as the limit of f(x) as x approaches xo from the right and is commonly denoted by f(xo+h). Thus,

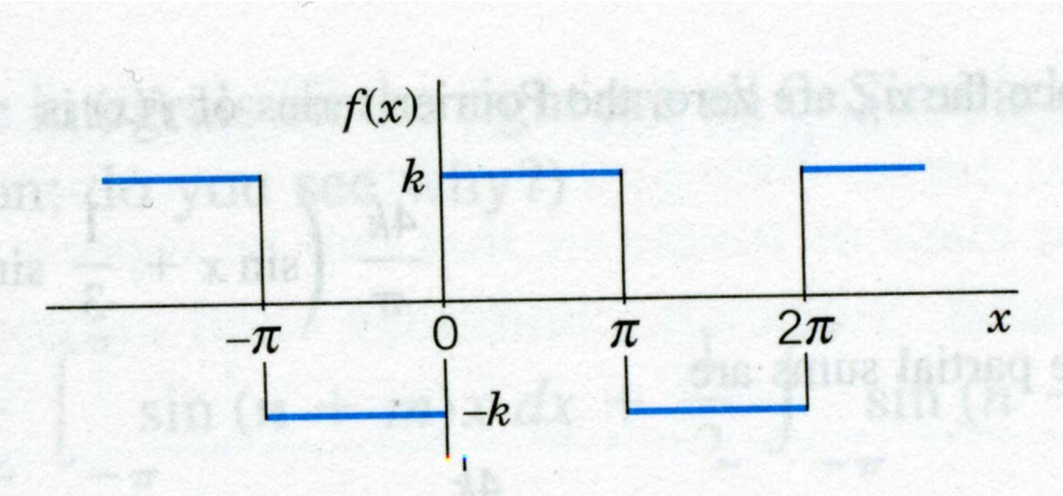
$$f(x_o + h) = \lim_{h \to 0} f(x_o + h) \text{ as } h \to 0 \text{ through positive values.}$$

Example

- Find the Fourier coefficients of the periodic function f(x) in the figure. The formula is:

$$f(x) = \begin{cases} -k & if \quad -\pi < x < 0 \\ k & if \quad 0 < x < \pi \end{cases}$$

$$and \quad f(x + 2\pi) = f(x)$$



Example

- Find the Fourier coefficients of the periodic function f(x) in the figure. The formula is:

$$f(x) = \begin{cases} -k & if \quad -\pi < x < 0 \\ k & if \quad 0 < x < \pi \end{cases}$$
$$and \quad f(x+2\pi) = f(x)$$

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx$$

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx = \frac{1}{2\pi}\int_{-\pi}^{0}(-k)dx + \frac{1}{2\pi}\int_{0}^{\pi}(k)dx$$
$$= \frac{1}{2\pi}(-kx)\Big|_{-\pi}^{0} + \frac{1}{2\pi}(kx)\Big|_{0}^{\pi} = -\frac{1}{2\pi}k\pi + \frac{1}{2\pi}k\pi = 0$$

- The above can also be seen without integration, since the area under the curve of f(x) between -π and π is zero.

Example

- From $$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx dx$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx dx = \frac{1}{\pi}\left[\int_{-\pi}^{0}(-k)\cos nx dx + \int_{0}^{\pi}(k)\cos nx dx\right]$$

$$= \frac{1}{\pi}\left[(-k\frac{\sin nx}{n})\Big|_{-\pi}^{0} + (k\frac{\sin nx}{n})\Big|_{0}^{\pi}\right] = 0$$

because $\sin nx = 0$ at $-\pi$, 0, and π for all n = 1,2,....Similarly:

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx dx$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx dx = \frac{1}{\pi}\left[\int_{-\pi}^{0}(-k)\sin nx dx + \int_{0}^{\pi}(k)\sin nx dx\right]$$

$$= \frac{1}{\pi}\left[k\frac{\cos nx}{n}\Big|_{-\pi}^{0} - k\frac{\cos nx}{n}\Big|_{0}^{\pi}\right]$$

Since $\cos(-\alpha) = \cos\alpha$ and $\cos 0 = 1$, this yields:

Example

$$b_n = \frac{k}{n\pi}\left[\cos 0 - \cos(-n\pi) - \cos n\pi + \cos 0\right]$$

$$= \frac{k}{n\pi}\left[2 - 2\cos n\pi\right]$$

$$= \frac{2k}{n\pi}\left(1 - \cos n\pi\right)$$

Now $\cos(\pi) = -1$, $\cos 2\pi = 1$, $\cos(3\pi) = -1$ etc in general

$$\cos n\pi = \begin{cases} -1 \text{ for odd n} \\ 1 \text{ for even n} \end{cases}$$

and thus $1 - \cos n\pi = \begin{cases} 2 \text{ for odd n} \\ 0 \text{ for even n} \end{cases}$

Example

Hence the Fourier coefficients b_n of our function are:

$$b_1 = \frac{4k}{\pi}, \quad b_2 = 0, \quad b_3 = \frac{4k}{3\pi}, \quad b_4 = 0, \quad b_5 = \frac{4k}{5\pi}, \cdots$$

Since the a_n are zero, the Fourier series of *f(x)* is:

$$\frac{4k}{\pi}\left(\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \cdots\right)$$

The partial sums are:

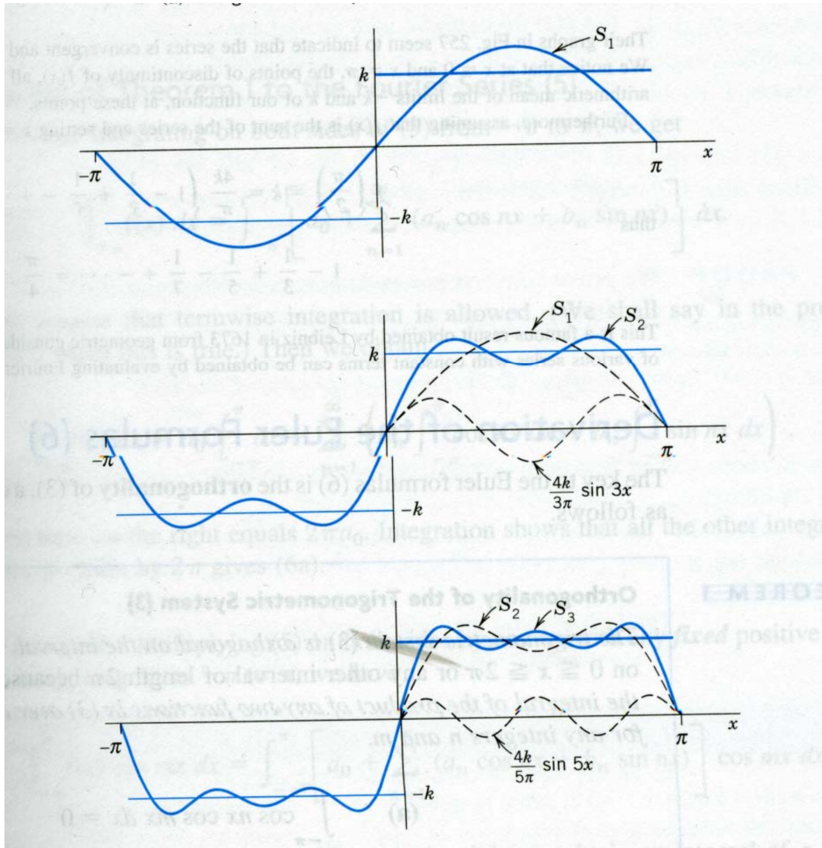
$$S_1 = \frac{4k}{\pi}\sin x$$

$$S_2 = \frac{4k}{\pi}\left(\sin x + \frac{1}{3}\sin 3x\right) \textit{ etc.}$$

- Their graph seems to indicate that the series is convergent and has the sum f(x), the given function.

Example

- We notice that at x=0 and x=π, the points of discontinuity of f(x), all partial sums have the value zero, the arithmetic mean of the limits k and –k of our function at these points.



(b) The first three partial sums of the corresponding Fourier series

Derivation of the Euler formulas

- The key to the Euler formulas is the orthogonality of

$$1, \quad \cos x, \quad \sin x, \quad \cos 2x, \quad \sin 2x, \cdots, \quad \cos nx, \quad \sin nx$$

a concept of basic importance as follows:

THEOREM 1: The trigonometric system above is orthogonal on the interval $-\pi \leq x \leq \pi$ (hence also on $0 \leq x \leq 2\pi$ or any other interval of length 2π because of periodicity); that is the integral of the product of any two functions in $1, \quad \cos x, \quad \sin x, \quad \cos 2x, \quad \sin 2x, \cdots, \quad \cos nx, \quad \sin nx$ over that interval is zero, so that for any integers n and m,

$$(a) \int_{-\pi}^{\pi} \cos nx \cos mx dx = 0 \quad (n \neq m)$$

$$(b) \int_{-\pi}^{\pi} \sin nx \sin mx dx = 0 \quad (n \neq m)$$

$$(c) \int_{-\pi}^{\pi} \sin nx \cos mx dx = 0 \quad (n \neq m \quad or \quad n = m)$$

Fourier series expansion of any period p=T

- The functions considered so far had period 2π, for the simplicity of the formulas. However, we will mostly use the variable time t and work with functions x(t) with period T. We now show that the transition from 2π to period T is quite simple. The idea is simply to find and use a change of scale that gives from a function f(x) of period 2π to a function of period T.
- In the equations below we can write the change of scale as: x=kt with n such that the old period x=2 π gives for the new variable t the new period t=T. Thus

2 π=kT hence k=2 π/T and x=kt= 2πt/T. This implies dx= 2πdt/T which upon substitution into

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx \qquad\qquad a_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)dt$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nxdx \quad n=1,2,\cdots \qquad\Rightarrow\qquad a_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos\frac{2\pi nt}{T}dt \quad n=1,2,\cdots$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nxdx \quad n=1,2,\cdots \qquad\qquad b_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin\frac{2\pi nt}{T}dt \quad n=1,2,\cdots$$

Fourier series expansion

- Since we will mostly use the variable time t and in the frequency domain $2\pi n/T$, the equation

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx \right)$$

can be written as follows:

$$x(t) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{2\pi nt}{T} + b_n \sin \frac{2\pi nt}{T} \right)$$

The coefficients in this case can be written as shown on the rhs rather than the lhs.

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx \quad n = 1,2,\cdots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx \quad n = 1,2,\cdots$$

$$a_0 = \frac{1}{T} \int_{-T/2}^{T/2} x(t) dx$$

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos \frac{2\pi nt}{T} dt \quad n = 1,2,\cdots$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin \frac{2\pi nt}{T} dt \quad n = 1,2,\cdots$$

Fourier series expansion of the periodic loading

- For a function *x(t)* defined on the interval [-τ/2, τ/2], we have the representation on that interval

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos\left[\left(\frac{2\pi n}{\tau}\right) t\right] + \sum_{n=1}^{\infty} b_n \sin\left[\left(\frac{2\pi n}{\tau}\right) t\right]$$

- The coefficients are obtained as follows: Consider the integral

$$I_m = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t) \cos\left[\left(\frac{2\pi m}{\tau}\right) t\right] dt$$

- When *x(t)* is substituted from the first equation, this integral breaks down into $I_m^{(1)}$, $I_m^{(2)}$ and $I_m^{(3)}$. The first is:

$$I_m^{(1)} = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} a_0 \cos\left[\left(\frac{2\pi m}{\tau}\right) t\right] dt = \left[\frac{a_0 \tau}{2\pi m} \sin\left[\left(\frac{2\pi m}{\tau}\right) t\right]\right]_{-\frac{\tau}{2}}^{\frac{\tau}{2}}$$

Fourier series expansion of the periodic loading

$$= \frac{a_0 \tau}{2\pi m}\left[\sin(\pi m) - \sin(-\pi m)\right]$$

$$= \frac{a_0 \tau}{\pi m}\sin(\pi m)$$

but this is zero as *sin*(π*m*)=0 for all *m*. The second integral is:

$$I_m^{(2)} = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \left[\sum_{n=1}^{\infty} a_n \cos\left[\left(\frac{2\pi n}{\tau}\right)t\right]\right] \cos\left[\left(\frac{2\pi m}{\tau}\right)t\right] dt$$

- Assuming we can change the order of integration and summation we obtain

$$I_m^{(2)} = \sum_{n=1}^{\infty} a_n \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \cos\left[\left(\frac{2\pi n}{\tau}\right)t\right] \cos\left[\left(\frac{2\pi m}{\tau}\right)t\right] dt$$

- Using the identity $2\cos A \cos B = \cos(A+B) + \cos(A-B)$

Fourier series expansion of the periodic loading

$$I_m^{(2)} = \sum_{n=1}^{\infty} \frac{a_n}{2} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \left(\cos\left[\left(\frac{2\pi(n+m)}{\tau}\right) t\right] + \cos\left[\left(\frac{2\pi(n-m)}{\tau}\right) t\right]\right) dt$$

$$= \sum_{n=1}^{\infty} \frac{a_n}{2} \left[\frac{\tau}{2\pi(n+m)} \sin\left[\left(\frac{2\pi(n+m)}{\tau}\right) t\right] + \frac{\tau}{2\pi(n-m)} \sin\left[\left(\frac{2\pi(n-m)}{\tau}\right) t\right]\right]_{-\frac{\tau}{2}}^{\frac{\tau}{2}}$$

$$= \sum_{n=1}^{\infty} \frac{a_n}{2} \left[\frac{\tau}{\pi(n+m)} \sin\left[\pi(n+m)\right] + \frac{\tau}{\pi(n-m)} \sin\left[\pi(n-m)\right]\right]$$

- Now if *n* and *m* are different integers then *n-m* and *n+m* are both nonzero integers and the sine terms in the last expression vanish. If *n* and *m* are equal, we have a problem with the second term above. We could use a limit argument but it is simpler to go back to the first equation with *n=m*.

$$I_m^{(2)} = \frac{a_m}{2} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \left(\cos\left[\left(\frac{4\pi m}{\tau}\right) t\right] + 1\right) dt$$

$$a_m \left[\frac{\tau}{8\pi m} \sin\left[\left(\frac{4\pi m}{\tau}\right) t\right] + \frac{t}{2}\right]_{-\frac{\tau}{2}}^{\frac{\tau}{2}}$$

Fourier series expansion of the periodic loading

- This breaks down to:

$$I_m^{(2)} = a_m \frac{\tau}{2}$$

- An orthogonality relation has been proved.

$$\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \cos\left[\left(\frac{2\pi n}{\tau}\right)t\right] \cos\left[\left(\frac{2\pi m}{\tau}\right)t\right] dt = 0 \qquad \text{for} \quad m \neq n$$

$$= \frac{\tau}{2} \qquad \text{for} \quad m = n \neq 0$$

$$= \tau \qquad \text{for} \quad m = n = 0$$

- A similar analysis for the third integral gives

$$I_m^{(3)} = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \left[\sum_{n=1}^{\infty} b_n \sin\left[\left(\frac{2\pi n}{\tau}\right)t\right]\right] \cos\left[\left(\frac{2\pi m}{\tau}\right)t\right] dt$$

$$I_m^{(3)} = 0$$

Fourier series expansion of the periodic loading

- Derives along the way the orthogonality relation

$$\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \sin\left[\left(\frac{2\pi n}{\tau}\right)t\right] \cos\left[\left(\frac{2\pi m}{\tau}\right)t\right] dt = 0$$

- As

$$I_m = I_m^{(1)} + I_m^{(2)} + I_m^{(3)}.$$

- We have $$I_m = a_m \frac{\tau}{2} \qquad \text{for } m = n \neq 0$$

$$I_m = a_0 \tau \qquad \text{for } m = n = 0$$

- Or in terms of the original expression:

$$\frac{2}{\tau} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t) \cos\left[\left(\frac{2\pi m}{\tau}\right)t\right] dt = a_m \qquad \text{for } m = n \neq 0$$

Fourier series expansion

- If m=n=0 in the below equation:

$$a_m\tau = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t)\cos\left(\frac{2\pi m}{\tau}t\right)dt \qquad \text{for } m = n = 0 \qquad\qquad a_0 = \frac{1}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t)dt$$

- Performing the same operations using a multiplier of $sin(2\pi m/\tau)$ gives:

$$b_m = \frac{2}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t)\sin\left[\left(\frac{2\pi m}{\tau}\right)t\right]dt$$

via the orthogonality relation:

$$\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \sin\left[\left(\frac{2\pi n}{\tau}\right)t\right]\sin\left[\left(\frac{2\pi m}{\tau}\right)t\right]dt = \frac{\tau}{2}\delta_{mn}$$

where δ_{mn} is called the Kronecker delta and has the properties:

$$\delta_{mn} = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

Fourier series expansion of the periodic loading

Fourier Series Expansion in Exponential Form

- Recall the standard Fourier series in terms of

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega t) + \sum_{n=1}^{\infty} b_n \sin(n\omega t)$$

- Now suppose we apply de Moivres Theorem

$$\cos(n\omega t) = \frac{1}{2}(e^{in\omega t} + e^{-in\omega t})$$

$$\sin(n\omega t) = \frac{1}{2i}(e^{in\omega t} - e^{-in\omega t})$$

$$\sum_{n=1}^{\infty} a_n \cos(n\omega t) + \sum_{n=1}^{\infty} b_n \sin(n\omega t) = \frac{a_n}{2}(e^{in\omega t} + e^{-in\omega t}) + \frac{b_n}{2i}(e^{in\omega t} - e^{-in\omega t})$$

Fourier series expansion of the periodic loading

$$= \frac{1}{2}(a_n - ib_n)e^{in\omega t} + \frac{1}{2}(a_n + ib_n)e^{-in\omega t}$$

- This allows us to write the equation

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega t) + \sum_{n=1}^{\infty} b_n \sin(n\omega t)$$

in the following form:

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega t}$$

where

$$c_n = (a_n - ib_n)/2 \qquad n > 0$$

$$c_n = (a_n + ib_n)/2 = c_{-n}^* \qquad n < 0$$

$$c_0 = a_0 \qquad n = 0$$

Fourier series expansion

- Using the equations

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos\frac{2\pi nt}{T}\,dt \quad n = 1,2,\cdots$$

$$b_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin\frac{2\pi nt}{T}\,dt \quad n = 1,2,\cdots$$

in $c_n = (a_n - ib_n)/2 \qquad n > 0$

$$c_n = \frac{1}{2}(a_n - ib_n) = \frac{1}{2}\frac{2}{T}\left(\int_{-T/2}^{+T/2} x(t)\cos\frac{2\pi nt}{T}\,dt - i\int_{-T/2}^{+T/2} x(t)\sin\frac{2\pi nt}{T}\,dt\right)$$

$$= \frac{1}{T}\int_{-T/2}^{+T/2} x(t)\left(\cos\frac{2\pi nt}{T} - i\sin\frac{2\pi nt}{T}\right)dt$$

Gives:

$$c_n = \frac{1}{T}\int_{-T/2}^{+T/2} x(t)e^{-i2\pi nt/T}\,dt$$

Fourier series expansion

- Alternatively, using the orthogonality equation

$$\int_{-T/2}^{T/2} e^{in\omega t} e^{im\omega t} dt = 0 \quad \text{for} \;\; n+m \neq 0$$

$$= T \;\; \text{for} \;\; n+m = 0$$

and multiplying the equation

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega t}$$

by *exp(imωt),* and integrating directly gives:

$$c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-in\omega t} dt$$

Fourier transform

- If we want to look at the spectral content of nonperiodic signals we have to let $\tau \to \infty$ as all the interval $t \in [-\infty, \infty]$ contains important information. Recall the exponential form of the Fourier series

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega t}$$

where

$$c_n = \frac{1}{\tau} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t) e^{-in\omega t} dt$$

Combining the above two equations give:

$$x(t) = \sum_{n=-\infty}^{\infty} \left\{ \frac{1}{\tau} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t') e^{-in\omega t'} dt' \right\} e^{in\omega t}$$

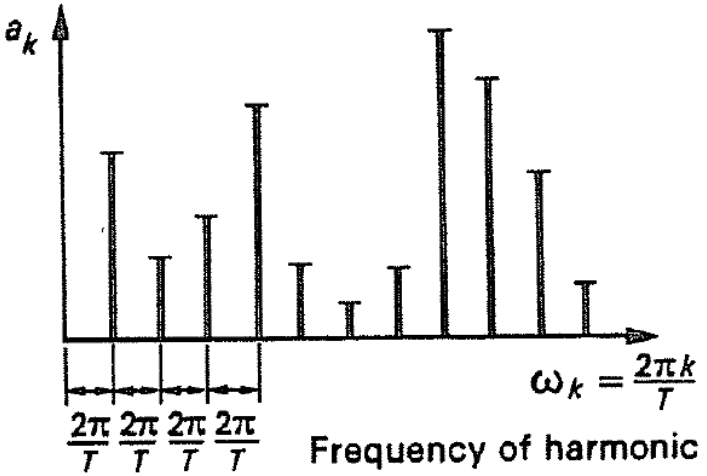
and we now have to let $\tau \to 0$

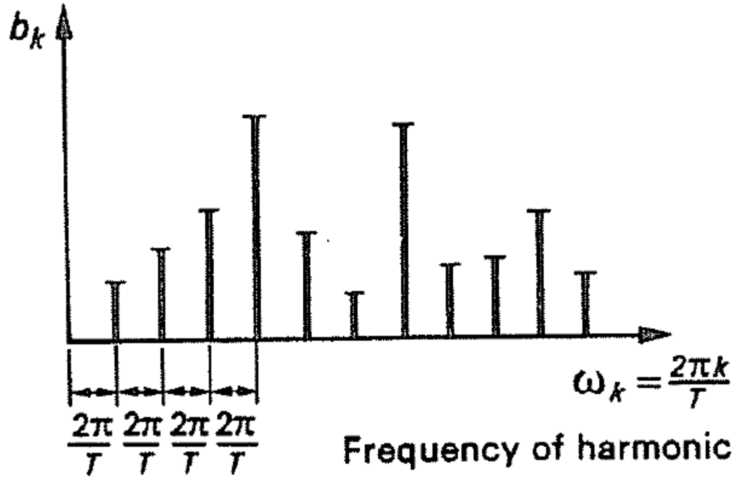
Fourier Transform

- If we suppose that the position of the t axis is adjusted so that the mean value of x(t) is zero. Then according to the first of the below equation the coefficient ao will be zero.

$$a_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)dx \qquad a_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos\frac{2\pi nt}{T}dt \quad n = 1,2,\cdots \qquad b_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin\frac{2\pi nt}{T}dt \quad n = 1,2,\cdots$$

- The remaining coefficients an and bn will in general all be different and their values may be illustrated graphically as shown.





Fourier Transform

Recall that the spacing between the frequency lines is

$$\omega = \Delta\omega = \frac{2\pi}{\tau}$$

so that the *kth* spectral line is at

$$\omega_k = k\Delta\omega$$

From the first equation, we see that

$$\frac{1}{\tau} = \frac{\Delta\omega}{2\pi}$$

The equation

$$x(t) = \sum_{n=-\infty}^{\infty} \left\{ \frac{1}{\tau} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t') e^{-in\omega t'} dt' \right\} e^{in\omega t}$$

becomes

$$x(t) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \left\{ \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t') e^{-i\omega_n t'} dt' \right\} e^{i\omega_n t} \Delta\omega$$

Fourier Transform

- As $\tau \to 0$, the ω_n become closer and closer together and the summation turns into an integral with $\Delta\omega = d\omega$ (assuming that *x(t)* is appropriately well behaved. In the limit

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x(t')e^{-i\omega t'}dt'\right\}e^{i\omega t}d\omega$$

- It follows that if we define

$$\mathcal{F}[x(t)] = X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt$$

where $\mathcal{F}$ denotes the Fourier transform then the first equation implies that

$$\mathcal{F}^{-1}[X(\omega)] = x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}d\omega$$

and this is the inverse Fourier transform.

Fourier Transform

- Note the formal similarity to the Laplace transform in fact we obtain the Fourier transform by letting $s = i\omega$ in the Laplace transform. The main difference between the two is the comparative simplicity of the inverse Fourier transform.

- $\{x(t), X(\omega)\}$ are a Fourier transform pair. As they are uniquely constructable from each other they must both encode the same information but in different domains. $X(\omega)$ expresses the frequency content of *x(t).* It is another form of spectrum. However note that it has to be a continuous function of ω in order to represent non-periodic functions.

Discrete Fourier Transform

- In reality not only will the signal be of finite duration but it will be sampled. It is not possible to store a continuous function on a computer as it would require infinite memory.
- What one usually does is take measurements from the signal at regular intervals say *t* seconds apart so the signal for manipulation takes the form of a finite vector of N samples

$$\{x_0, \ldots, x_{N-1}\} \qquad \text{where} \qquad x_i = x(t_i) = x(t_0 + i\Delta t)$$

where t is a reference time. If we take t_0=0 from now on we will have

$$t_i = i\Delta t$$

- How do we compute the spectrum of such a signal? We need the Discrete Fourier Transform DFT.

Discrete Fourier Transform

- Recall the exponential form of the Fourier series the spectral coefficients are

$$c_n = \frac{1}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t)e^{-\frac{2\pi i n}{\tau}t}dt$$

- In keeping with our notation for the Fourier transform we will relabel c_n by X_n from now on. Also the equation above is not in the most convenient form for the analysis so we will modify it slightly.
- Recall that *x(t)* is assumed periodic with period τ. Consider the integral

$$\frac{1}{\tau}\int_{-\frac{\tau}{2}}^{0} x(t)e^{-\frac{2\pi i n}{\tau}t}dt$$

Discrete Fourier Transform

- Let t′= t+τ, the integral becomes

$$\frac{1}{\tau}\int_{\frac{\tau}{2}}^{\tau 0} x(t'-\tau)e^{-\frac{2\pi i n}{\tau}(t'-\tau)}dt$$

but x(t′)=x(t′-τ) by periodicity. Also, $e^{-2\pi i n(t'-\tau)/\tau} = e^{-2\pi n i t/\tau}$

$$\frac{1}{\tau}\int_{-\frac{\tau}{2}}^{0} x(t)e^{-\frac{2\pi i n}{\tau}t}dt = \frac{1}{\tau}\int_{\frac{\tau}{2}}^{\tau} x(t)e^{-\frac{2\pi i n}{\tau}t}dt$$

$$X_n = \frac{1}{\tau}\int_{-\frac{\tau}{2}}^{0} x(t)e^{-\frac{2\pi i n}{\tau}t}dt + \frac{1}{\tau}\int_{0}^{\frac{\tau}{2}} x(t)e^{-\frac{2\pi i n}{\tau}t}dt$$

$$X_n = \frac{1}{\tau}\int_{0}^{\tau} x(t)e^{-\frac{2\pi i n}{\tau}t}dt$$

Discrete Fourier Transform

- Now, as we have only *x(t)* sampled at $t_r = r\Delta t$, we have to approximate the integral by a rectangular sum,

$$X_n = \frac{1}{\tau}\sum_{r=0}^{N-1} x(t_r)e^{-i\frac{2\pi n}{\tau}t_r}\Delta t = \frac{1}{\tau}\sum_{r=0}^{N-1} x_r e^{-i\frac{2\pi n}{\tau}r\Delta t}\Delta t$$

and as $\tau = N\Delta t$, this becomes,

$$X_n = \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-i\frac{2\pi n}{N}r}$$

- As we started off with only N independent quantities x_r we can only derive N independent spectral lines at most. This means we must have relations between the X_n . The simplest one is periodicity. Consider,

$$X_{n+N} = \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-i\frac{2\pi(n+N)}{N}r} = \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-i\frac{2\pi n}{N}r}e^{-i2\pi r} == \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-i\frac{2\pi n}{N}r} = X_n$$

Discrete Fourier Transform

- Therefore, it has been confirmed that we have at most N independent lines. In fact, there must be less than this as the X_n are the complex quantities. Given N real numbers, we can only specify $N/2$ complex numbers.
- Looking at the exponent in

$$X_n = \frac{1}{\tau}\int_0^{\tau} x(t)e^{-\frac{2\pi i n}{\tau}t}dt$$

if this is to be identified with the exponent $i\omega_n t$ of the Fourier transform, we must have,

$$\omega_n = n\Delta\omega = \frac{2\pi n}{\tau}$$

or

$$\Delta\omega = \frac{2\pi}{N\Delta t}$$

Discrete Fourier Transform

- Alternatively, if we specify the frequency spacing in Hertz

$$\Delta f = \frac{1}{N\Delta t}$$

- When n=0, the spectral line is given by: $X_0 = \frac{1}{N}\sum_{r=0}^{N-1} x_r$

  which is the arithmetic mean or DC component of the signal. Therefore X_0 corresponds to the frequency ω=0 as we might expect.

- This means that the highest frequency that we can represent is

$$\frac{N}{2}\Delta f = \frac{N}{2N\Delta t} = \frac{1}{2\Delta t} = \frac{f_s}{2}$$

  where f_s is the sampling frequency. This frequency is very important in signal processing and is called the **Nyquist Frequency.**

Discrete Fourier Transform

- This argument says that only the first half of the spectral lines are independent-so what are the second half? Consider:

$$X_{N-1} = \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-\frac{2\pi i(N-1)r}{N}} = \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-i2\pi r} e^{+i2\pi r/N}$$

but *r* is an integer, so $e^{-i2\pi r}=1$, this means,

$$X_{N-1} = \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{i2\pi r/N} = \left(\frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-i2\pi r/N}\right)^* = X_1^*$$

- A similar argument shows that generally,

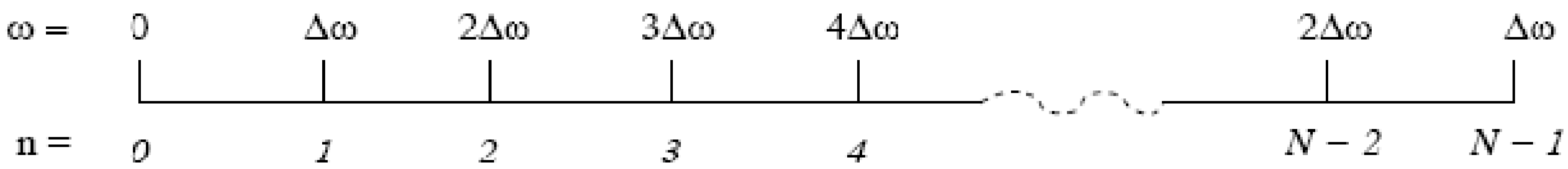
$$X_{N-k} = X_k^*$$

Discrete Fourier Transform

- Recall that it is a property of the Fourier Transform that

$$X(\omega)^* = X(-\omega)$$

- This means that the spectral coefficient X_{N-1} corresponds to the frequency $-\Delta\omega$ or more generally X_{N-k} corresponds to the frequency $-k\Delta\omega$. So the array X_n stores the frequency representation of the signal x_r as follows:



Discrete Fourier Transform

- What happens if there is a value of k for which $N\text{-}k=k?$ This implies that $N=2k$ so the number of sample points is even.

- Actually this turns out to be the most usual situation, in fact it is a requirement of the **Fast Fourier Transform** which we shall meet in the next lecture.

- In this case, we have

$$X^*_{N/2} = X_{N/2}$$

so this spectral line is real.

- This finally justifies our assertion that the maximum frequency represented is $N\Delta\omega/2$.

Discrete Fourier Transform

Proof of the inversion theorem

- Let us proove that

$$x_n = \sum_{r=0}^{N-1} X_r e^{i2\pi rn/N}$$

where

$$X_r = \frac{1}{N}\sum_{p=0}^{N-1} x_p e^{-i2\pi rp/N}$$

- If the second equation is substituted in the first equation, we obtain:

$$x_n = \sum_{r=0}^{N-1} \left\{ \frac{1}{N}\sum_{p=0}^{N-1} x_p e^{-i2\pi rp/N} \right\} e^{i2\pi rn/N}$$

Discrete Fourier Transform

- If we change the order of the summations, we obtain:

$$x_n = \sum_{p=0}^{N-1} \left\{ \frac{1}{N} \sum_{r=0}^{N-1} x_p e^{-i2\pi rp/N} e^{i2\pi rn/N} \right\}$$

which can be rewritten as:

$$x_n = \sum_{p=0}^{N-1} \left\{ \frac{1}{N} \sum_{r=0}^{N-1} x_p e^{i2\pi r(n-p)/N} \right\}$$

which can be reexpressed as follows provided the term in brackets is called *q*:

$$x_n = \sum_{p=0}^{N-1} \left\{ \frac{1}{N} \sum_{r=0}^{N-1} x_p \underbrace{\left[e^{i2\pi (n-p)/N} \right]}_{=q}{}^{r} \right\}$$

Discrete Fourier Transform

- Thus, if n≠p, within the equation

$$x_n = \sum_{p=0}^{N-1} \left\{ \frac{1}{N} \sum_{r=0}^{N-1} x_p \Big[\underbrace{e^{i2\pi(n-p)/N}}_{=q} \Big]^r \right\}$$

we have a geometric series as follows: $\sum_{r=0}^{N-1} q^r = 1 + q + q^2 + \ldots$

If we call the left hand side of the above equation s_r, we have:

$$s_r = 1 + q + q^2 + \ldots$$

If we multiply both sides of the above equation by q, and subtract the resulting equation from the above equation:

$$qs_r = q + \ldots . + q^r + q^{N-1+1}$$

$$s_r - qs_r = 1 - q^N$$

Discrete Fourier Transform

- We can obtain the value of s_r from here as:

$$s_r = \frac{1 - q^N}{1 - q}$$

Now, $q^N=1$, which can be proved as:

$q^N=e^{i2\pi(n-p)r}=\cos 2\pi r(n-p)+i\sin 2\pi r(n-p)=1+0=1$

which results in the below result for $n \neq p$:

$$s_r = \frac{1 - q^N}{1 - q} = 0$$

Discrete Fourier Transform

- For $n=p$, the term in brackets in the equation below becomes 1.

$$x_n = \sum_{p=0}^{N-1} \left\{ \frac{1}{N} \sum_{r=0}^{N-1} x_p \left[e^{i2\pi(n-p)/N} \right]^r \right\}$$

- The summation of all ones r times gives *N*. *N*'s cancel each other in the above equation and letting *p=n* gives:

$$\frac{1}{N} \sum_{p=0}^{N-1} x_p N \delta_{pn} = x_n$$

- Thus, the above equality is proved to be satisfied which consequently proves the inversion theorem. This takes us to the final proven Discrete Fourier Transform formulas in the next slide.

Discrete Fourier Transform

- Discrete Fourier Transform

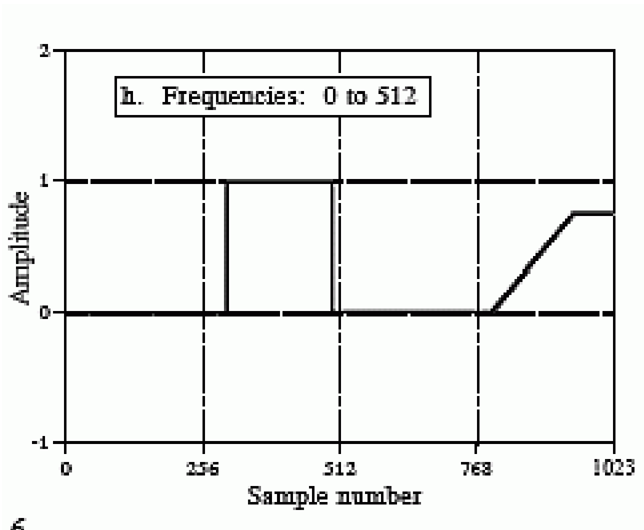
$$X_n = \sum_{r=0}^{N-1} x_r e^{-i2\pi nr/N}$$

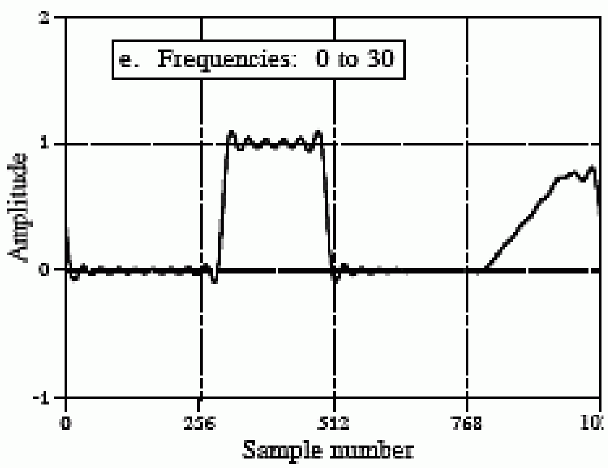
- Inverse Discrete Fourier Transform

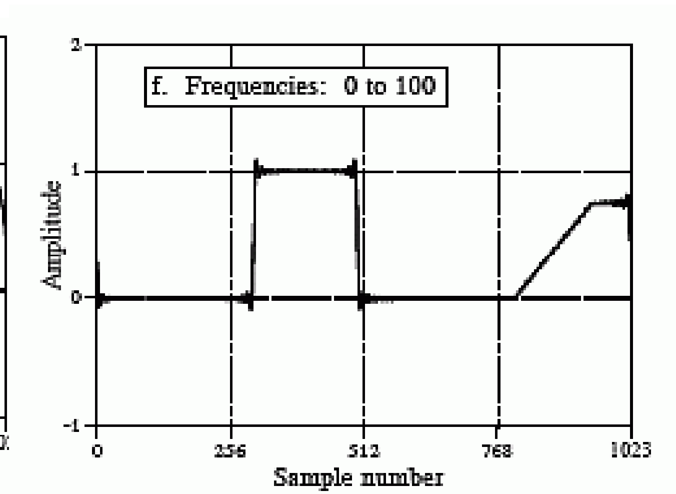
$$x_n = \frac{1}{N}\sum_{r=0}^{N-1} X_r e^{i2\pi nr/N}$$

Gibb’s phenomenon

- The first figure shows a time domain signal being synthesized from sinusoids. The signal being reconstructed is shown in the graph on the left hand side. Since this signal is 1024 points long, there will be 513 individual frequencies needed for a complete reconstruction. The figure on the right shows a reconstructed signal using frequencies 0 through 100. This signal was created by taking the DFT of the signal on the left hand side, setting frequencies 101 through 512 to a value of zero, and then using the Inverse DFT to find the resulting time domain signal. The figure in the middle shows a reconstructed signal using frequencies 0 through 30.

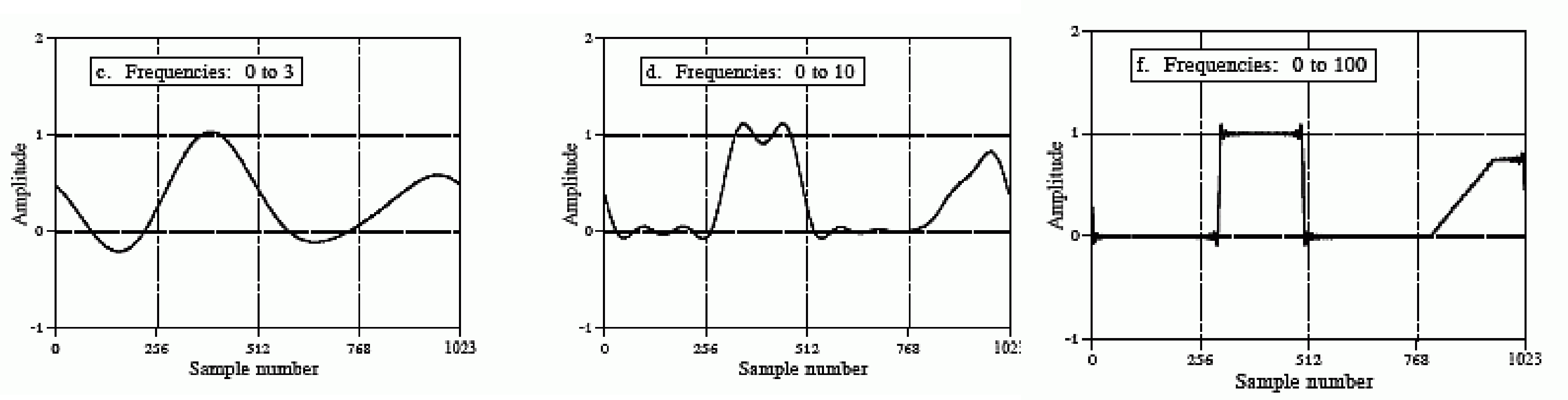






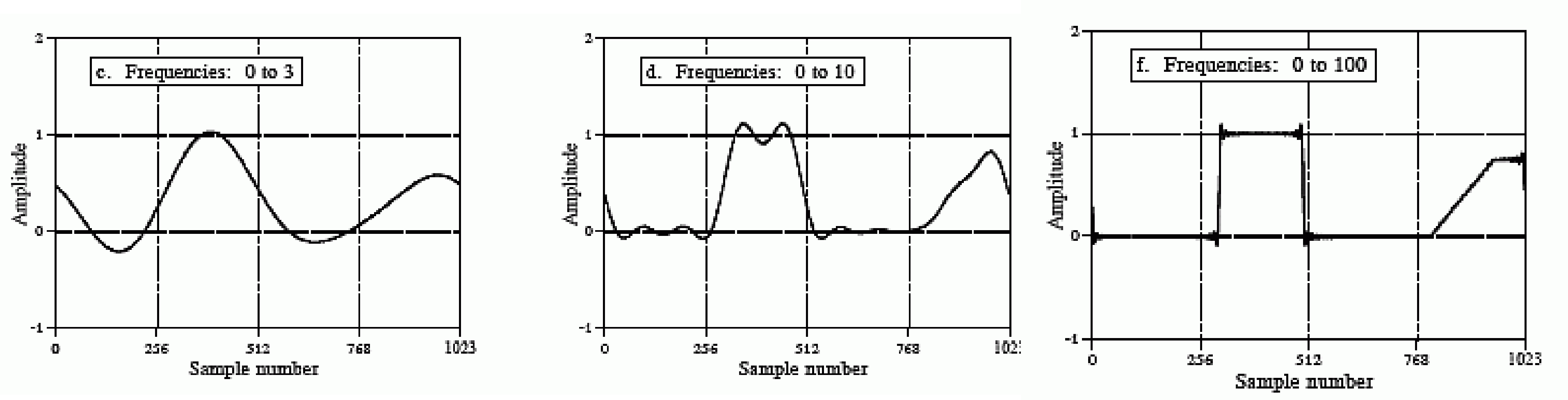
Gibb's phenomenon

- When only some of the frequencies are used in the reconstruction, each edge shows *overshoot* and *ringing* (decaying oscillations). This overshoot and ringing is known as the **Gibbs effect,** after the mathematical physicist Josiah Gibbs, who explained the phenomenon in 1899.
- The critical factor in resolving this puzzle is that the *width* of the overshoot becomes smaller as more sinusoids are included. The overshoot is still present with an infinite number of sinusoids, but it has *zero* width. Exactly at the discontinuity the value of the reconstructed signal converges to the midpoint of the step. As shown by Gibbs, the summation converges to the signal in the sense that the *error* between the two has zero energy.



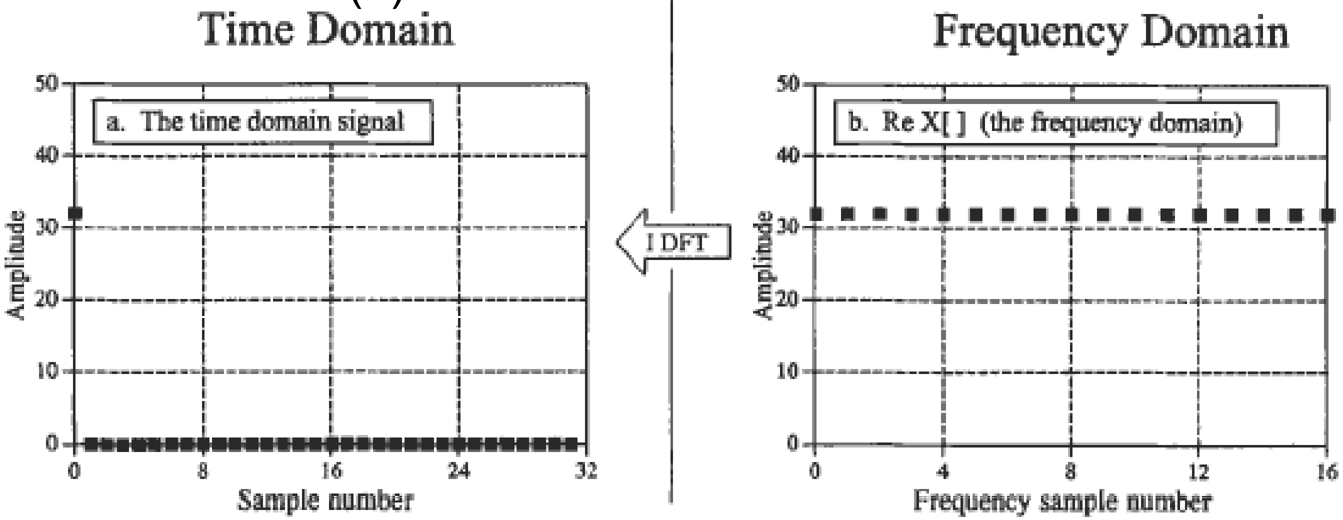
Gibb’s phenomenon

- Problems related to the Gibbs effect are frequently encountered in DSP. For example, a low-pass filter is a *truncation* of the higher frequencies, resulting in overshoot and ringing at the edges in the *time domain*. Another common procedure is to truncate the ends of a time domain signal to prevent them from extending into neighboring periods. By duality, this distorts the edges in the *frequency domain*. These issues will resurface in future chapters on filter design.



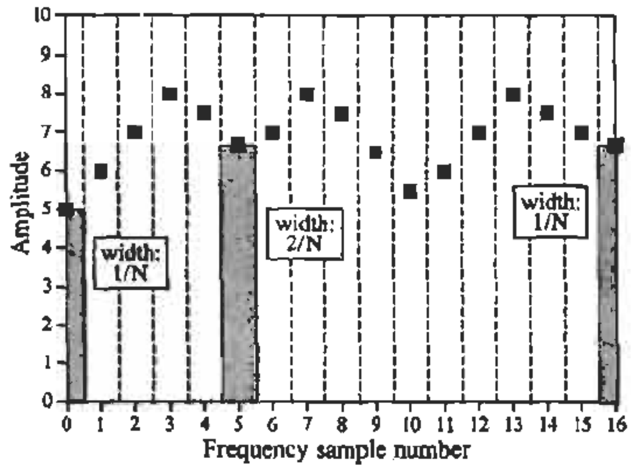
Example: An important DFT pair

- Figure is an example signal we wish to synthesize, an impulse at sample zero with an amplitude of 32. Figure b shows the frequency domain representation of this signal. The real part of the frequency domain is a constant value of 32. The imaginary part (not shown) is composed of all zeros.
- As discussed in the next chapter, this is an important DFT pair: an impulse in the time domain corresponds to a constant value in the frequency domain. For now, the important point is that (b) is the ***DFT*** of (a), and (a) is the ***Inverse DFT*** of (b).



Bandwidth

- As shown in the figure, the bandwidth can be defined by drawing dividing lines between the samples. For instance, sample number *5* occurs in the band between 4.5 and *5.5;* sample number *6* occurs in the band between 5.5 and *6.5,* etc. Expressed as a fraction of the total bandwidth (i.e., N/2), the bandwidth of each sample is *2/N.* An exception to this is the samples on each end, which have one-half of this bandwidth, *1IN.* This accounts for the *2/N* scaling factor between the sinusoidal amplitudes and frequency domain, as well as the additional factor of two needed for the first and last samples.

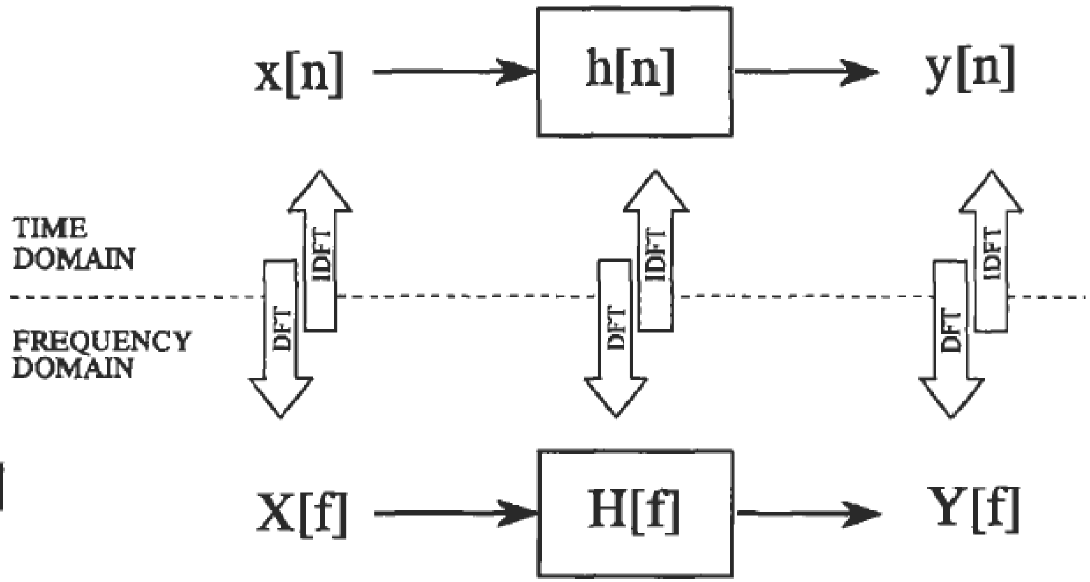


- DFT can be calculated by the fast Fourier transform (FFT), which is an ingenious algorithm that decomposes a DFT with *N* points, into *N* DFTs each with a single point.

Frequency response-Impulse response

- A system’s frequency response is the Fourier transform of its impulse response.
- Keeping with standard DSP notation, impulse responses use lower-case variables, while the corresponding frequency responses are upper case. Since *h*[ ] is the common symbol for the impulse response, *H*[ ] is used for the frequency response. That is, ***convolution*** in the time domain corresponds to ***multiplication*** in the frequency domain.

$$x[n] * h[n] = y[n]$$



$$X[f] \times H[f] = Y[f]$$

How much resolution can you obtain in the frequency response?

- The answer is: ***infinitely*** high, if you are willing to pad the impulse response with an ***infinite*** number of zeros. In other words, there is nothing limiting the frequency resolution except the length of the DFT.

- Even though the impulse response is a ***discrete*** signal, the corresponding frequency response is ***continuous.*** *An N* point DFT of the impulse response provides *N/2* + 1 *samples* of this continuous curve. If you make the DFT longer, the resolution improves, and you obtain a better idea of what the continuous curve looks like.

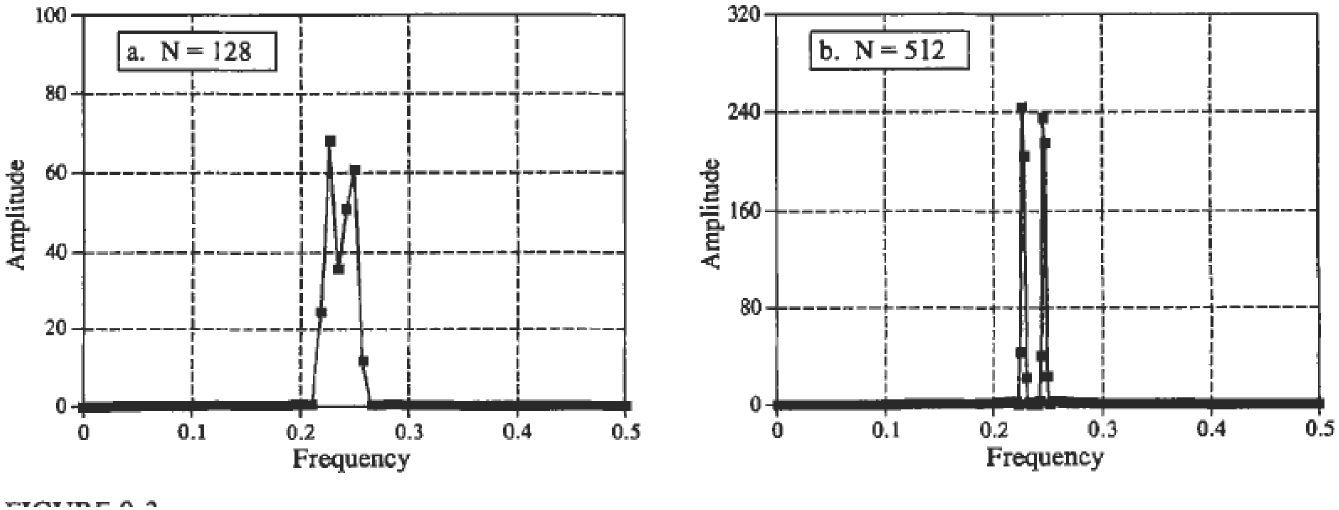
- This can be better understood by the **discrete time Fourier transform (DTFT).** Consider an ***N*** sample signal being run through an ***N*** point DFT, producing an ***N/2*** + 1 sample frequency domain. DFT considers the time domain signal to be ***infinitely long*** and ***periodic.*** That is, the ***N*** points are repeated over and over from negative to positive infinity. Now consider what happens when we start to pad the time domain signal with an ever increasing number of zeros, to obtain a finer and finer sampling in the frequency domain.

How much resolution can you obtain in the frequency response?

- Adding zeros makes the period of the time domain ***longer,*** while simultaneously making the frequency domain samples ***closer together.***
- Now we will take this to the extreme, by adding ***an infinite*** number of zeros to the time domain signal. This produces a different situation in two respects.
- First, the time domain signal now has an infinitely long period. In other words, it has turned into an ***aperiodic*** signal.
- Second, the frequency domain has achieved an infinitesimally small spacing between samples. That is, it has become a ***continuous signal.*** This is the DTFT, the procedure that changes a discrete aperiodic signal in the time domain into a frequency domain that is a continuous curve. In mathematical terms, a system's frequency response is found by taking the DTFT of its impulse response. Since this cannot be done in a computer, the DFT is used to calculate a ***sampling*** of the true frequency response. This is the difference between what you do in a computer (the DFT) and what you do with mathematical equations (the DTFT).

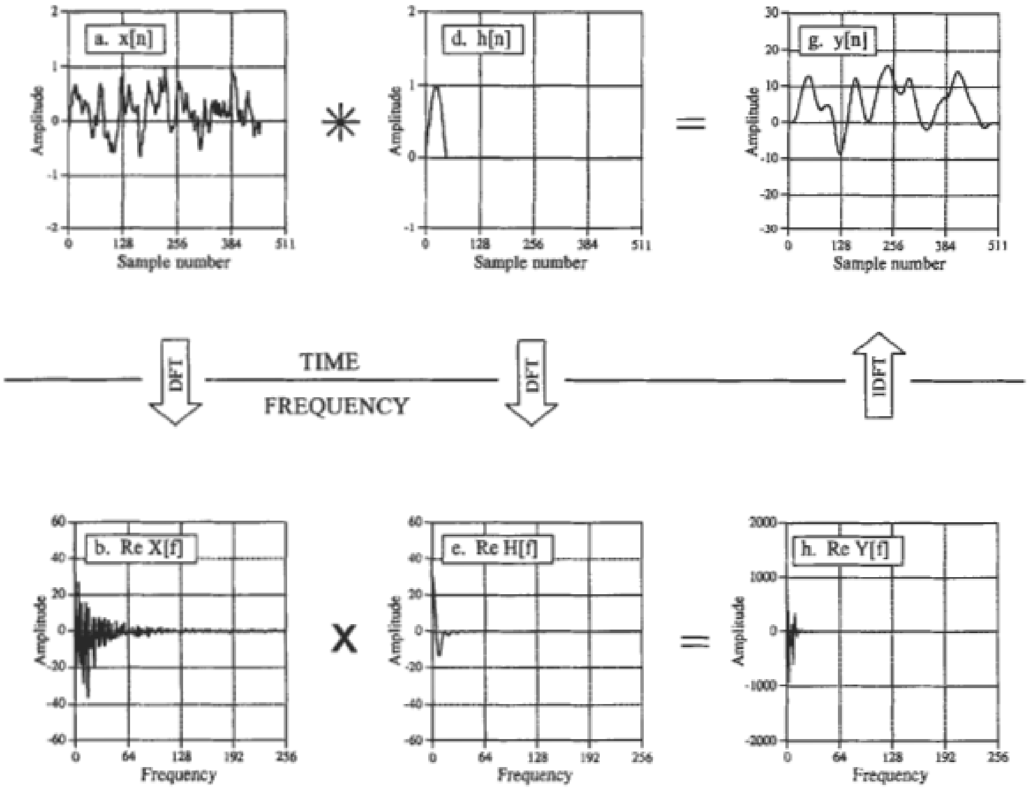
Close peaks

- Suppose there are peaks very close together, such as shown in the figure**.** There are two factors that limit the frequency resolution that can be obtained-that is,how close the peaks can be without merging into a single entity. The first factor is the length of the DFT. The frequency spectrum produced by an ***N*** point DFT consists of ***N/2*** + 1 samples equally spaced between zero and one half of the sampling frequency. To separate two closely spaced frequencies, the sample spacing must be ***smaller*** than the distance between the two peaks. For example, a 512-point DFT is sufficient to separate the peaks in the figure,while a 128-point DFT **is** not.



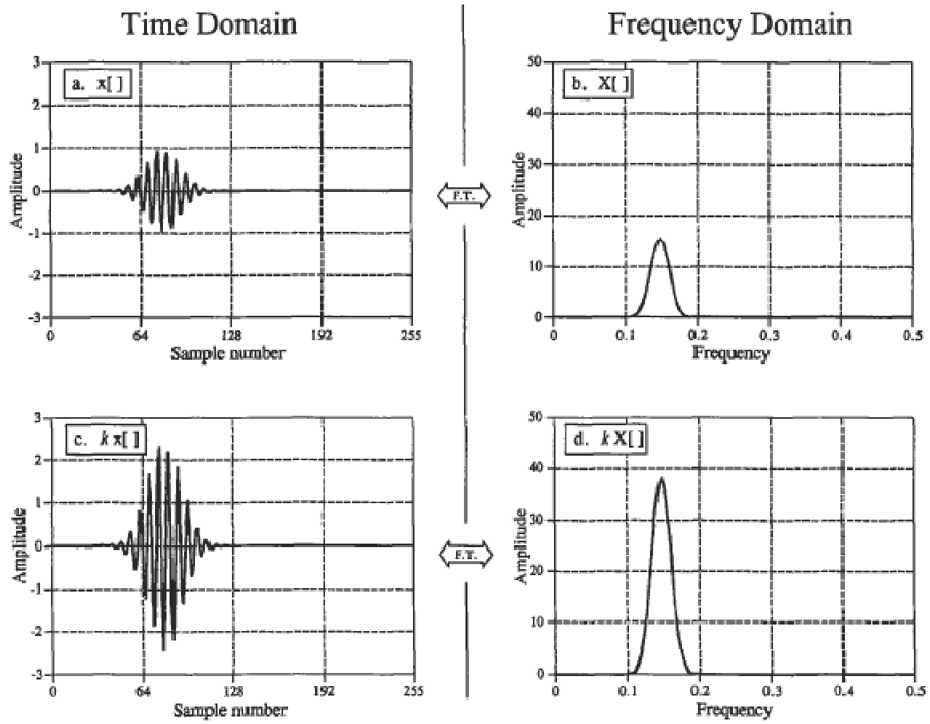
Output of a system

- What are you going to do if given an input signal and impulse response, and need to find the resulting output signal? Transform the two signals into the frequency domain, multiply them, and then transform the result back into the time domain.



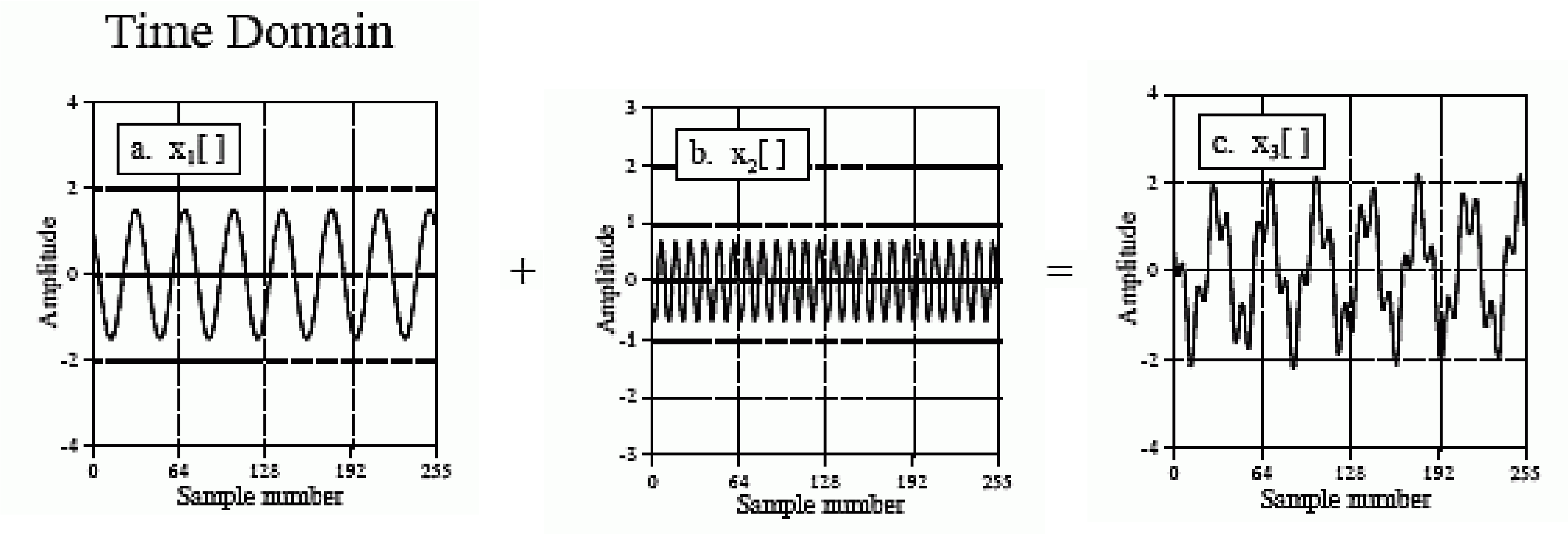
Properties of the Fourier Transform

- ***Homogeneity*** means that a change in amplitude in one domain produces an identical change in amplitude in the other domain. When the amplitude of a time domain waveform is changed, the amplitude of the sine and cosine waves making up that waveform must also change by an equal amount. In mathematical form, if *x*[ ] and *X*[ ] are a Fourier Transform pair, then *kx*[ ] and *kX*[ ] are also a Fourier Transform pair, for any constant *k*.



Properties of the Fourier Transform

- ***Additivity*** of the Fourier transform means that ***addition*** in one domain corresponds to ***addition*** in the other domain. An example of this is shown in the figure. In this illustration, (a) and (b) are signals in the time domain called $x_1[n]$ and $x_2[n]$, respectively. Adding these signals produces a third time domain signal called $x_3[n]$, shown in (c).



Properties of the Fourier Transform

- In spite of being linear, the Fourier transform is *not* shift invariant. In other words, a shift in the time domain *does not* correspond to a shift in the frequency domain.
- Let *h(t)* be the impulse response, i.e., the system's response to a Dirac impulse, it can be proved that the response *g(t)* of the system to an input *f(t)* is the convolution of *f(t)* and *h(t):*

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

- The convolution theorem in Fourier analysis states that *convolution* in one domain corresponds to *multiplication* in the other domain.
- Hence the frequency response function *H(f)* is the ratio between the response and the input as a function of the frequency.

$$g(t) = h(t) * f(t)$$

$$G(f) = H(f)F(f)$$

Properties of the Fourier Transform

- Differentiation in the time domain is equivalent to multiplication by a factor $i\omega$ in the frequency domain.

$$F\left[\dot{x}(t)\right] = (i\omega)X(\omega)$$

- Integration in the time domain is equivalent to division by $i\omega$ in the frequency domain.

$$F\left[\int_{-\infty}^{t} x(\tau)d\tau\right] = \frac{X(\omega)}{i\omega}$$

- If the input is the harmonic probe $e^{i\omega t}$, the output is $e^{i\omega t}$ multiplied by the FRF evaluated at ω.

$$y(t) = H(\omega)e^{i\omega t}$$

Fourier transform

- Measured signals are time domain functions. It is important to investigate the signals in the frequency domain in order to study their frequency content. The Fourier tansform is a tool to transform signals from the time domain to the frequency domain.

$$\mathcal{F}[x(t)] = X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt$$

- The signal can be transformed back to the time domain using the inverse Fourier transform:

$$\mathcal{F}^{-1}[X(\omega)] = x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega$$

Fast Fourier Transform

- The FFT is a clever algorithm for rapidly calculating the DFT.

- The N point DFT of an N point sequence x_n is given by:

$$X_k = \sum_{r=0}^{N-1} x_n W_N^{nk} \quad \text{where} \quad W_N = e^{-j2\pi/N}$$

- Because x_n may be either real or complex, evaluating X_k requires on the order of N complex multiplications and N complex additions for each value of k. Therefore, because there are N values of X_k, computing an N point DFT reqires N^2 complex multiplications and additions.

- The basic strategy that is used in the FFT algorithm is one of **divide and conquer**, which involves decomposing an N point DFT into succesively smaller DFTs.

Fast Fourier Transform

- Suppose that the length of x_n is even, (i.e., N is divisible by 2). If x_n is decimated into two sequences of length N/2, computing the $N/2$ point DFT of each of these sequences requires approximately $(N/2)^2$ multiplications and the same number of additions.Thus, the two DFTs require $2*(N/2)^2=N^2/2$ multiplications and the same number of additions.

- Therefore, if it is possible to find the N point DFT of x_n from these two $N/2$ point DFTs, great computational time will be saved because $N^2/2$ operations will be required instead of N^2.

- Let x_n be a sequence of length $N=2^v$, and suppose that x_n is split (decimated) into two subsequences, each of length $N/2$.

- The first sequence g_n is formed from the even-index terms,

$$g_n = x_{2n} \qquad n = 0,1,\ldots,\frac{N}{2}-1$$

Fast Fourier Transform

- The second sequence h_n is formed from the odd-index terms,

$$h_n = x_{2n+1} \qquad n = 0,1,\ldots,\frac{N}{2}-1$$

- In terms of these sequences, the N-point DFT of x_n is

$$X_k = \sum_{n=0}^{N-1} x_n W_N^{nk} = \sum_{n \text{ even}} x_n W_N^{nk} + \sum_{n \text{ odd}} x_n W_N^{nk}$$

$$= \sum_{l=0}^{\frac{N}{2}-1} g_l W_N^{2lk} + \sum_{l=0}^{\frac{N}{2}-1} h_l W_N^{(2l+1)k}$$

Fast Fourier Transform

- Since

$$W_N = e^{-j2\pi/N} \longrightarrow W_N^{2lk} = W_{N/2}^{lk}$$

- Substituting the above result into the equation:

$$X_k = \sum_{n=0}^{N-1} x_n W_N^{nk} = \sum_{n \text{ even}} x_n W_N^{nk} + \sum_{n \text{ odd}} x_n W_N^{nk}$$

$$= \sum_{l=0}^{\frac{N}{2}-1} g_l W_N^{2lk} + \sum_{l=0}^{\frac{N}{2}-1} h_l W_N^{(2l+1)k}$$

- We obtain:

$$X_k = \sum_{l=0}^{\frac{N}{2}-1} g_l W_{N/2}^{lk} + W_N^k \sum_{l=0}^{\frac{N}{2}-1} h_l W_{N/2}^{lk}$$

Fast Fourier Transform

- Note that in the equation

$$X_k = \sum_{l=0}^{\frac{N}{2}-1} g_l W_{N/2}^{lk} + W_N^k \sum_{l=0}^{\frac{N}{2}-1} h_l W_{N/2}^{lk}$$

the first term is the *N*/2 point DFT of g_n and the second is the *N*/2 point DFT of h_n:

$$X_k = G_k + W_N^k H_k \qquad k = 0,1,\ldots,N-1$$

- Although the N/2 point DFTs of g_n and h_n are sequences of length *N*/2, the periodicity of the complex exponentials allows us to write:

$$G_k = G_{k+N/2}$$
$$H_k = H_{k+N/2}$$

Fast Fourier Transform

- Therefore, X_k may be computed from the $N/2$ point DFTs G_k and H_k.

- If $N/2$ is even, g_n and h_n may again be decimated. For example, G_k may be evaluated as follows:

$$G_k = \sum_{n=0}^{\frac{N}{2}-1} g_n W_{N/2}^{nk} = \sum_{n \text{ even}}^{\frac{N}{2}-1} g_n W_{N/2}^{nk} + \sum_{n \text{ odd}}^{\frac{N}{2}-1} g_n W_{N/2}^{nk}$$

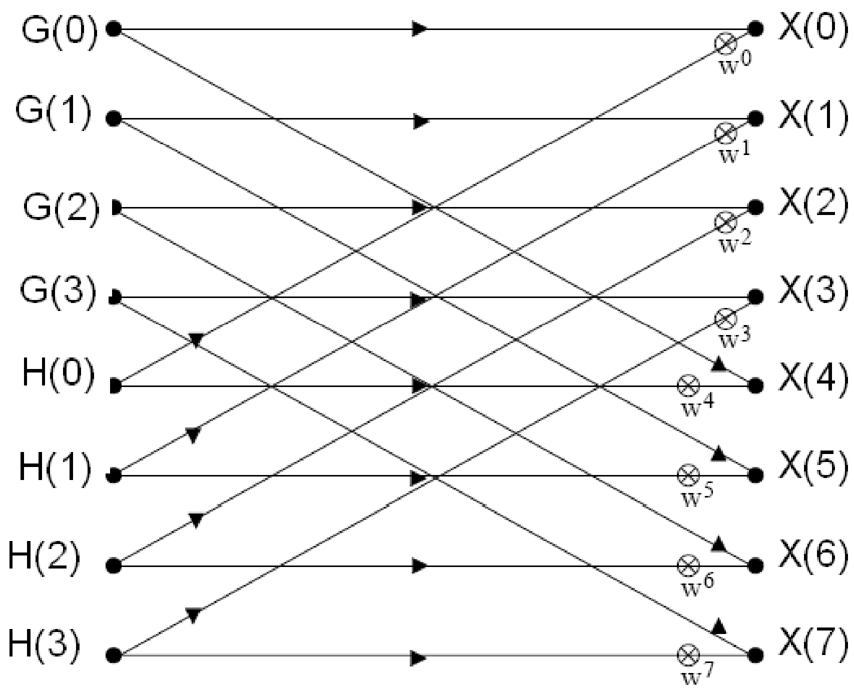
- As before, this leads to

$$G_k = \sum_{n=0}^{\frac{N}{4}-1} g_{2n} W_{N/4}^{nk} + W_{N/2}^{k} \sum_{n=0}^{\frac{N}{4}-1} g_{2n+1} W_{N/4}^{nk}$$

where the first term is the N/4 point DFT of the even samples of g_n, and the second is the N/4 point DFT of the odd samples.

Fast Fourier Transform

- If N is a power of 2, the decimation may be continued until there are only two-point DFTs. A block diagram showing the computations that are necessary for the first stage of an eight point decimation in time is shown in the figure. This diagram is called the **FFT butterfly**.



Fast Fourier Transform

- Computing an N point DFT using FFT is much more effective than calculating the DFT directly.

- For example, if $N=2^v$, there are $\log_2 N=v$ stages of computation. At each of these stages, we are required to carry out N multiplications. The total cost of the algorithm is then $N\log_2 N$.

- The saving in moving from the DFT to FFT is:

$$\frac{N^2}{N\log_2 N}=\frac{N}{\log_2 N}=\frac{N}{v}$$

- Suppose N=1024, we get a saving of computational effort of the order 100:1, and this saving increases with N.

Problems associated with analog-to-digital conversion

- *Sampling*
- *Aliasing*

Problems associated with ADC

- We will now explore the issues of aliasing and apparent signal distortion associated with choosing a sample rate for digital conversion of analog signals. This process is known as **analog to digital conversion (A/D).**
- The process of **sampling** reduces an infinite set to a finite set of data resulting in a loss of information which is responsible for the issues detailed in this section.
- Since the original analog signal contains infinite information (knowledge of the signal at any point in time), the frequencies within the signal are also known.
- Information lost in the sampling process is also lost in the frequency representation. We will explain how this leads to the problem of **leakage** where high frequency components can not be distinguished from low frequency components.

Sampling

- Nearly all data acquisition systems sample data with uniform time intervals. For evenly sampled data, time can be expressed as:

$$T = (N-1)\Delta t$$

where N is the sampling index which is the number of equally spaced samples. For most Fourier analyzers N is restricted to a power of 2.

- The sample rate or the sampling frequency is:

$$f_s = \frac{1}{\Delta t} = (N-1)\Delta f$$

Sampling

- Sampling frequency is the reciprocal of the time elapsed Δt from one sample to the next.

- The unit of the sampling frequency is cycles per second or Hertz (Hz), if the sampling period is in seconds.

- The sampling theorem asserts that the uniformly spaced discrete samples are a complete representation of the signal if the bandwidth f_{max} is less than half the sampling rate. The sufficient condition for exact reconstructability from samples at a uniform sampling rate f_s (in samples per unit time) ($f_s \geq 2 f_{max}$).

Aliasing

- One problem encountered in A/D conversion is that a high frequency signal can be falsely confused as a low frequency signal when sufficient precautions have been avoided.
- This happens when the sample rate is not fast enough for the signal and one speaks of **aliasing**.
- Unfortunately, this problem can not always be resolved by just sampling faster, the signal's frequency content must also be limited.
- Furthermore, the costs involved with postprocessing and data analysis increase with the quantity of data obtained. Data acquisition systems have finite memory, speed and data storage capabilities. Highly oversampling a signal can necessitate shorter sample lengths, longer time on test, more storage medium and increased database management and archiving requirements.

Aliasing

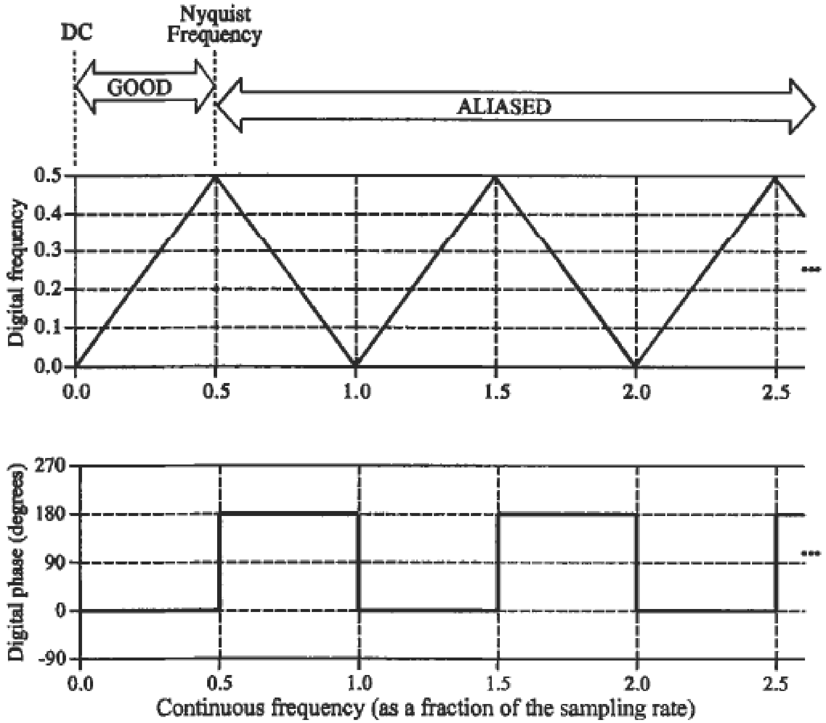
- The central concept to avoid aliasing is that the sample rate must be at least twice the highest frequency component of the signal ($f_s \geq 2f_{max}$). We define the Nyquist or cut-off frequency

$$f_N = \frac{f_s}{2} = \frac{1}{2\Delta t}$$

- The concept behind the cut-off frequency is often referred to as Shannon's sampling criterion. Signal components with frequency content above the cut-off frequency are aliased and can not be distinguished from the frequency components below the cut-off frequency.

Aliasing

- Conversion of analog frequency into digital frequency during sampling is shown in the figure. Continuous signals with a frequency less **than** one-half of the sampling rate are directly converted into the corresponding digital frequency. Above one-half of the sampling rate, aliasing takes place, resulting in the frequency being misrepresented in the digital data. Aliasing always changes a higher frequency into a lower frequency between 0 and 0.5. In addition, aliasing may also change the phase of the signal by 180 degrees.



Aliasing

- What happens if the original signal actually has a component above the Nyquist frequency?

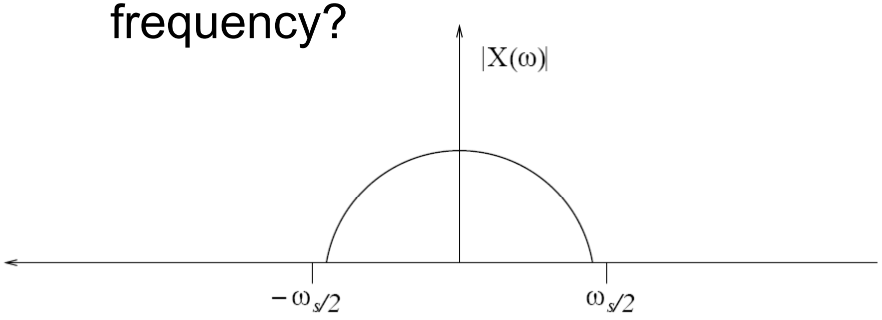


Figure 3. Spectrum of continuous signal $x(t)$.

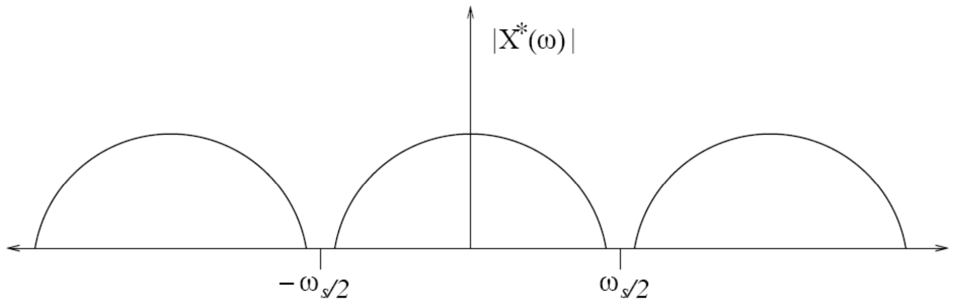


Figure 4. Spectrum of sampled signal $x_s(t)$.

- Now if the spectrum of the continuous signal extends beyond the Nyquist frequency we see overlap

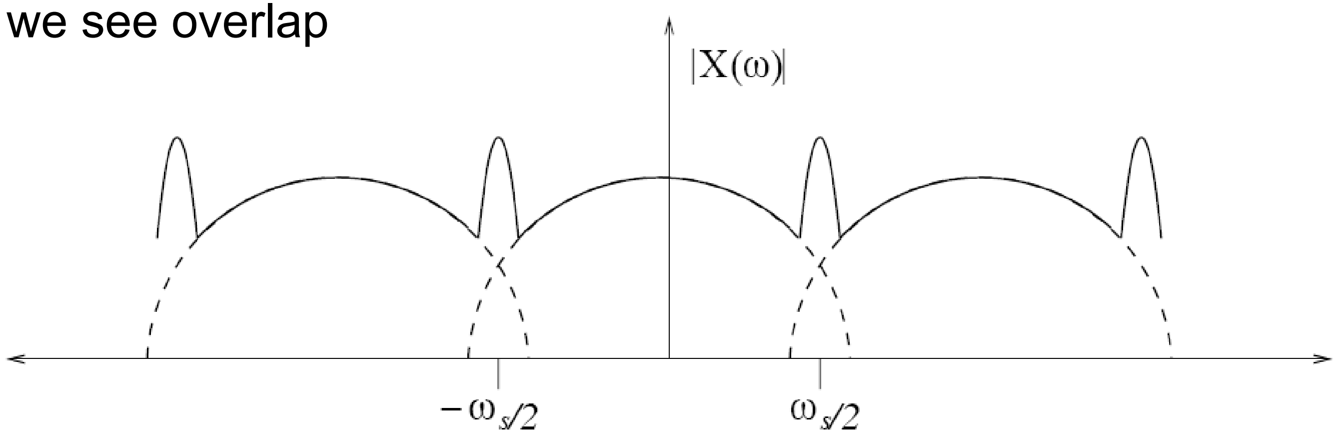


Figure 5. Spectrum of sampled signal $x_s(t)$ with overlap - aliasing.

Aliasing

- If any energy in the original signal extends beyond the Nyquist frequency, it is folded back into the Nyquist interval in the spectrum of the sampled signal. This folding is called **aliasing**

fs≥2fmax

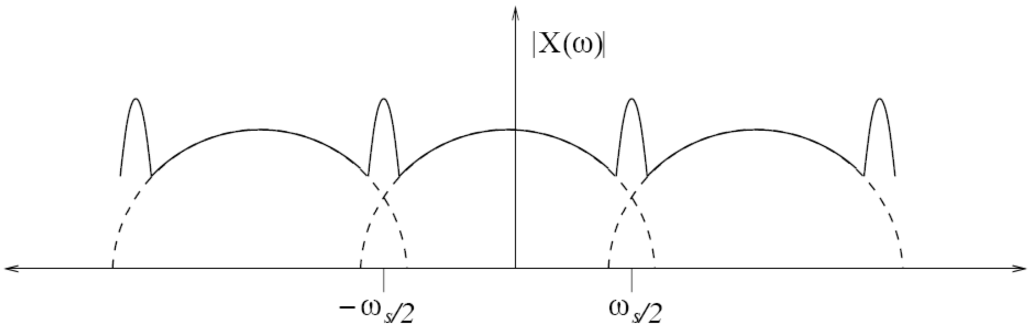
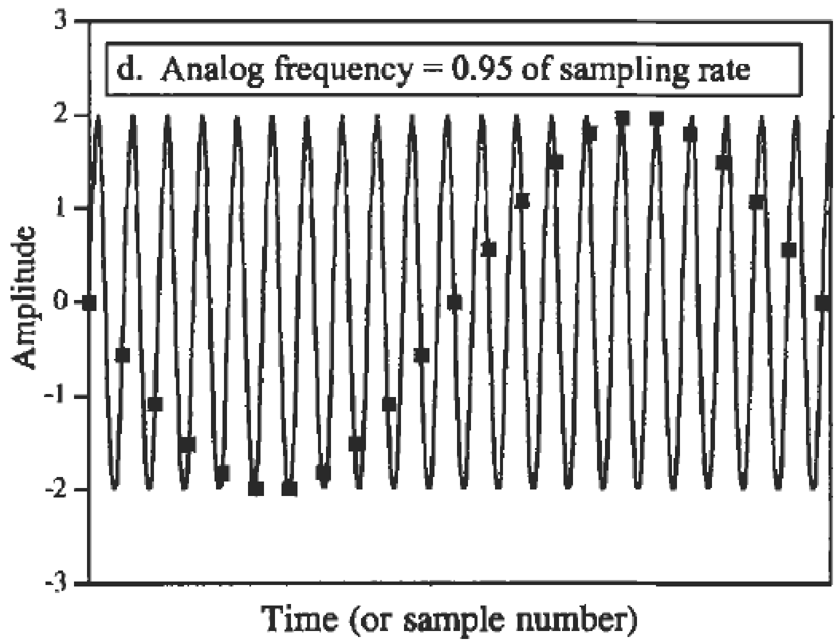


Figure 5. Spectrum of sampled signal $x_s(t)$ with overlap - aliasing.

Aliasing

- Just as aliasing can change the frequency during sampling, it can also change the **phase.** For example, the aliased digital signal in the figure is **inverted** from the original analog signal; one is a sine wave while the other is a negative sine wave. In other words, aliasing has changed the frequency and introduced a 180" phase shift. Only two phase shifts are possible: 0" (no phase shift) and 180" (inversion).



$f_s \geq 2f_{max}$

Leakage

- When converting a signal from the time domain to the frequency domain, the Fast Fourier Transform (FFT) is used.
- The Fourier Transform is defined by the equation:

$$X(f) = F\{x(t)\} = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

which requires a signal sample from –∞ to ∞.

- The Fast Fourier Transform however only requires a finite number of samples (which must be a value of 2^n where n is an integer. i.e. 2, 4, 8, 16 … 512, 1024). The FFT is defined as:

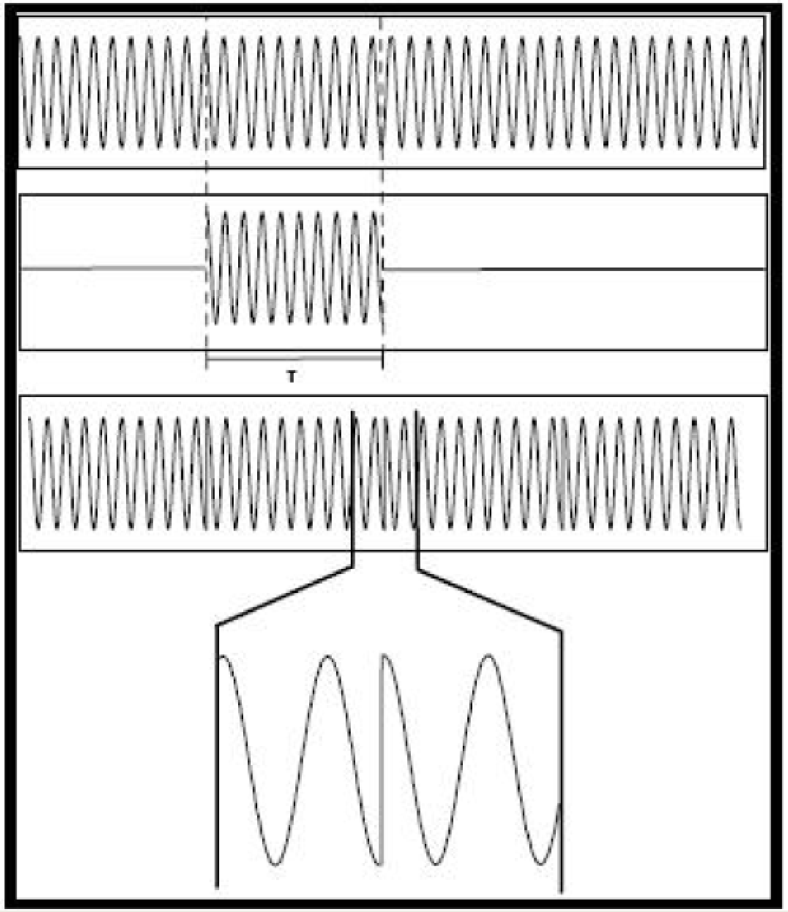
$$X_k = \sum_{i=0}^{n-1} x_i e^{-j2\pi ik/n} \qquad \text{for } k = 0,1,2,\dots,n-1$$

Leakage

- The Fast Fourier Transform is commonly used because it requires much less processing power than the Fourier Transform. Like all shortcuts, there are some compromises involved in the FFT.
- The signal must be **periodic** in the sample window or **leakage** will occur.
- The signal must start and end at the same point in its cycle.
- **Leakage** is the smearing of energy from the true frequency of the signal into adjacent frequencies.
- Leakage also causes the amplitude representation of the signal to be less than the true amplitude of the signal.

Leakage

- An example of a nonperiodic signal can be seen in the Figure.



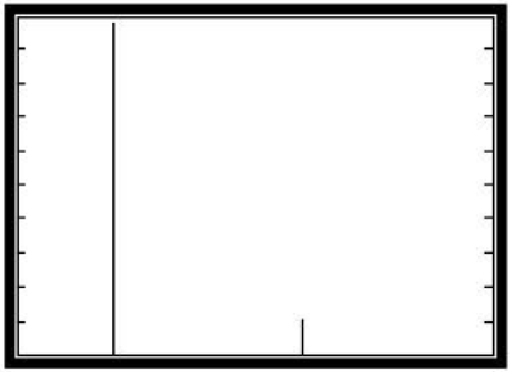


Figure 2: FFT of periodic Signal

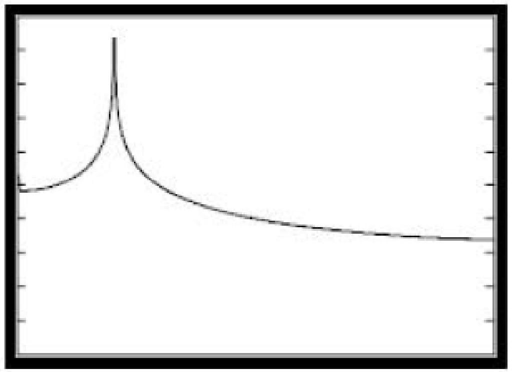


Figure 3: FFT of Non-Periodic Signal.

Leakage

- By comparing the Figures, it can be seen that the frequency content of the signal is smeared into adjacent frequencies when the signal is not periodic.

- In addition to smearing, the amplitude representation of the signal is less than the true value.

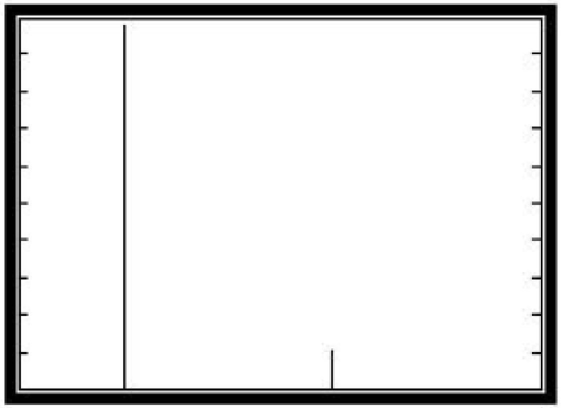


Figure 2: FFT of periodic Signal

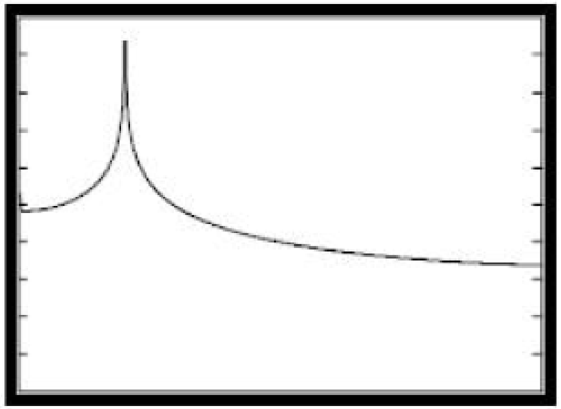


Figure 3: FFT of Non-Periodic Signal.

Leakage

- The only solution to the **leakage problem**, is to make sure that the signal is **periodic** or **completely observed** within the observation window.

- Generally, this is very difficult to achieve. For systems with a perfect linear behaviour, it can be accomplished by exciting the structure with a **periodic signal**. Excitation signals as **burst random** also minimize this problem.

- **Decreasing the frequency step Δf** increases the observation time T and hence will **improve the periodicity** of the signal.

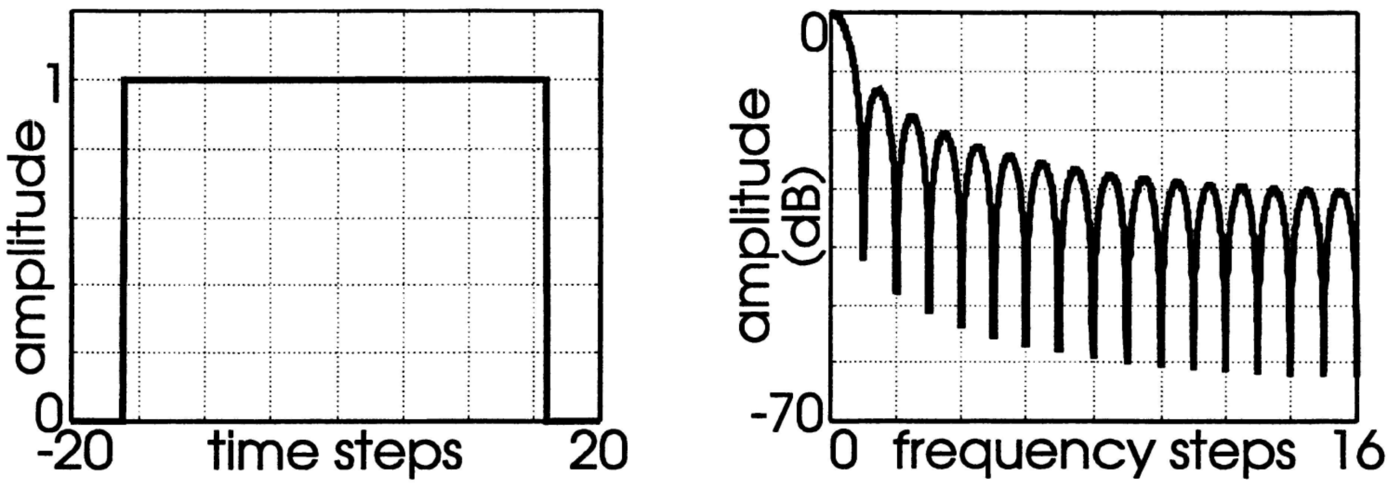
- The use of a **time window other than a rectangular one** offers an approximate solution to the leakage problem.

Windowing

- In signal processing, a **window function** is a function that is zero-valued outside of some chosen interval.
- Applications of window functions include **spectral analysis** and **filter design**.
- The first type of window is called the “**rectangular**” window; it does not weight the signal in any way and is equivalent to saying that no window was used.
- **Rectangular window** is used whenever frequency resolution is of high importance. This window can have up to 36% amplitude error if the signal is not periodic in the sample interval. It is good for signals that inherently satisfy the **periodicity requirement** of the FFT process.

Windowing

- A **rectangular window** is a function that is constant inside the interval and zero elsewhere, which describes the shape of its graphical representation. When another function or a signal (data) is multiplied by a window function, the product is also zero-valued outside the interval: all that is left is the "view" through the window. It can be shown that there is no window with a narrower main lobe than the rectangular window.



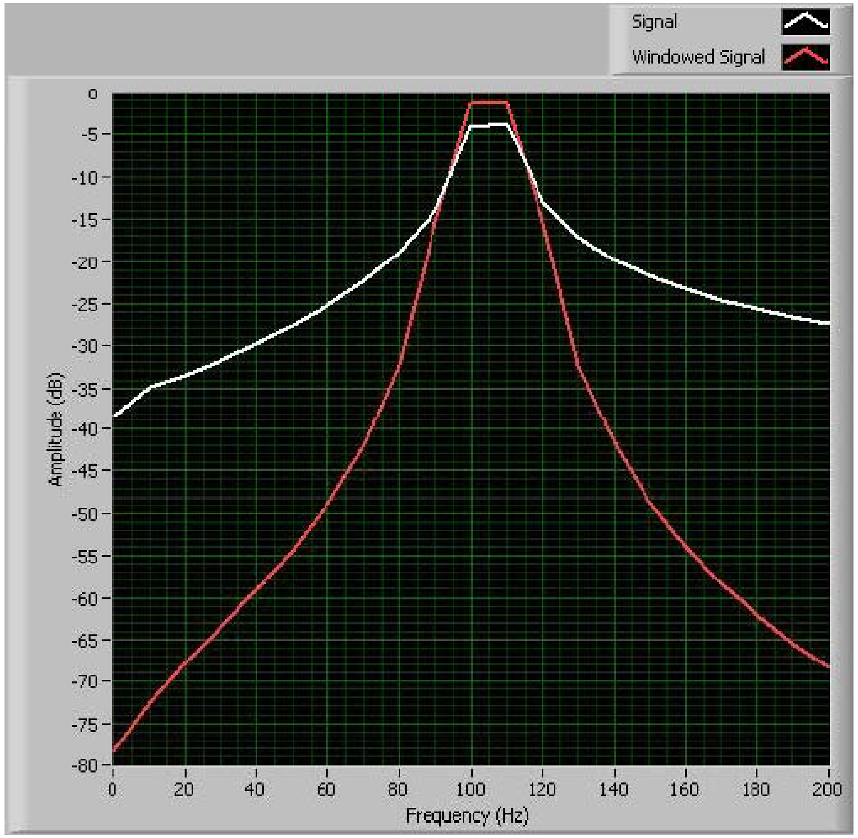
w=1

Windowing

- Windows work by weighting the start and end of a sample to zero while at the same time increasing the amplitude of the signal at the center as to maintain the average amplitude of the signal.

- The effect of a **Hanning window** on a non-periodic signal in the Frequency Domain can be seen in the Figure.

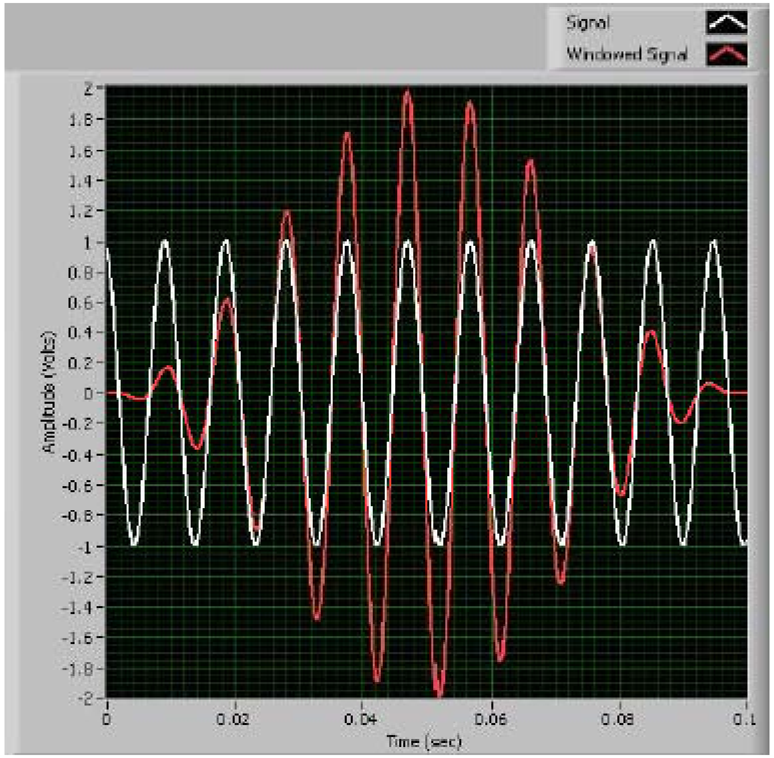
- Figure shows that the **window reduces smearing** and better preserves the amplitude of the signal.



Windowing

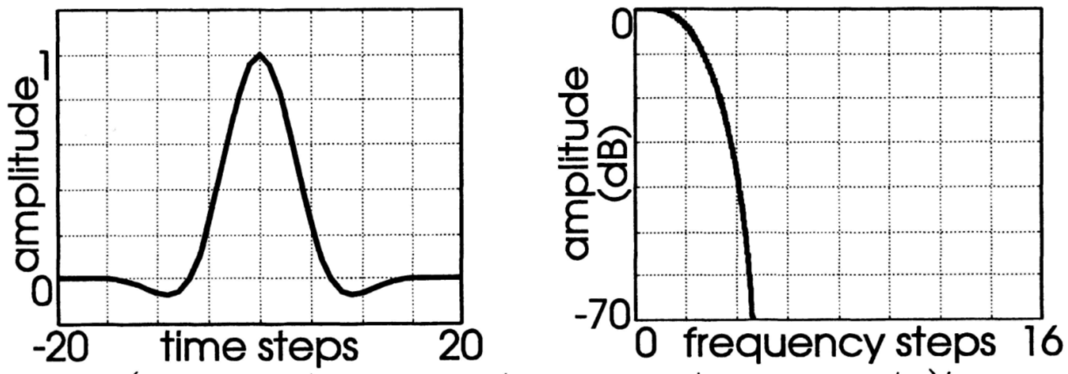
- The effect of the same **Hanning Window** on the time domain signal can be seen in the Figure.

- Figure shows how the **Hanning window** weights the beginning and end of the sample to zero so that it is more **periodic** during the FFT process.



Windowing

- **The Flat Top window** is used whenever signal amplitude is of very high importance. The **flat top window preserves the amplitude** of a signal very well; however it has **poor frequency resolution** so that the exact frequency content may be hard to determine, this is particularly an issue if several different frequency signals exist in close proximity to each other. The flat top window will have at most 0.1% amplitude error.

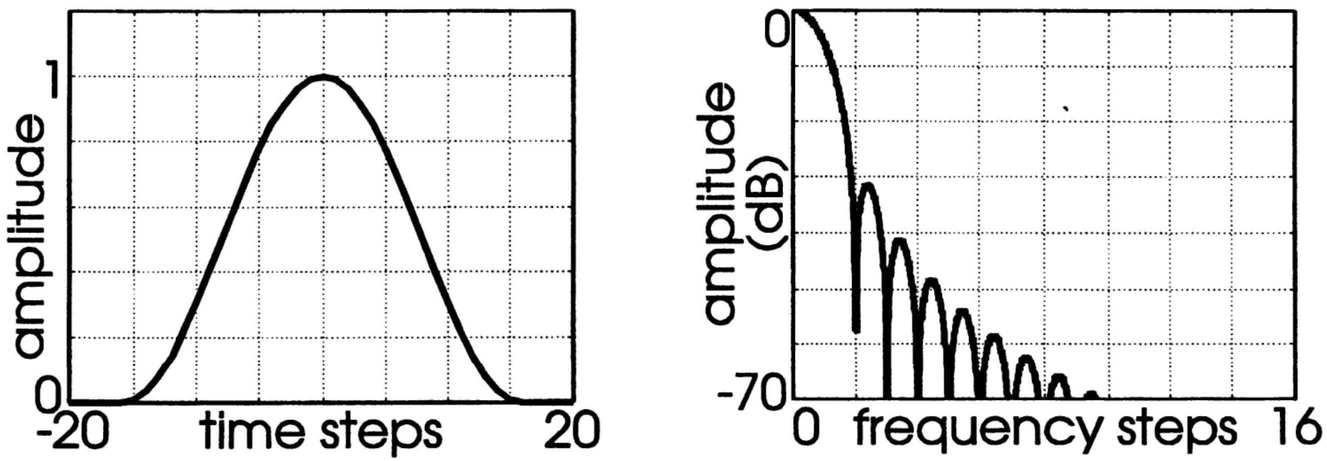


$$w = \left(1 + 1.93\cos(2\pi t/T) + 1.29\cos(4\pi t/T) + .388\cos(6\pi t/T) + .028\cos(8\pi t/T)\right)/4.636$$

Fig.a.2.11e: Flat-top window and transform

Windowing

- **The Hanning window** is a compromise between the Flat Top and Rectangular windows. It helps to maintain the amplitude of a signal while at the same time maintaining frequency resolution. This window can have up to a 16% amplitude error if the signal is not periodic.

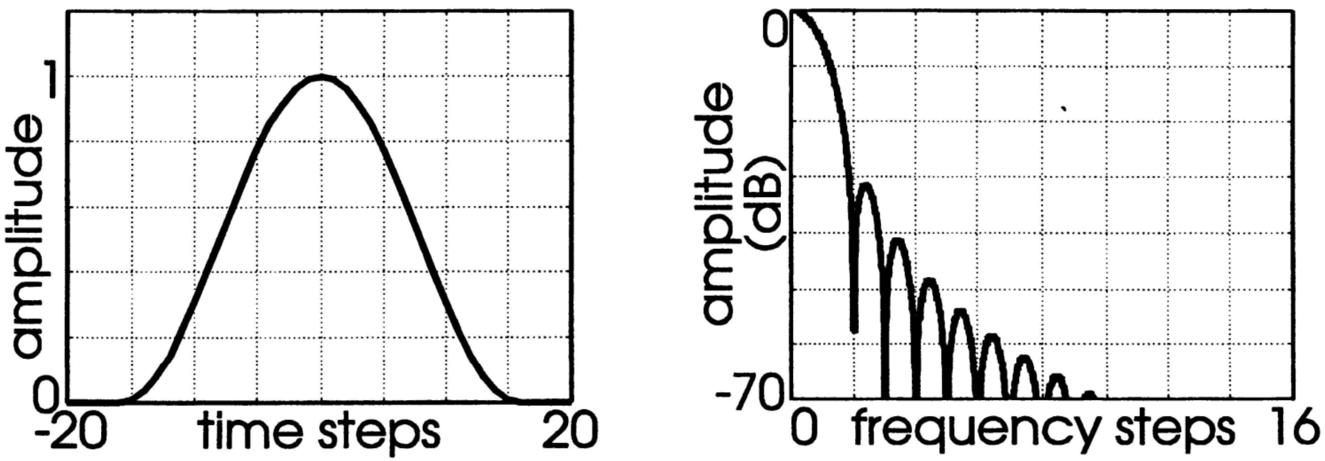


$w=0.5+0.5\cos(2\pi t/T)$

Fig.a.2.11b: Hanning window and transform

Windowing

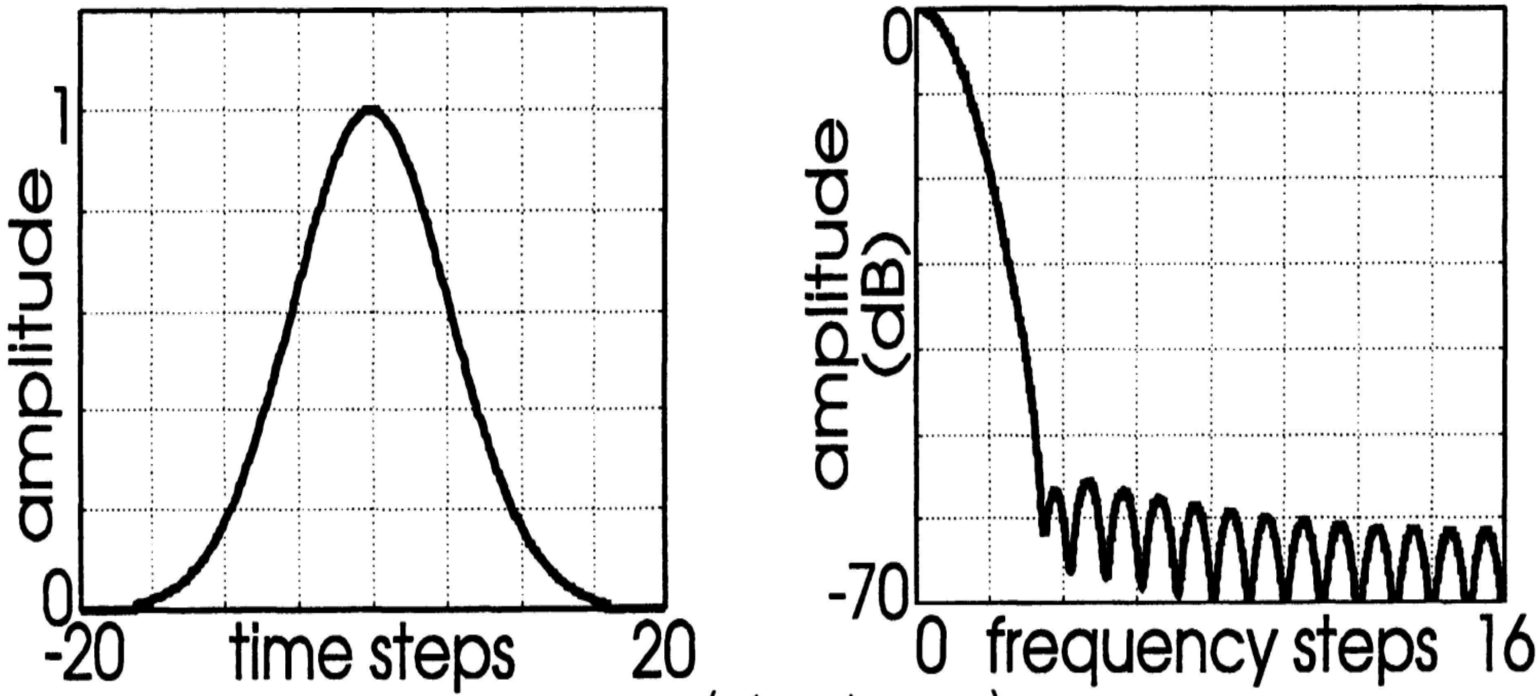
- **The most common window used for random excitations exerted by shakers is the Hanning window.**



w=0.5+0.5 cos(2πt/T)

Fig.a.2.11b: Hanning window and transform

Windowing



$$w = \exp\left(-\left(\beta^2/2\right)\left(2t/T\right)^2\right)$$

Fig.a.2.11d: Gaussian window (β=3.0) and transform

Windowing

- The **exponential window** is used to make a measurement from a vibrating structure more accurate.
- It is used when the “ringing” of a structure does not attenuate adequately during the sample interval.
- An example of the **exponential window** can be seen in the figure.

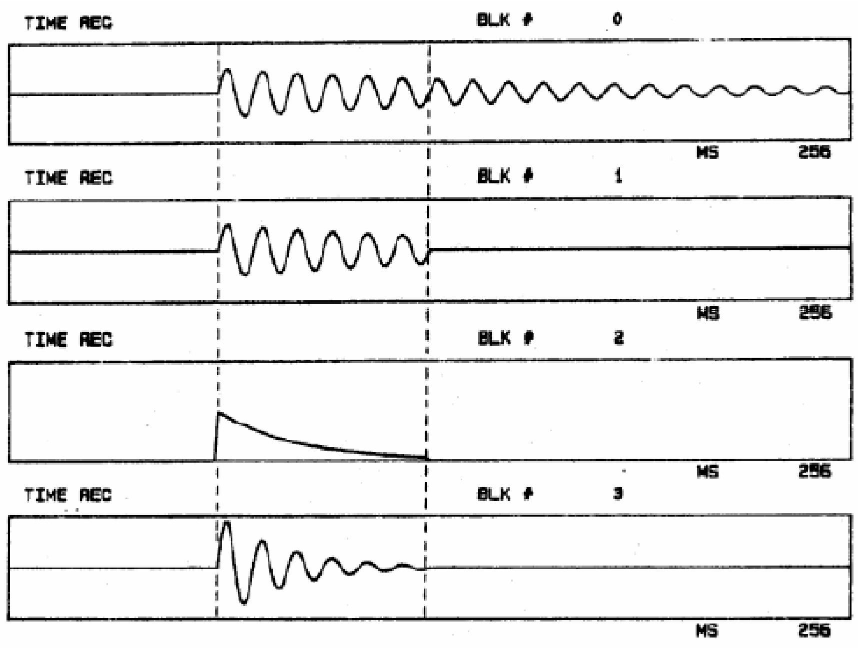


Figure 7: Exponential Window.

Windowing

- The **exponential window** can cause some problems if not used properly.

- As an example, a very simple lightly damped structure was subjected to an impact test. The signal processing parameters were selected for a 400 Hz bandwidth which resulted in a 1 sec time window.

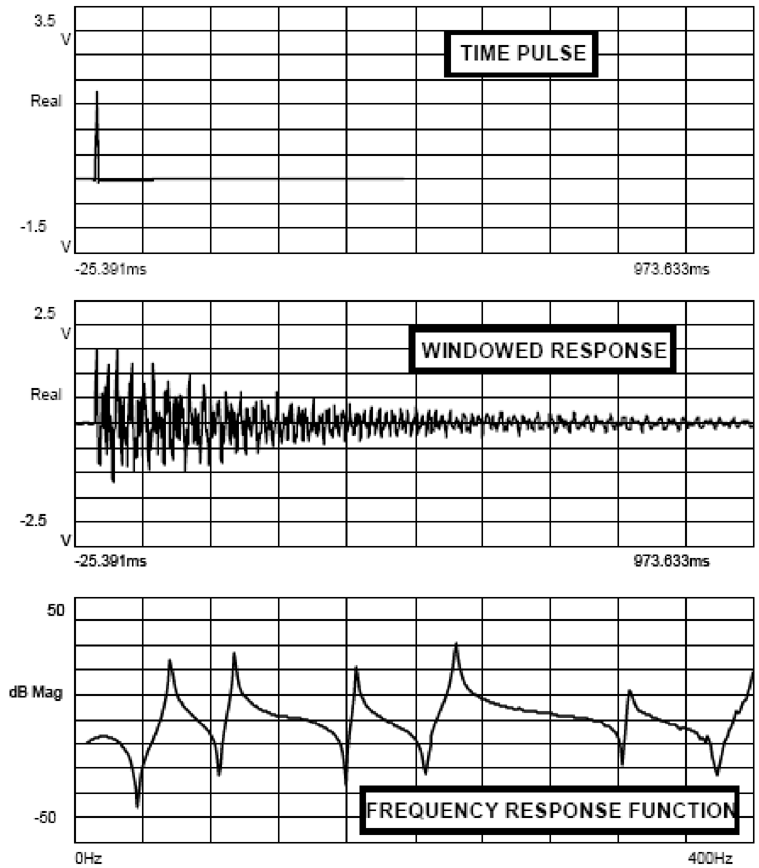


Figure 1 - FRF with slightly too much damping

Exponential Window

- On the right more damping is applied and the peaks are much wider now!

- If an excessive amount of damping is needed to minimize the effects of leakage, then you run the risk of missing **closely spaced modes**.

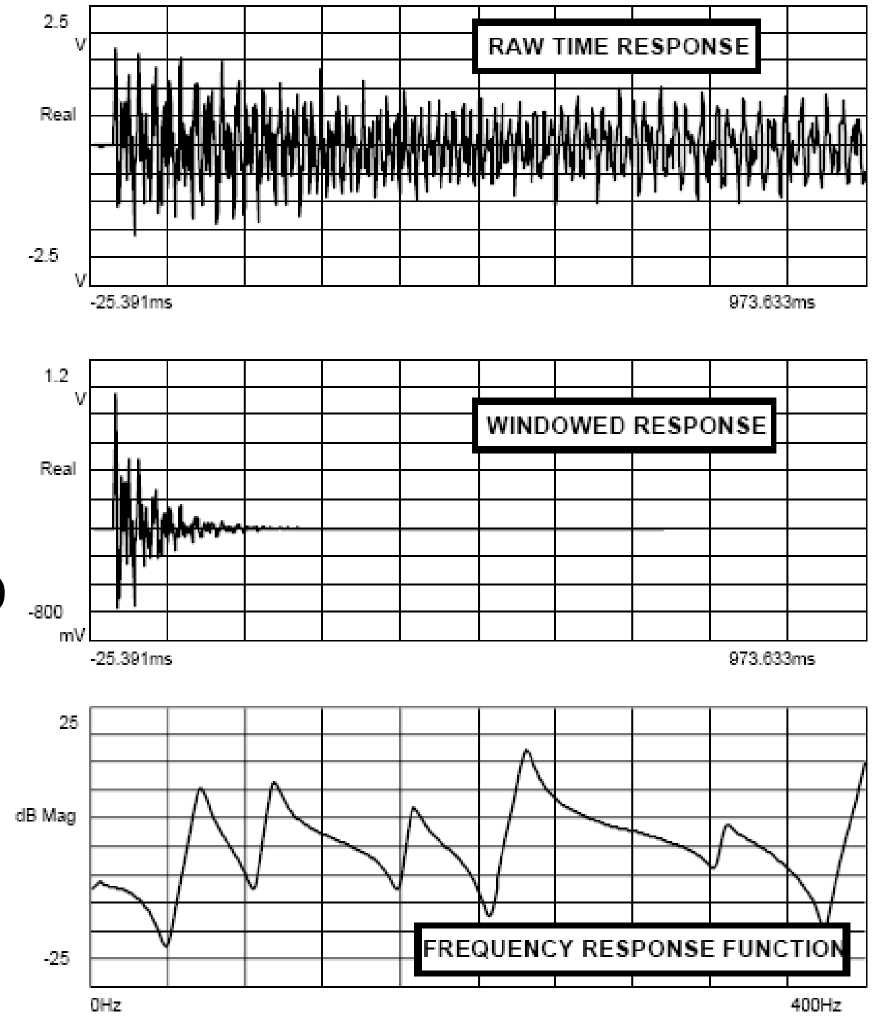


Figure 2 - FRF with too much damping

Exponential Window

Before any window is applied, it is advisable to try alternative approaches to minimize the leakage in the measurement such as:

- Increasing the number of spectral lines
- Halving the bandwidth

which both result in increased total time for measurement.

$$f_s = \frac{1}{\Delta t} = (N-1)\Delta f$$

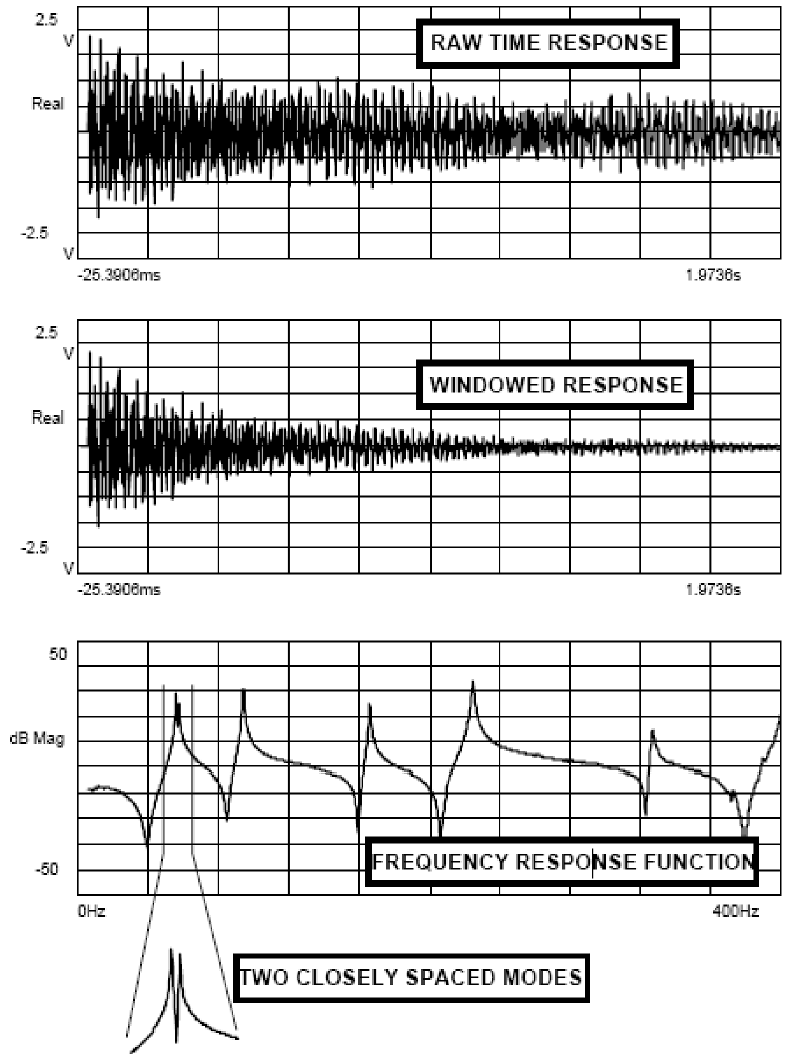


Figure 3 - FRF with increased time/spectral resolution

Windowing

- Impact testing always causes some type of transient response that is the summation of damped exponential sine waves.
- If the entire transient event can be captured such that the FFT requirements can be met, leakage will not be a problem.
- But for lightly damped structures, in many impact testing situations, the use of an **exponential window** is necessary.
- However, the use of exponential window can cause some difficulties when evaluating structures with light damping and closely spaced modes.
- The use of windows may also hide or distort the modes in the measurement.

Windowing

- The effects of leakage can only be minimized through the use of a window. It can never be eliminated!
- All windows distort data!
- Almost all the time when performing a modal test, the input excitation can be selected such that the use of windows can be eliminated. e.g.,signals such as pseudo random, burst random, sine chirp, and digital stepped sine.

Spectra and FRFs

- *Definition*
- *Pads and filters*

Averaging

- Suppose that now we want to estimate the spectrum of a random signal. In the limit as $\tau \longrightarrow \infty$, we would get an accurate spectrum but for finite τ , we have a problem. Any finite realisation of a random process will not represent exactly the long term frequency content precisely because it is random.
- Assuming no problems with aliasing we will find

$$\mathcal{F}[x_\tau(t)] = X(\omega) + \in_\tau (\omega)$$

where X is the true spectrum and $\in_\tau$ is an error term associated with the finite sample size. Now for each spectral line $\in_\tau$ is a random variable and is just as likely to cause an underestimate as an overestimate.

- This means we can remove it by averaging

$$E[\mathcal{F}[x_\tau(t)]] = E[X(\omega)] + E[\in_\tau (\omega)] = X(\omega)$$

Averaging

- The averaging can be implemented by taking N segments of time data $x_i(t)$ and transforming to

$$X_i(\omega) = X(\omega) + \in_{ri}(\omega)$$

then

$$X(\omega) \approx E[X_i(\omega)] = \frac{1}{N}\sum_{i=1}^{N} X_i(\omega)$$

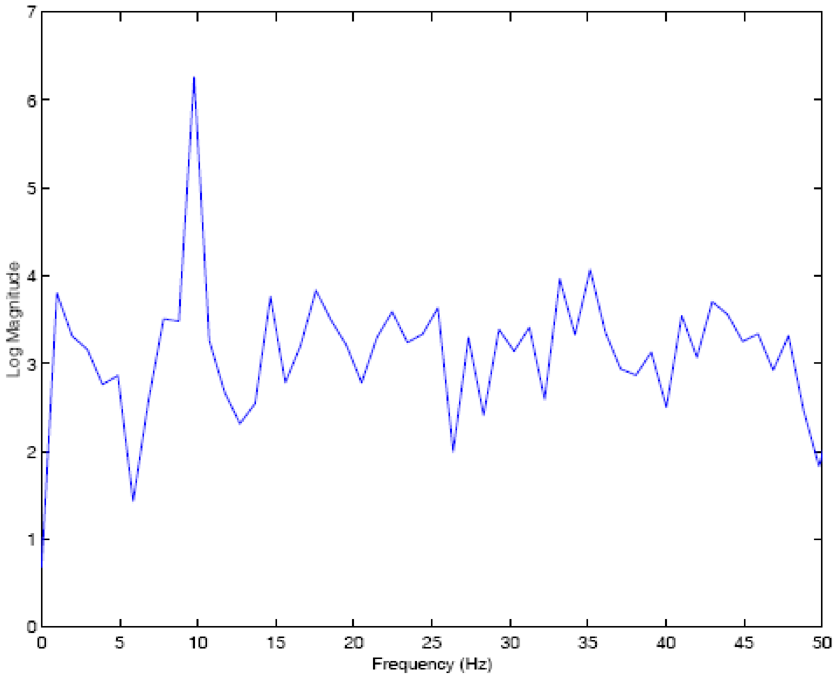
- For a signal

$$x(t) = sin(\omega t) + 0.25N(0,1)$$

- The frequency of the sine wave is chosen such that it is periodic over the window, so we don't have to worry about leakage from the sine wave.

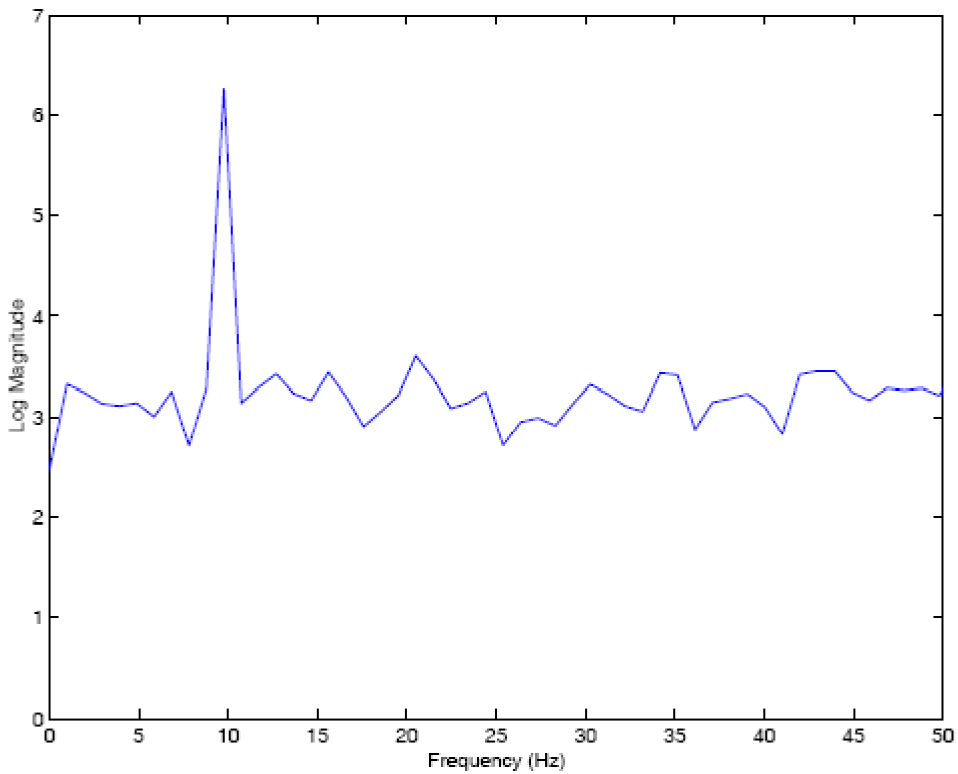
Averaging

- The first is one average- a one-shot measurement. Although the sine wave (at 10.24 Hz) is visible, there is a lot of background noise from the single average.



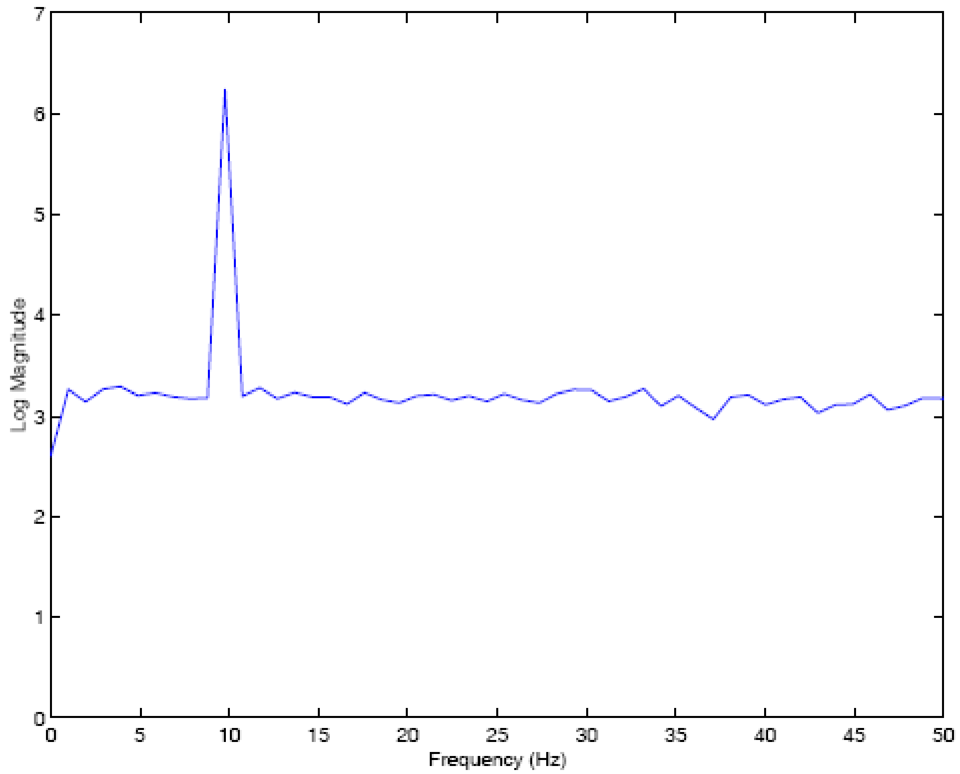
Averaging

- The next figure shows the result of taking 10 averages.



Averaging

- Finally, we see the effect of taking 100 averages.



Power spectral density

- Often we plot $E[|X|]^2$ as this is proportional to power. It is called the power spectral density and is denoted

$$S_{xx}(\omega) = E[|X(\omega)|^2] = E[X(\omega)X(\omega)^*]$$

- The autocorrelation of a random process is defined by:

$$\phi_{xx}(\tau) = E[x(t+\tau)x(t)]$$

- It is a measure of how much a signal looks like itself when shifted by an amount τ. It is used to find regularities in data. Suppose that $x(t)=\sin(2\pi t/\tau')$, then there will be regular peaks in $\phi_{xx}(\tau)$ when $\tau=n\tau'$. So the autocorrelation function can also be used to detect periodicities.

Power spectral density

- If *x(t)* is zero mean, then

$$\phi_{xx}(0) = E[x(t)^2] = \sigma_x^2$$

and if x is not zero mean, $\phi_{xx}(0)$ is the mean square of the process.

- As *x(t)* is stationary, we can change the origin of *t* to *t*- τ without changing the autocorrelation, i.e.

$$\phi_{xx}(\tau) = E[x(t+\tau)x(t)] = E[x(t)x(t-\tau)] = E[x(t-\tau)x(t)] = \phi_{xx}(-\tau)$$

- So $\phi_{xx}(\tau)$ is an even function of τ

Filters

- *Definition*
- *Pads and filters*

Filters

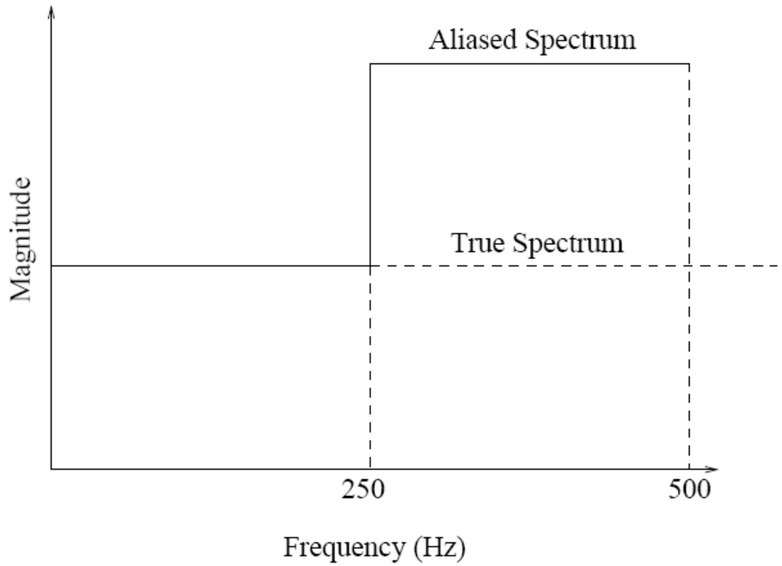
- Assume that we are trying to build a Fourier transforming device which can give us the spectrum of a given time signal. Suppose that we have a maximum sampling frequency of 1000 Hz i.e. a Nyquist frequency of 500 Hz.

- If the time signal has a broadband spectrum which is flat up to 750 Hz, what will the estimated spectrum look like? So energy is aliased into the range 250-500 Hz from the range 500-750 Hz and we obtain a completely fictitious spectrum.

- How can we help this?Suppose we had a device which removed the part of the signal at frequencies between 500 and 750 Hz. Then we would have changed the signal admittedly but the FFT would at least give us an accurate spectrum all the way up to 500 Hz.

Filters

- Such a device which passes parts of a signals frequency content and suppresses others is called a **filter**. The particular filter described above is called an **antialiasing filter** for obvious reasons.



Filters

- A **filter** is a function that in the frequency domain has a value close to 1 in the range of frequencies that the analyst wishes to retain and close to zero in the range of frequencies that the analyst wishes to eliminate.

- The filter can be applied in the time domain, by convolution of its transform with the time history, or in the **frequency domain** by **multiplying the filter frequency response function with the Fourier amplitude spectrum (FAS) of the time history**, and then obtaining the filtered time history through the inverse Fourier transform.

$$y(t) = IDFT\left[H(f)X(f)\right]$$

Filters

- Equally unimportant is the choice of the actual generic filter: users are faced with a wide range of filters to choose from, including Ormsby, elliptical, Butterworth, Chebychev and Bessel.
- The correct application of the chosen filter is much more important than the choice of a particular filter.
- The terminology used to describe filters can be confusing, especially for engineers more accustomed to thinking in terms of periods than frequencies.
- A filter that removes high frequencies (short periods) is usually referred to as a low-pass filter because motion at lower frequencies gets through and higher frequencies are, in effect, blocked by the filter. For such a filter civil engineers prefer the term high-cut, which refers directly to the frequencies being removed.

Classification of filters

- If it is judged that there is significant high-frequency noise in the record, or if for some other reason it is desirable to reduce or remove high frequencies introduced by interaction effects at the recording station, this can be easily achieved by the application of filters.

- Filters can be applied in the frequency domain or the time domain but their function is best understood in the frequency domain.

- If the filter is a mechanical or electrical device which operates on the continuous time physical signal it is called an **analogue filter**.

- If the filter is a numerical algorithm or mechanical device which operates on sampled data it is called a **digital filter**.

High pass-low pass filters

- The purpose of a low pass filter is to remove that part of the signal that is judged to be heavily contaminated by high-frequency noise that is often observed in strong-motion records.

- A low-pass filter passes low-frequency components in the signal with minimal distortion and attenuates the high-frequency components.

- The so-called cutoff or corner frequency divides the pass band and the stop band. A low pass causal Butterworth filter is an all pole filter with a squared magnitude response given by:

$$|H(\omega)|^2 = \frac{1}{1 + (\frac{\omega}{\omega_c})^{2n}}$$

High pass-low pass filters

- Two considerations are important when applying a highcut filter.

  – The first is that the application of the filter will act in a contradictory manner to any instrument correction and at least in some frequency ranges the two will counteract each other.

  – The second consideration is that an upper frequency limit on the usable range of high frequencies in the motion is imposed by the sampling rate: the Nyquist frequency, which is the highest frequency at which characteristics of the motion can be correctly determined, is equal to ($1/2\Delta t$) where Δ t is the sampling interval.

  – A high-cut filter applied at frequencies greater than the Nyquist will have no effect on the record.

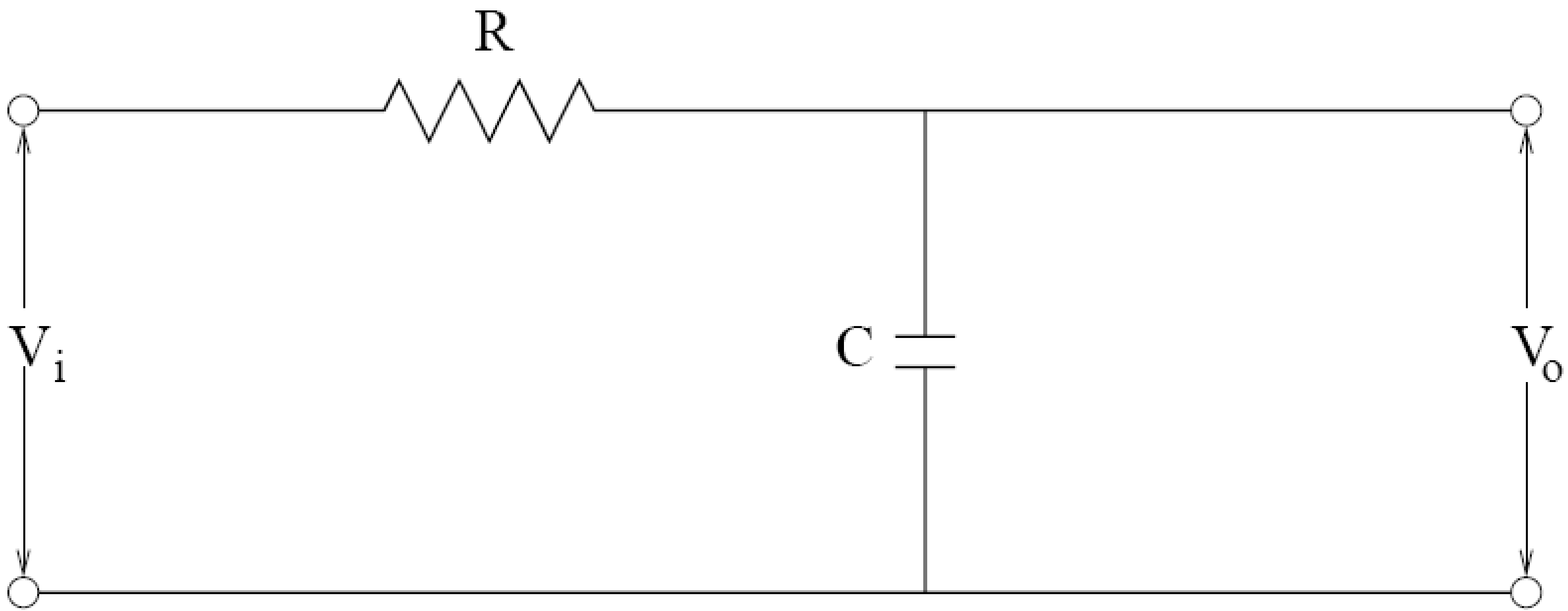
Analogy between mechanical and electrical systems

	Electrical system	Mechanical system
Input	Voltage, e(t)	Force, F(t)
Output	Charge q(t)	Displacement y(t)
	Current i(t)=dq/dt	Velocity v(t)=dy/dt
Constant parameters	Inductance, L	Mass,m
	Resistance, R	Damping,c
	Capacitance,C	Compliance, 1/k

Analogue filters

- We will start the discussion with an electrical analogue filter. Consider the circuit below with an alternating voltage input,

$$V_i(t) = V_i \cos(\omega t)$$



Analogue filters

- Elementary circuit theory gives the output voltage $V_o(t)$ as the solution of the differential equation:

$$RC\frac{dV_o}{dt}+V_o=V_i(t)$$

where R is the resistance and C is the capacitance. Passing to the frequency domain gives:

$$iRC\omega V_o(\omega)+V_o(\omega)=V_i(\omega)$$

- So

$$V_o(\omega)=H(\omega)V_i(\omega)$$

where

$$H(\omega)=\frac{1}{1+iRC\omega}$$

Analogue filters

- The gain of the system is:

$$\left|H(\omega)\right| = \frac{1}{\sqrt{1+R^2C^2\omega^2}}$$

and the phase is :

$$\angle H(\omega) = -\tan^{-1}(RC\omega)$$

- When $RC\omega=0.1$, $|H(\omega)|=0.995$
- When $RC\omega=10$, $|H(\omega)|=0.0995$
- As we have a filter that attenuates high frequencies and passes low frequencies, it is called a low-pass filter.

Digital low pass filters

- How can we implement a filter of the sort described above on sampled data?
- First let us adopt a more general notation. Let x(t) or x_i be the input to the filter and y(t) or y_i be the output. The differential equation of the electrical analogue filter is then,

$$RC\frac{dy}{dt} + y = x(t)$$

- Now suppose that x and y are sampled with an interval Δt, so $x(t) \longrightarrow x_i = x(t_i) = x(i\,\Delta t)$ and $y(t) \longrightarrow y_i = y(t_i) = y(i\,\Delta t)$. The derivative above can be approximated by:

$$\frac{dy_i}{dt} = \frac{y_i - y_{i-1}}{\Delta t}$$

Digital low pass filters

- This results in a difference equation

$$\frac{RC}{\Delta t}(y_i - y_{i-1}) + y_i = x_i$$

- With a bit of arrangement:

$$y_i = \left(\frac{RC}{\frac{RC}{\Delta t}+1}\right) y_{i-1} + y_i = \left(\frac{1}{\frac{RC}{\Delta t}+1}\right) x_i$$

$$y_i = a_1 y_{i-1} + b_0 x_i$$

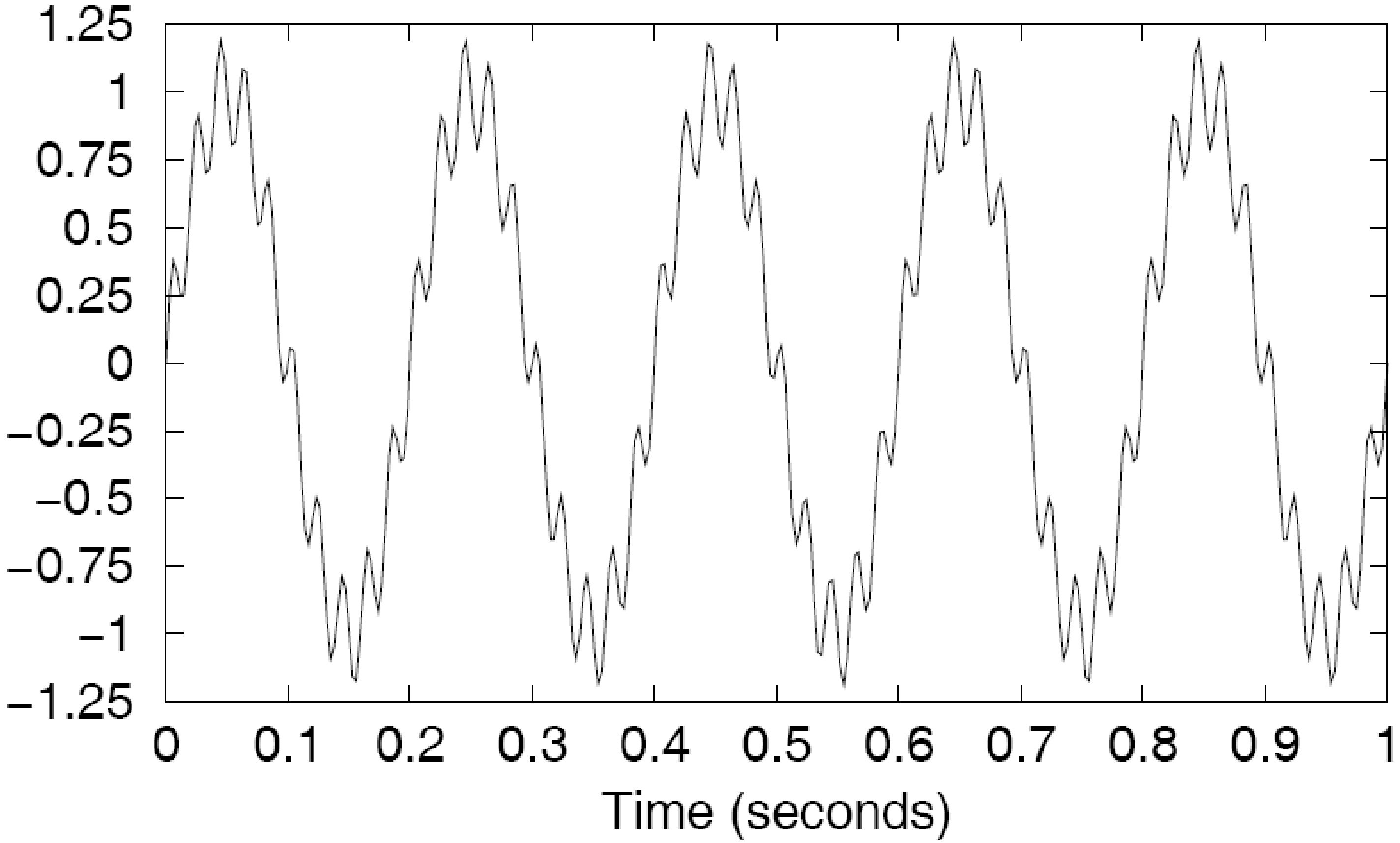
with appropriate definitions for a_1 and b_0. Consider the signal

$$x(t) = \sin(2\pi .5t) + \sin(2\pi .50t)$$

sampled with Δt=0.003 as shown in the next slide.

Digital low pass filters

- Noisy sine wave

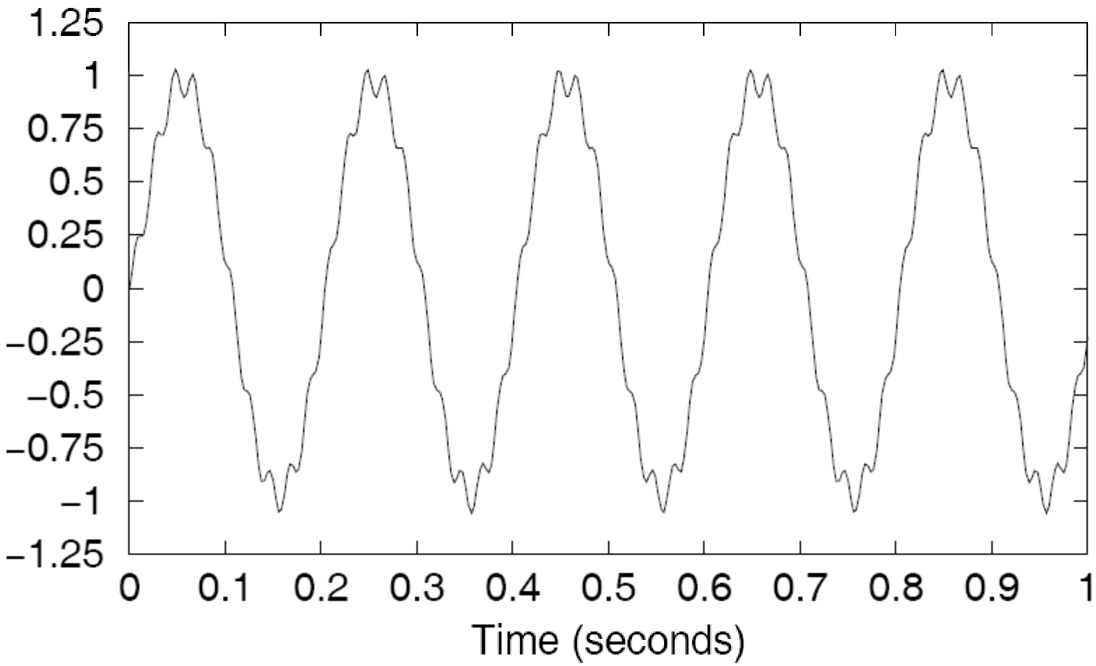


Digital low pass filters

- After one pass through the digital filter

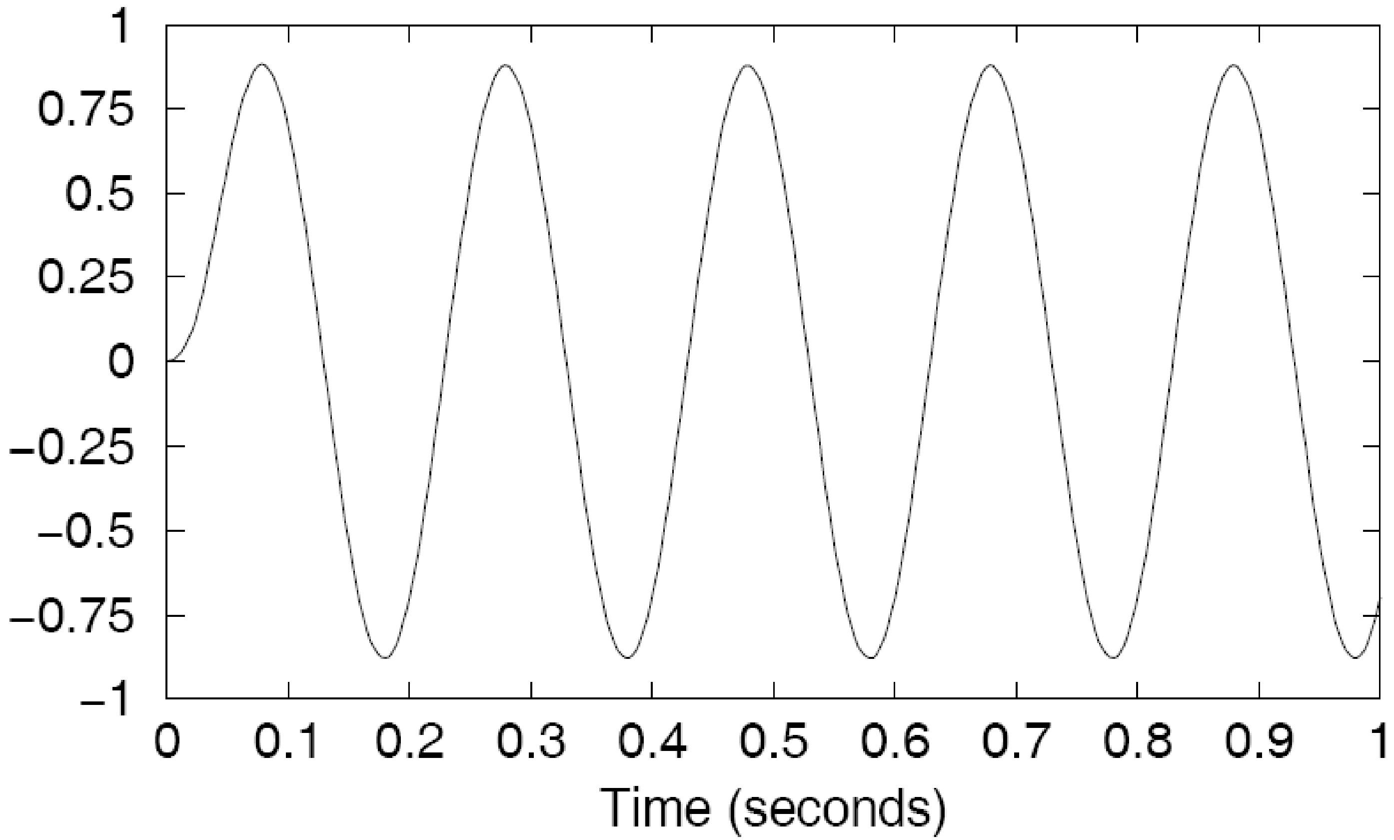
$$y_i = 0.6655 y_{i-1} + 0.3345 x_i$$

- The resulting noisy sine wave after one pass through the filter is:



Digital low pass filters

- The resulting noisy sine wave after five passes through the digital filter is:



Digital low pass filters

- So the filter is able to remove the high frequency sine-wave effectively. However, note that the **amplitude** of the carrier signal has also been **attenuated**. Also, importantly, the **phase** of the signal has been **altered**.

- The next question we should ask is-what is the frequency response of the new digital filter derived from the original analogue filter?

- We use a theorem from an earlier lecture which says that if the input to a system is $x(t)=e^{iwt}$, then the response is $y(t)=H(w)\, e^{iwt}$. Now define the backward shift operator Z^{-1} by:

$$Z^{-1} y_i = y_{i-1}$$

Digital low pass filters

- This allows us to write the equation

$$y_i = a_1 y_{i-1} + b_0 x_i$$

- As:

$$(1 - a_1 Z^{-1}) y_i = b_0 x_i$$

- Let

$$x_i = e^{i\omega t_i}$$

- Then

$$y_i = H(\omega) e^{i\omega t_i}$$

$$Z^{-1} e^{i\omega t_i} = e^{i\omega t_{i-1}} = e^{i\omega (t_i - \Delta t)} = e^{-i\omega \Delta t} e^{i\omega t_i}$$

Digital low pass filters

- Then

$$(1-a_1Z^{-1})y_i = b_0x_i$$

becomes:

$$(1-a_1e^{-i\omega\Delta t})y_i = b_0x_i$$

- It follows that the FRF is:

$$H(\omega) = \frac{b_0}{1-a_1e^{-i\omega\Delta t}}$$

- In terms of the FRF derived above, we have enough now to obtain a general result. A general digital filter would then be:

$$y_i = \sum_{j=1}^{n_y} a_j y_{i-j} + \sum_{j=0}^{n_x} b_j x_{i-j}$$

A high pass filter

- Recall the formula for the low pass filter:

$$H(\omega) = \frac{1}{1 + iRC\omega}$$

- The desired properties for a low pass filter are that:

$$\left|H(\omega)\right| \to 1 \quad \text{as} \quad \omega \to 0$$
$$\left|H(\omega)\right| \to 0 \quad \text{as} \quad \omega \to \infty$$

- The desired properties for a high pass filter would be:

$$\left|H(\omega)\right| \to 0 \quad \text{as} \quad \omega \to 0$$
$$\left|H(\omega)\right| \to 1 \quad \text{as} \quad \omega \to \infty$$

A high pass filter

- Now we can obtain this characteristic by making a simple transformation on the below equation:

$$H(\omega) = \frac{1}{1 + iRC\omega}$$

- Namely:

$$i\omega \to \frac{1}{i\omega} \quad \text{or} \quad \omega \to -\frac{1}{\omega}$$

- The FRF of the low pass filter becomes:

$$H(\omega) = \frac{1}{\frac{RC}{i\omega} + 1} = \frac{\frac{i\omega}{RC}}{1 + \frac{i\omega}{RC}} \quad \text{and the gain is} \quad \left| H(\omega) \right| = \frac{\frac{\omega}{RC}}{\sqrt{1 + \frac{\omega^2}{R^2C^2}}}$$

A high pass filter

- This filter is high pass as required.

$$\text{When } \frac{\omega}{\mathrm{RC}} = 0.1, \qquad |H(\omega)| = 0.0995$$

$$\text{When } \frac{\omega}{\mathrm{RC}} = 10, \qquad |H(\omega)| = 0.995$$

- One of the most useful families of analog filters is that of **Butterworth filters**. These are controlled by two parameters for the low-pass filter. The FRF gain is specified as:

$$|H(\omega)|^2 = \frac{1}{1+\left(\frac{\omega}{\omega_c}\right)^{2n}} = \frac{1}{1+\left(\frac{f}{f_c}\right)^{2n}}$$

where ω_c is the cut-off frequency and n is a steepness factor which specifies how fast the signal should die away after the cut off frequency.

High pass-low pass filters

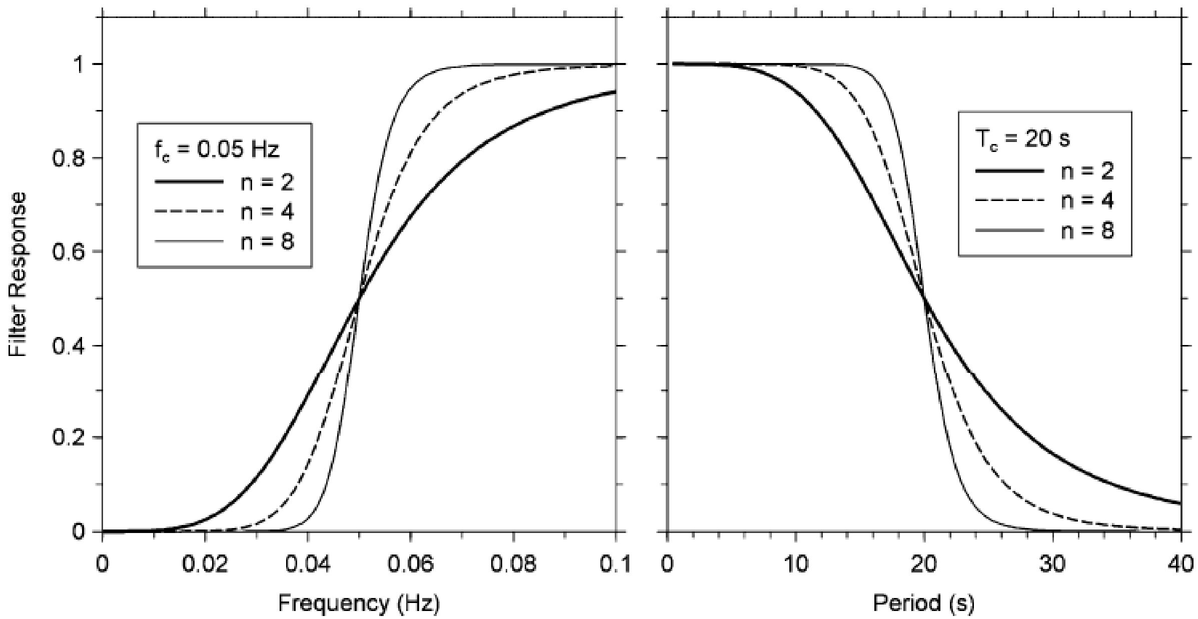
$$|H(\omega)|^2 = \frac{1}{1 + (\frac{\omega}{\omega_c})^{2n}}$$

where n is the order of the filter (number of poles in the system function) and ω_c is the cut off frequency of the Butterworth filter which is the frequency where the magnitude of the causal filter $|H(\omega_c)|$ is $1/\sqrt{2}$ regardless of the order of the filter.

- The purpose of a low-cut filter is to remove that part of the signal that is judged to be heavily contaminated by long-period noise. The key issue is selecting the period beyond which the signal-to-noise ratio is unacceptably low. Applying a filter that abruptly cuts out all motion at periods above the desired cut-off can lead to severe distortion in the waveform, and therefore a transition—sometimes referred to as a ramp or a rolloff— is needed between the pass-band, where the filter function equals unity, and the period beyond which the filter function is equal to zero.

High pass-low pass filters

- The figure shows an illustration of a low-cut Butterworth filter as a function of frequency and period. The filter frequency is 0.05 Hz, which means that periods above 20 s are at least partially removed. The different curves are for different orders of filter: the higher the order, the more abrupt the cut-off (the more rapid the roll-off (but with increased filter-response oscillations for the higher order filters)). For the lower order filters, information will be removed from periods as low as 10 s.



Causal-acausal filters

- Although the choice of filter type is less important, the way in which the filter is applied to the accelerogram has been shown to be very important.

- The fundamental choice is between **causal** and **acausal** filters, the distinguishing feature of the latter being that they do not produce any **phase distortion** in the signal, whereas **causal filters** do result in phase shifts in the record.

- The **zero phase shift** is achieved in the **time domain** by passing the transform of the filter along the record **from start to finish and then reversing the order** and passing the filter from the end of the record to the beginning.

Causal-acausal filters

- The imlementation of a zero phase filtering scheme is shown in the figure.

$$u[n]=v[-n]\ , \qquad y[n]=w[-n]$$

High pass-low pass filters

- $|H(\omega_c)|$ takes the value of 0.5 for acausal filters. The mathematical reasoning behind this can be explained as follows: If we let $H_c(\omega)$ be the frequency response of the causal Butterworth filter given by the second equation, this filter has unity gain at frequency $\omega = 0$. The cutoff frequency ω_c is the frequency at which the power of the filter ouput is half the power of the filter input,i.e. $|H_c(\omega_c)|^2 = 1/2$. The frequency response of an acausal Butterworth filter $H_a(\omega)$ is given by:

$$H_a(\omega) = H_c(\omega)H_c(-\omega) \qquad |H(\omega)|^2 = \frac{1}{1+(\frac{\omega}{\omega_c})^{2n}}$$

- The previous equation shows that $|H_a(\omega_c)|=1/2$.
- The frequency response of a Butterworth filter decreases monotonically with increasing frequency, and as the filter order increases, the transition band becomes narrower.

Classification of filters

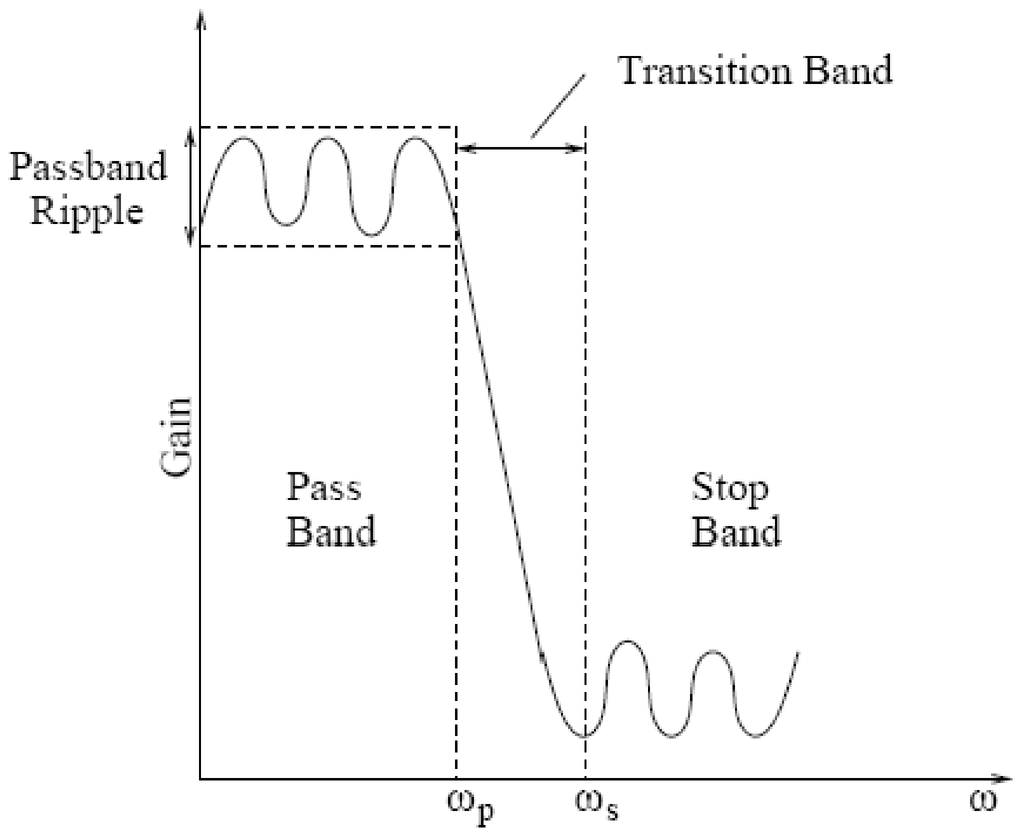


Figure 9. Gain of generic low-pass filter.

Classification of filters

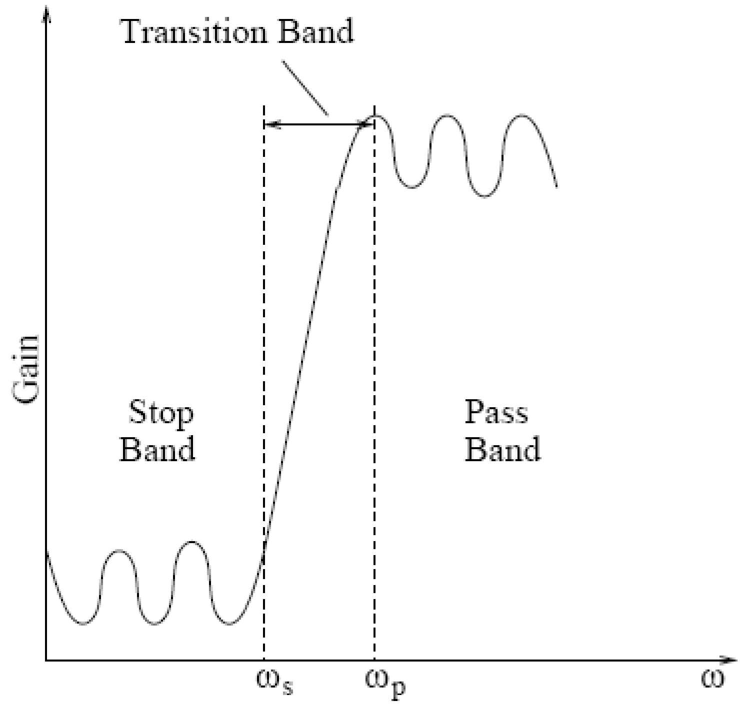


Figure 10. Gain of generic high-pass filter.

Classification of filters

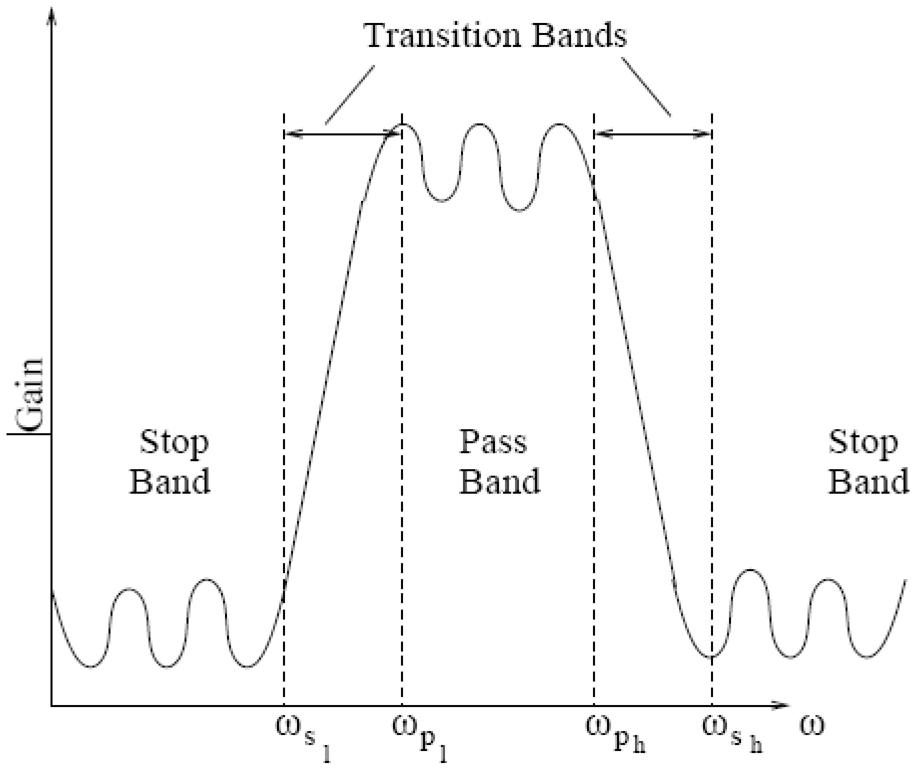


Figure 11. Gain of generic band-pass filter.

Classification of filters

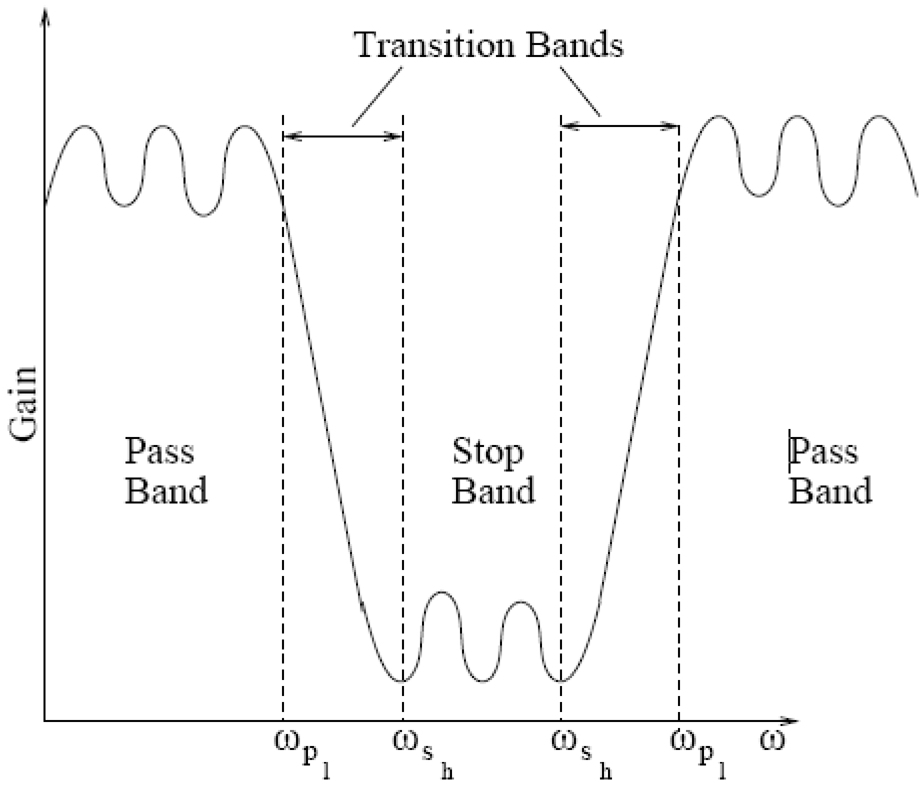


Figure 12. Gain of generic band-stop filter.

Overview of excitation signals

Classification

- The commonly used **excitation signals** can be categorized in several ways. For practical purposes, it is easy to consider two main groups: **broad band** or **single frequency signals**.

- The signal frequency group contains:
  - Swept sine
  - Stepped sine
- The broadband group consist of three subgroups:
  - Transients
  - Periodic
  - Nonperiodic

Classification

- Transients
  - Burst random
  - Burst chirp (or burst swept sine)
  - İmpact excitation
- Periodic
  - Pseudo random
  - Periodic random
  - Chirp (fast swept sine)
- Nonperiodic
  - Pure random

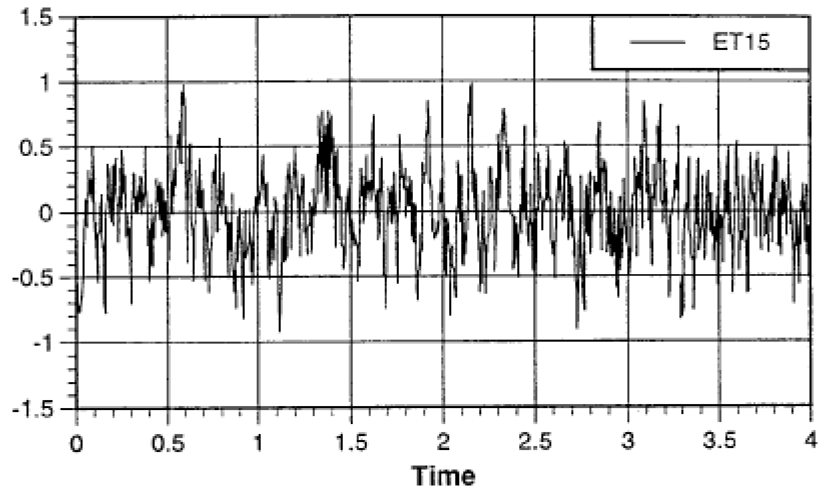
Classification

- Random signals can only be defined by their statistical properties.
- For **stationary** random signals, these properties do not vary with respect to translations in time.
- All random excitation signals are of the **ergodic** random type, which means that a time average on any particular subset of the signal is the same for any arbitrary subset of the random signal.

Pure random

- Pure random is a nonperiodic stochastic signal with a Gaussian probability distribution. Averaging is essential when estimating the frequency spectrum.

- The main problem of the pure random signal is **leakage**. Since the signal is not periodic within the observation time window, this error can not be avoided. The application of dedicated time windows (e.g. Hanning) to the input and output signals can not completely remove the effects of leakage without causing undesired side effects such as a decreased frequency resolution.



Pure random

- Pure random easily averages out noncoherent noise.
- It yields the best linear approximation of nonlinear systems, since in each averaged time record, the nonlinear distortions will be different and tend to cancel with sufficient averaging.
- Test time is relatively long due to the necessary number of averages.
- However, the total time becomes shorter when using **overlap averaging**. In the overlap averaging procedure, each averaged time record will contain the last part of the previous one.

Pseudo random

- The **pseudo random** is an ergodic stationary signal with a spectrum consisting of integer multiples of the discrete Fourier transform frequency increment. Hence it is perfectly periodic within the sample time window.

- Due to the periodicity of the signal, no leakage problem exists.

- However, since the same time block is repeated for averaging, pseudo random excites the nonlinearities the same way in each average. Therefore, averaging will not remove distortion caused by nonlinearities.

- For linear structures, only a few averages are necessary in general. Hence this excitation signal may be very fast.

Periodic random

- Periodic random excitation is simply a different use of a pseudo random signal, so that non-linearities can be removed with spectrum averaging.
- For periodic random testing, **a new pseudo random sequence is generated for each new spectrum average**.
- The advantage of this is that when multiple spectrum averages of different random signals are averaged together, randomly excited non-linearities are removed.
- Although periodic random excitation overcomes the disadvantage of pseudo random excitation, it takes **at least three times longer** to make the same measurement. This extra time is required between spectrum averages to allow the structure to reach a new steady-state response to the new random excitation signal.

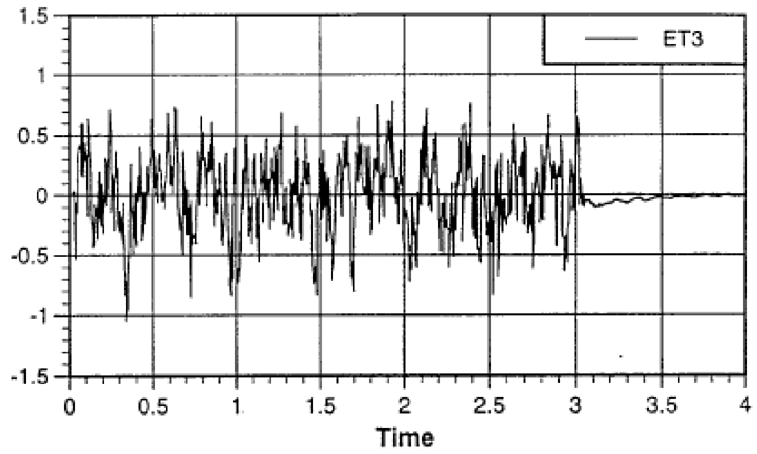
Periodic random

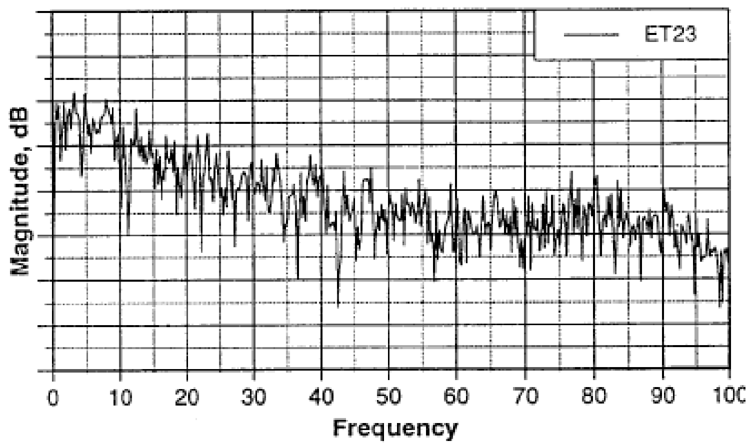
- Other advantages are:
  - Signals are periodic in the sampling window, so measurements are leakage free.
  - Removes non-linear behavior when used with spectrum averaging.

- Disadvantages are:
  - Slower than other random test methods.
  - Special software required for implementation

Burst random

- **Burst random excitation** is similar to periodic random testing,but faster.
- In burst random testing, a true random signal can be used, but it is **turned off** prior to the end of the sampling window time period.
- This is done in order to allow the structural response to decay within the sampling window. This insures that both the excitation and response signals are completely contained within the sampling window.
- Hence, they are **periodic in the sampling window.**





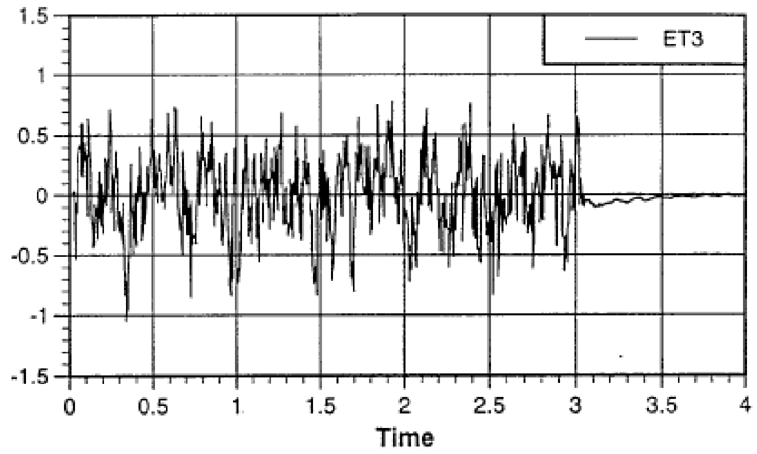
Burst random

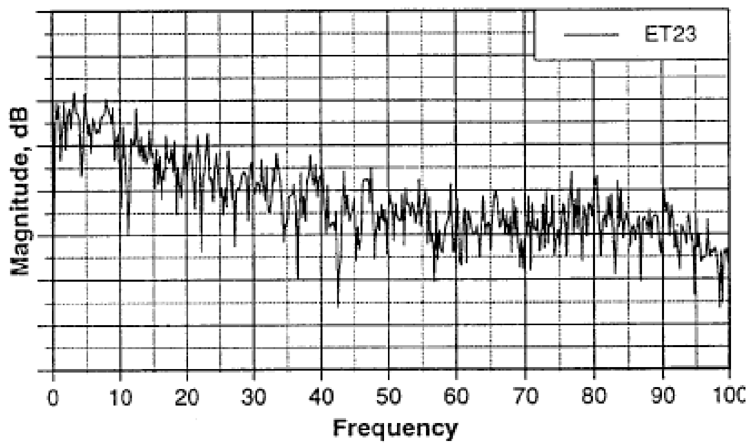
Advantages :

- Signals are periodic in the sampling window, so measurements are leakage free.
- Removes non-linear behavior when used with spectrum averaging.
- Fast measurement time.

Disadvantages :

- Special software required for implementation.





Burst random

- **The most commonly** used excitation for modal testing!
- In order to have the entire transient be captured, the length of the excitation burst can be reduced.
- Generally, the use of **windows** for this type of excitation technique is **not required!**

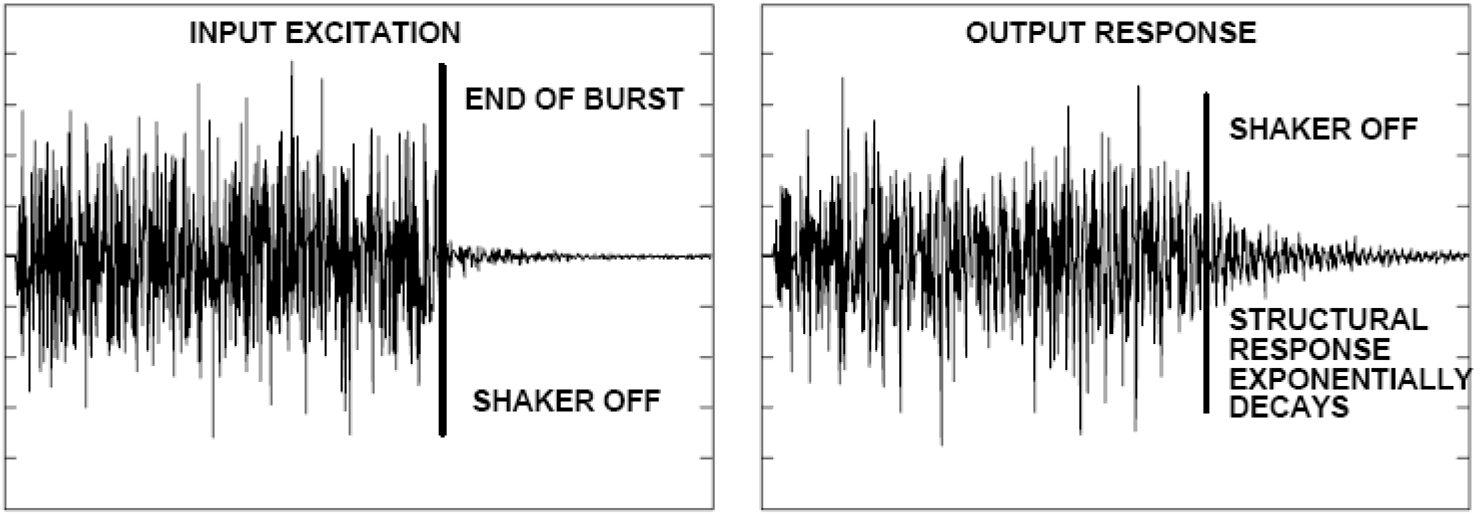
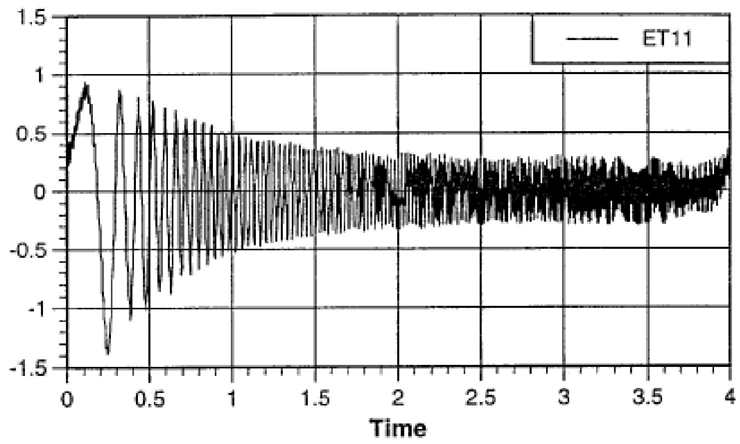


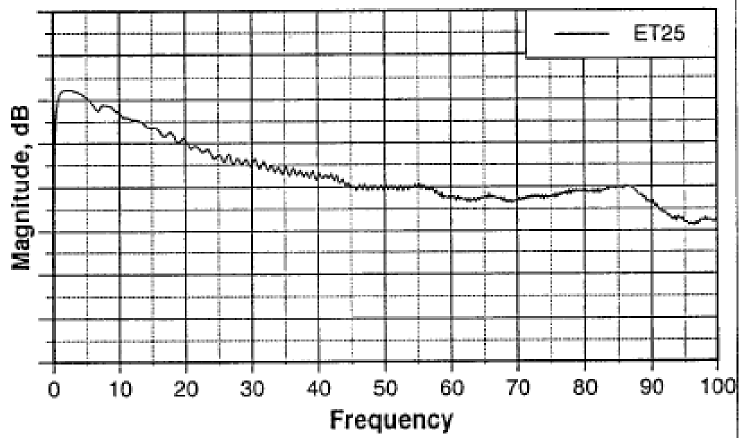
Figure 3 - Typical Burst Random Measurement Sequence

Chirp and Burst Chirp

- A swept sine excitation signal can also be synthesized in an FFT analyzer to coincide with the parameters of the sampling window, in a manner similar to the way a pseudo random signal is synthesized.



- Since the sine waves must sweep from the lowest to the highest frequency in the spectrum, over the relatively short sampling window time period (**T**), this fast sine sweep often makes the test equipment sound like a bird chirping, hence the name chirp signal.



Chirp and Burst Chirp

- A burst chirp signal is the same as a chirp, except that it is **turned off** prior to the end of the sampling window, just like burst random.

- This is done to ensure that the measured signals are **periodic in the window**.

Chirp and Burst Chirp

- The advantage of burst chirp over chirp is that the structure has returned to rest before the next average of data is taken.

- This insures that the measured response is only caused by the measured excitation, an important requirement for any multichannel measurement such as a FRF.

Chirp and Burst Chirp

Advantages of Burst Chirp Excitation

- High signal–to-noise and RMS-to-peak ratios.
- Signals are periodic in the sampling window, so measurements are leakage free.
- Fast measurement time.

Disadvantages of Burst Chirp Excitation

- Special software required for implementation.
- Doesn’t remove non-linear behavior.

Swept sine

- The sine wave excitation signal has been used since the early days of structural dynamic measurement. It was the only signal that could be effectively used with traditional analog instrumentation.

- Even broad band testing methods (like impact testing), have been developed for use with FFT analyzers, sine wave excitation is still useful in some applications. The primary purpose for using a sine wave excitation signal is to put energy into a structure at a specific frequency.

- Slowly sweeping sine wave excitation is also useful for characterizing non-linearities in structures.

Swept sine

Advantages of Sine Testing

- Best signal-to-noise and RMS-to-peak ratios of any signal.
- Controlled amplitude and bandwidth.
- Useful for characterizing non-linearities.
- Long history of use.

- **Disadvantages of Sine Testing**
- Distortion due to over-excitation.
- Extremely slow for broad band measurements.

Stepped sine

- Stepped sine excitation is a modern version of the swept sine technique that makes maximum use of the developments in DSP during the last two decades.
- Instead of a continuously varying frequency, stepped sine consists of a stepwise changing frequency.
- It remains a rather slow procedure due to the frequency scan and wait periods needed for the transients to decay. This can be overcome by multi-channel acquisition.
- The application of stepped sine excitation requires special soft and hardware.
- The digital processing allows for varying frequency spacing, yielding data condensation and testing time reduction, and for a better control against aliasing and leakage problems.
- Useful for characterizing non-linearities.
- Excellent signal-to-noise ratios.

ERASMUS Teaching (2008), Technische Universität Berlin

Processing strong-motion records

The focus of this section is on the effects of noise in accelerograms, and the effects of 'correction' procedures, on the peak ground-motion amplitudes and on the ordinates of acceleration and displacement response spectra.

Processing of earthquake records

- High-pass filters are an effective way of removing the low-frequency noise that is present in many, if not most, analog and digital strong-motion recordings.
- This low frequency noise usually appears as drifts in the displacements derived from double integration of acceleration, making it difficult to determine the true peak displacement of the ground motion.
- It can never be claimed that a complete and accurate description of the ground shaking can be obtained from accelerograms.

Processing of earthquake records

- For engineering uses of strong-motion data it is important to be able to estimate the level of noise present in each accelerogram and the degree to which this may affect different parameters that are derived from the records.

- The main parameters of interest for engineering application are:

  – The ordinates of response spectra, both of acceleration and displacement.

  – The peak ground acceleration (PGA),although of limited significance from both geophysical and engineering perspectives, is also a widely used parameter in engineering.

  – The peaks of velocity (PGV),and displacement (PGD), measured from the time-histories obtained by integration of the acceleration, are also important parameters.

Processing of earthquake records

- It is important for users of strong-motion data to appreciate that digitized accelerograms are never pure and complete reproductions of the seismic signal.

- From the outset, however, it is important to be clear that it is not possible to identify, separate and remove the noise in order to recover the unadulterated seismic signal.

- The best that can be achieved in general is to identify those portions of the frequency content of the record where the signal-to-noise ratio is unacceptably low and to thus identify that portion of the record, in the frequency domain, that can be used with some confidence.

Analog accelerograms

- In light of these considerations, it is not appropriate to refer to most of the processing procedures described herein as ‘corrections’, since the term implies that the real motion is known and furthermore that it can be recovered by applying the procedures.

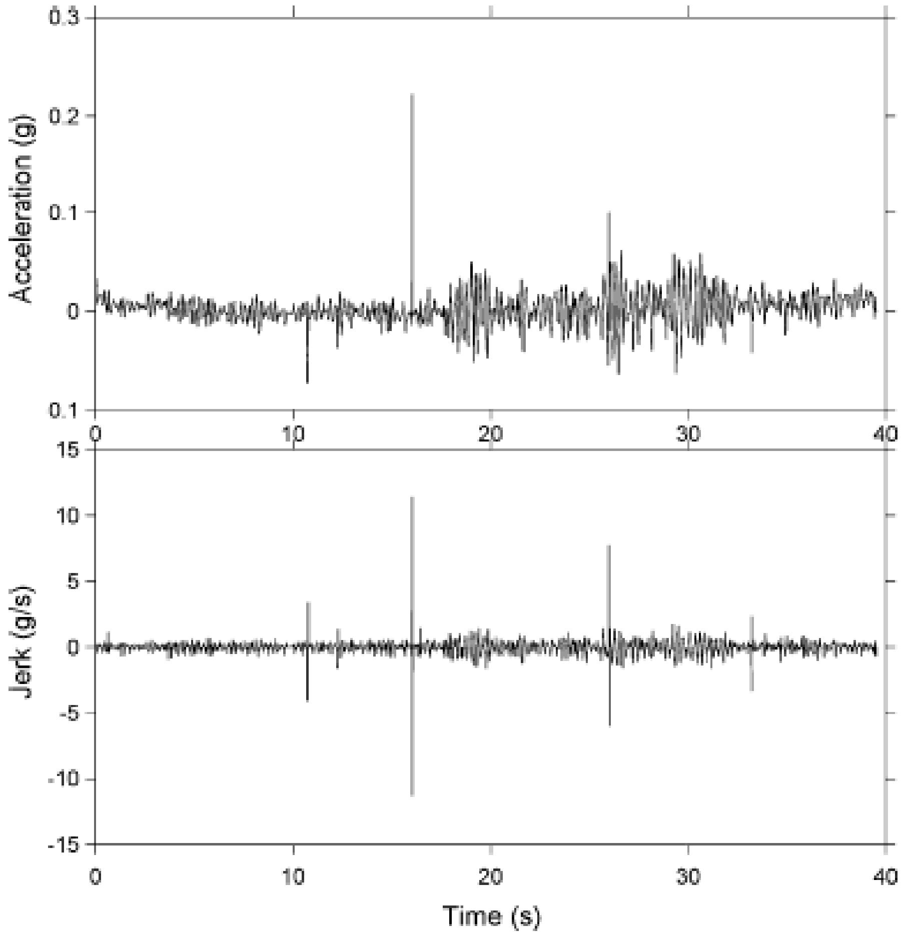
- In order to estimate the signal-to-noise ratio, a model of the noise in the digitized record is required. Most analog accelerographs, such as the SMA-1, produce two fixed traces on the film together with the three traces of motion (two horizontal, one vertical) and the time marks. If these fixed traces are digitized together with the motion, then any ‘signal’ they contain can be interpreted as being composed entirely of noise since the traces are produced by infinitely stiff transducers that experience no vibration during the operation of the instrument.

- Unfortunately, the fixed traces are very often not digitized or else the digitized fixed traces are not kept and distributed with the motion data, hence it is rare that a model of the noise can be obtained from this information.

Standard vs nonstandard noise

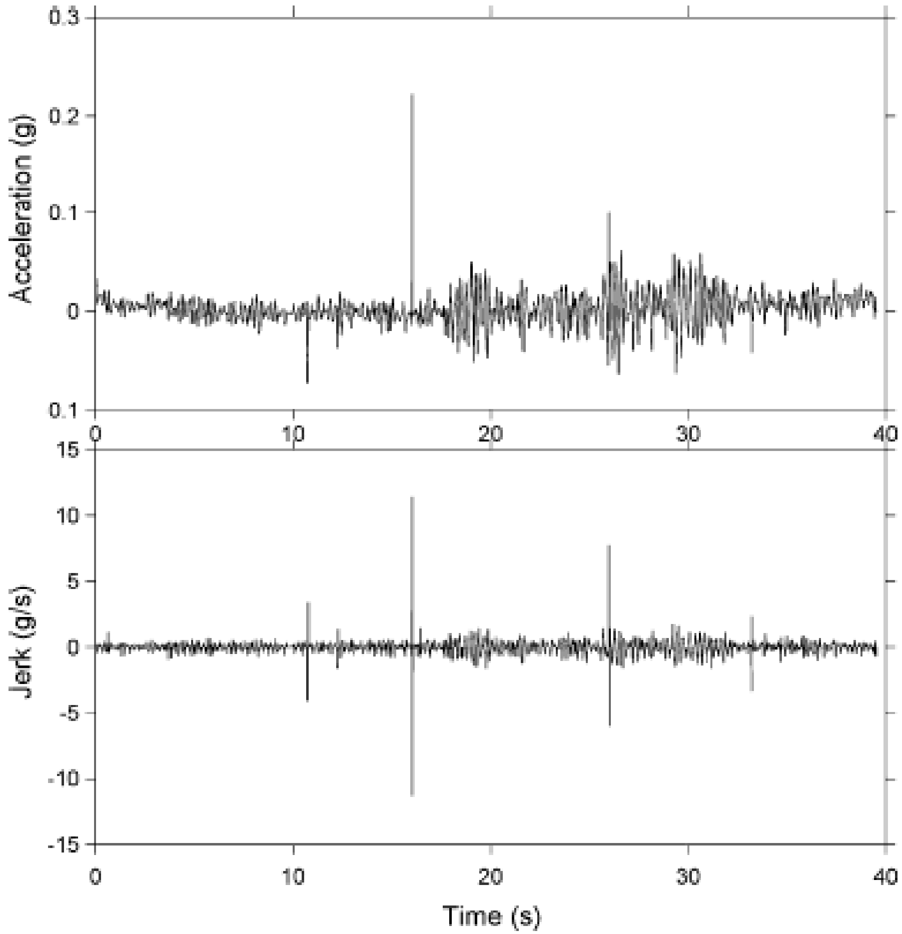
- Some types of noise, particularly step changes in the baseline, can also be identified from the ‘**jerk**’, which is the first derivative of the acceleration trace.

- Figure shows horizontal component of the Bajestan recordings of the 1978 Tabas earthquake in Iran; spurious spikes are obvious in the acceleration record at 10.8 and 16 s. The derivative of the acceleration trace (to produce the quantity called ‘jerk’) will convert a spike into a double sided pulse, making it easier to identify spikes. By doing this (bottom panel), spikes at 12.3, 26 and 33.2 s are also identified.



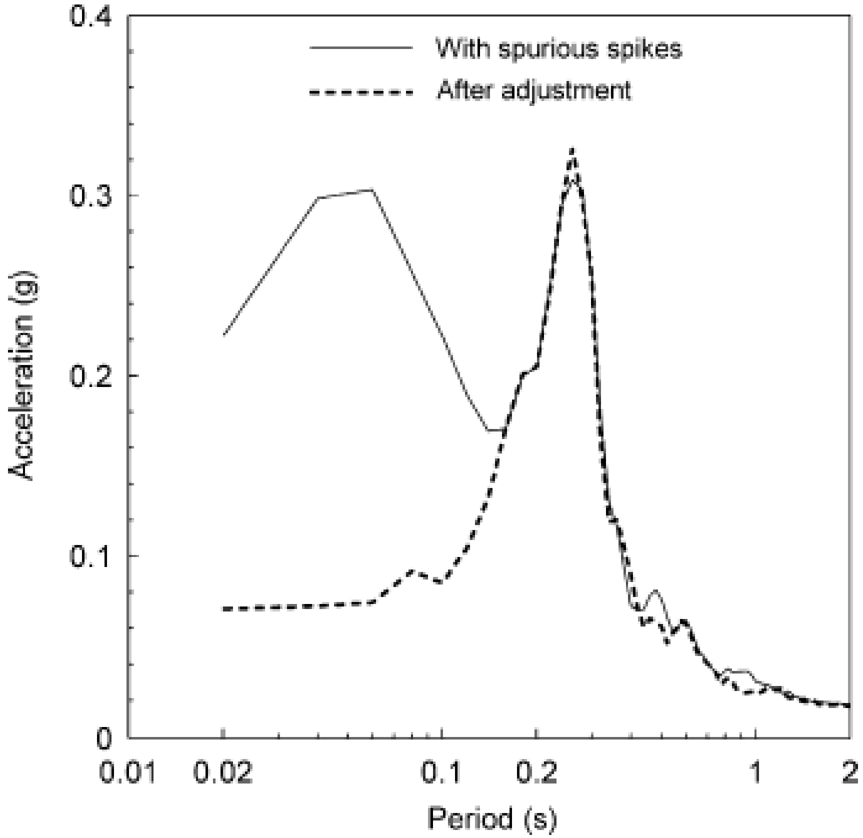
Standard vs nonstandard noise

- In this particular case, the spurious nature of these spikes was confirmed by comparison with a reproduction of the original analog record.
- The origin of the spikes has not been ascertained, although a possible cause in this instance was the misplacement of the decimal point in transcribing the digitized values.
- Once the spikes have been identified as erroneous, they should be removed from the digitized record; one way to achieve this is replace the acceleration ordinate of the spike with the mean of the accelerations of the data points either side.



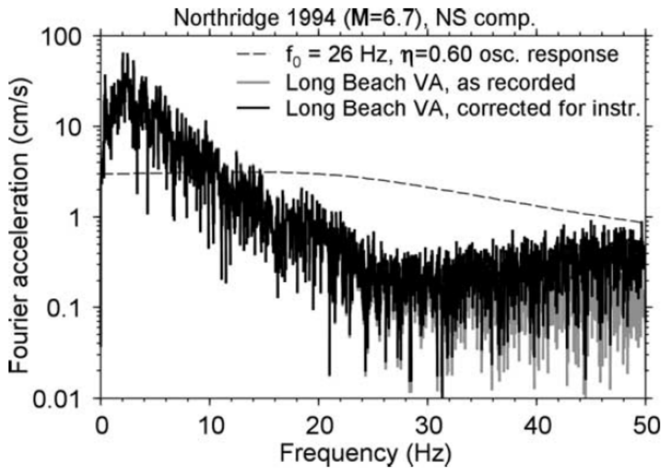
Standard vs nonstandard noise

- The spectra in the Figure were obtained with the record shown in the figure in the last slide and after the spikes were removed.
- The spikes clearly constituted a serious noise contamination at short periods but it is also noted that their elimination appears to have led to slight modifications in the spectrum at long periods (spikes are broadband and have energy content at long as well as short periods).
- If the misplacement of decimal points is identified as the cause of the errors, then an exact correction could be made.



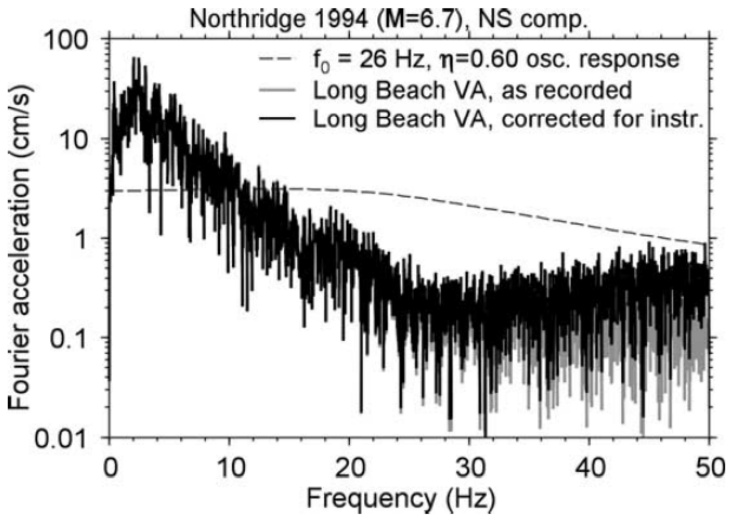
Instrument correction

- As noted earlier, the transducer frequency in analog instruments is limited to about 25 Hz, and this results in distortions of amplitudes and phases of the components of ground motion at frequencies close to or greater than that of the transducer.
- The digitization process itself can also introduce high-frequency noise as a result of the random error in the identification of the exact mid-point of the film trace as shown in the figure.



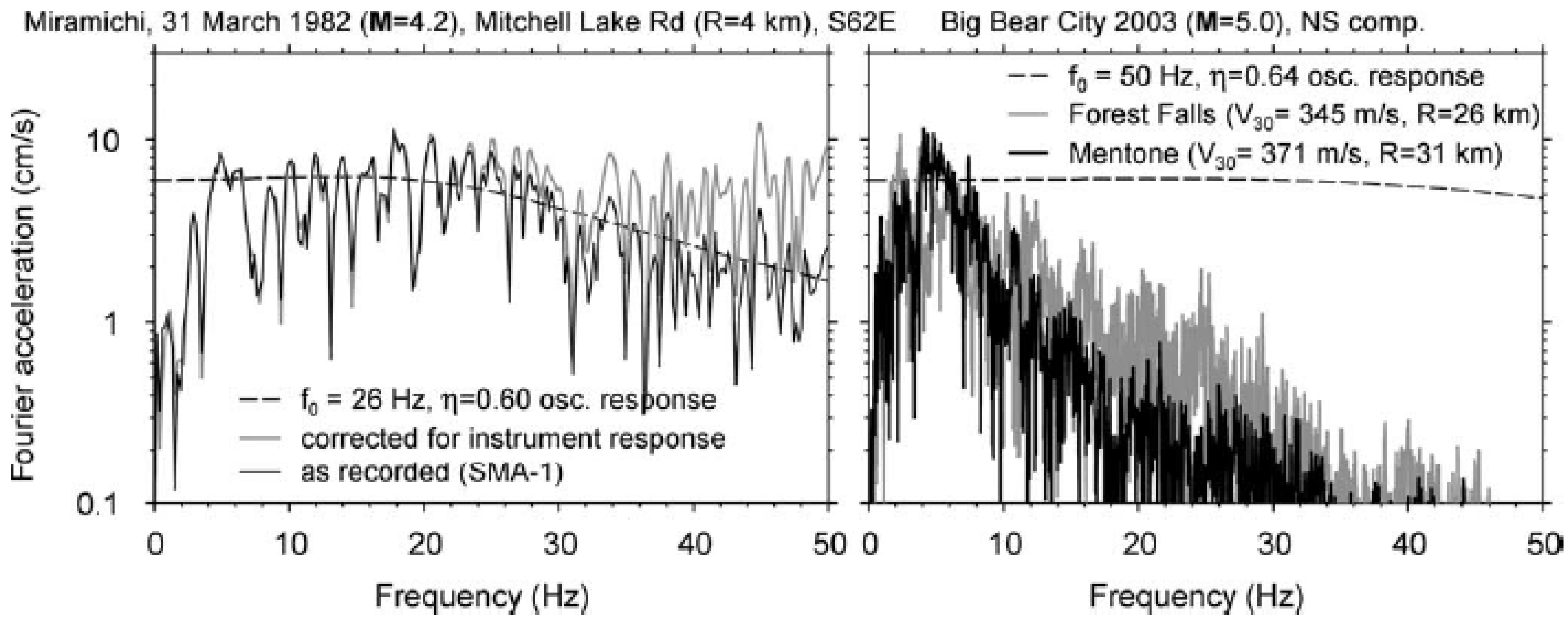
Instrument correction

- Fourier acceleration spectrum of an analog recording at a site underlain by thick sediments is shown in the figure. Natural processes along the propagation path have removed energy at frequencies much below those affected by the instrument response (see dashed line; the instrument response has been shifted vertically so as not to be obscured by the data), leading to the decreasing spectral amplitudes with increasing frequency up to about 26 Hz (coincidentally the same as the instrument frequency), at which point noise produces an increase in spectral amplitudes. Instrument correction only exacerbates the contamination of the signal by high frequency noise.



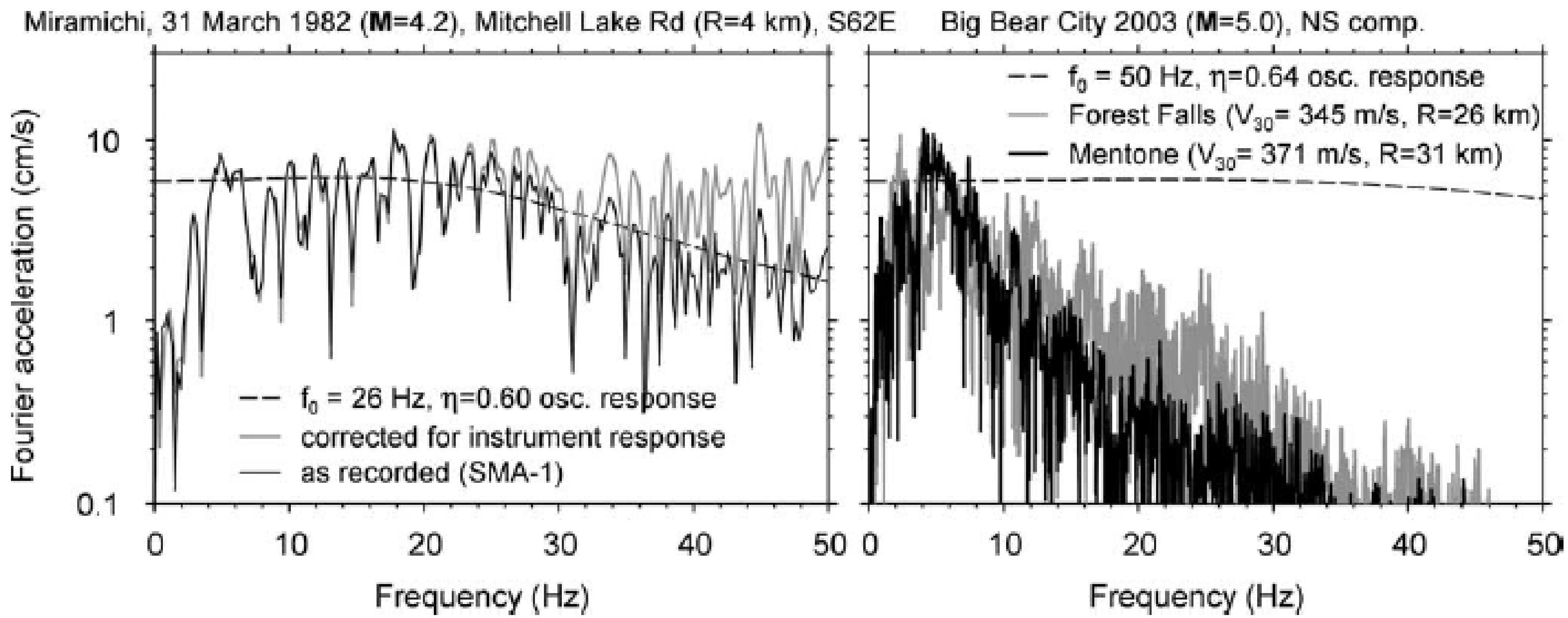
Fourier Spectra

- The left-hand plot in the figure shows an example of the Fourier spectra of high-frequency ground motion obtained at a very hard rock site in Canada at a distance of 4 km from the source of a small magnitude earthquake. **Softer sites**, even those classified as ‘rock’ such as class B in the 2003 NEHRP guidelines, will tend to **filter out such high frequency motion**.
- **Very high-frequency motions will also tend to attenuate rapidly with distance** and hence will not be observed at stations even a few tens of kilometers from the fault rupture.



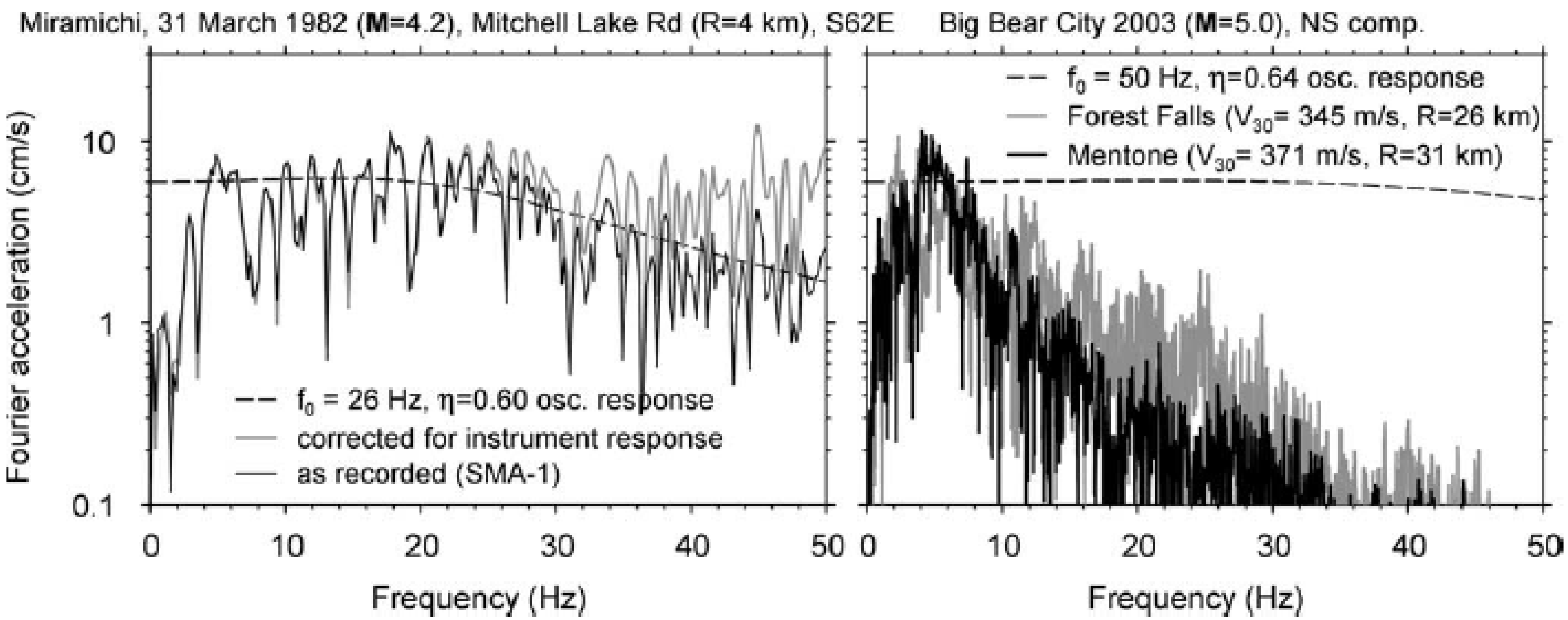
Fourier Spectra

- The figure shows the Fourier acceleration spectra of earthquakes recorded in eastern and western North America (left and right graphs, respectively). The eastern North America recording has much higher frequency content than that from western North America, even without instrument correction. The record from Miramichi was recorded on an analog instrument, whereas those from the Big Bear City earthquake were recorded on digital instruments (the response curves of the instruments are shown by the dashed lines and have been shifted vertically so as not to be obscured by the data).



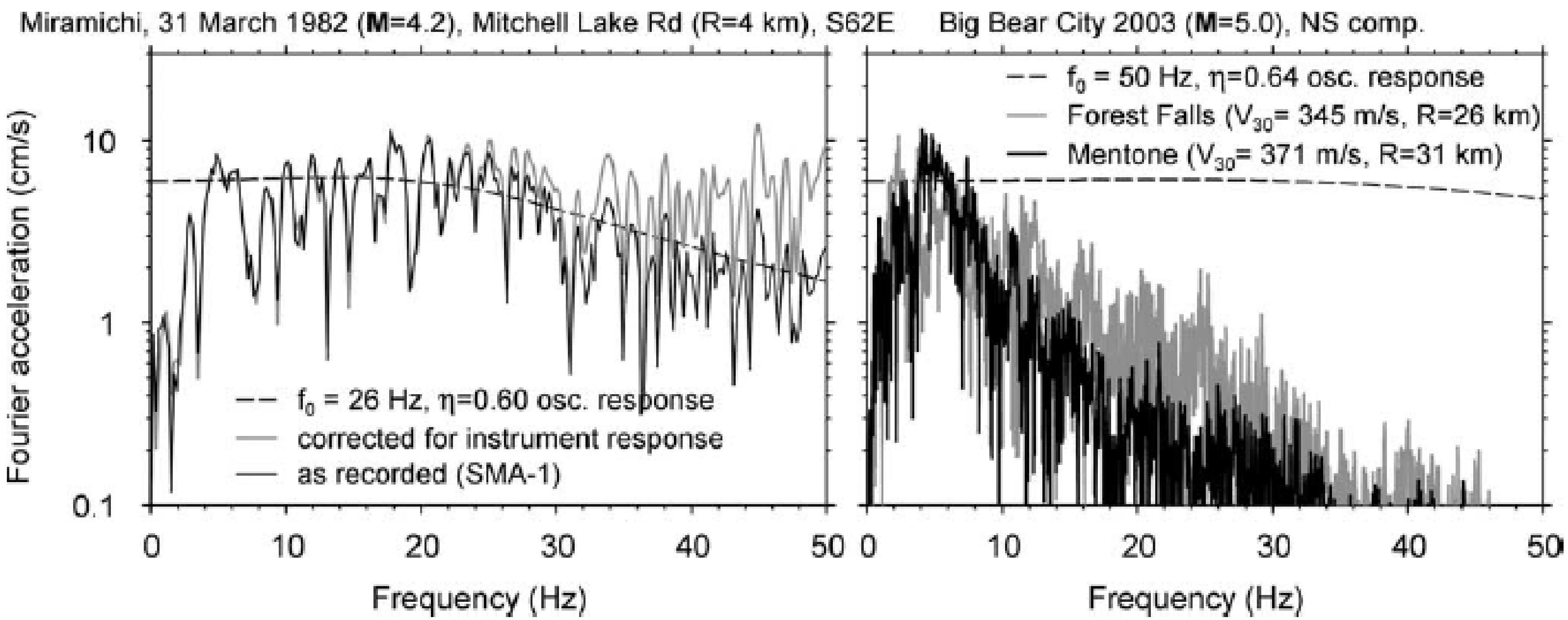
Fourier Spectra

- The plot in the figure also shows the typical transducer response for the instrument (SMA-1) on which the record was obtained, and the **effect of applying a correction for the instrument characteristics**, which is to **increase slightly the amplitudes at frequencies greater than 30 Hz**. The nature of such motions, at periods of less than 0.03 s, will only be relevant to particular engineering problems, such as the response of plant machinery and nonstructural components.



Fourier Spectra

- The right-hand plot in the figure show the Fourier spectra of more typical ground motions obtained at soil sites during a moderate magnitude earthquake in California. These records were obtained on digital instruments and are lacking in very high frequency motion mainly because of the attenuating effect of the surface geology at these sites compared to the very hard site in Canada. The plot also shows the transducer response for these digital instruments, which is almost flat to beyond 40 Hz.

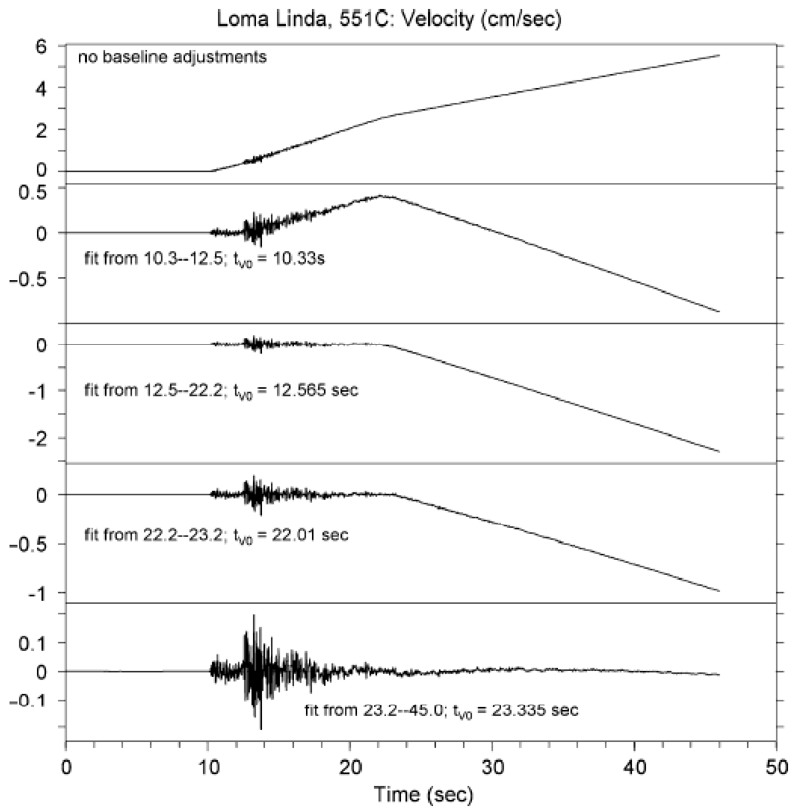


Corrections for transducer characteristics

- For **digital recordings**, **instrument corrections should not be necessary**. For analog recordings, if the engineering application is concerned with motions at frequencies above 20 Hz and the site characteristics are sufficiently stiff for appreciable amplitudes at such frequencies to be expected, a correction should be considered.
- However, it should be borne in mind that the **instrument corrections essentially amplify the high-frequency motions**; if the digitization process has introduced high-frequency noise into the record, then the **instrument correction will amplify this noise**.
- Unless there are compelling reasons for applying a correction for the instrument characteristics, we recommend that no attempt should be made to do so. The one exception to this may be the very earliest recordings obtained in the US with accelerographs that had natural frequencies of the order of 10 Hz.

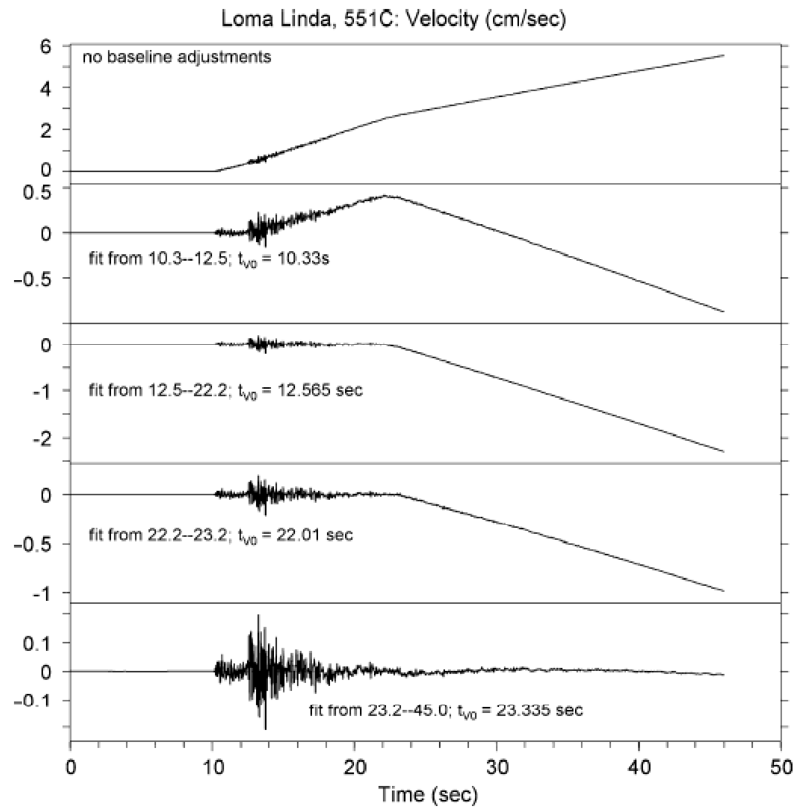
Baseline adjustments

- A major problem encountered with both analog and digital accelerograms are distortions and shifts of the reference baseline, which result in unphysical velocities and displacements.
- One approach to compensating for these problems is to use **baseline adjustments**, whereby one or more baselines, which may be straight lines or low-order polynomials, are subtracted from the acceleration trace.



Baseline adjustments

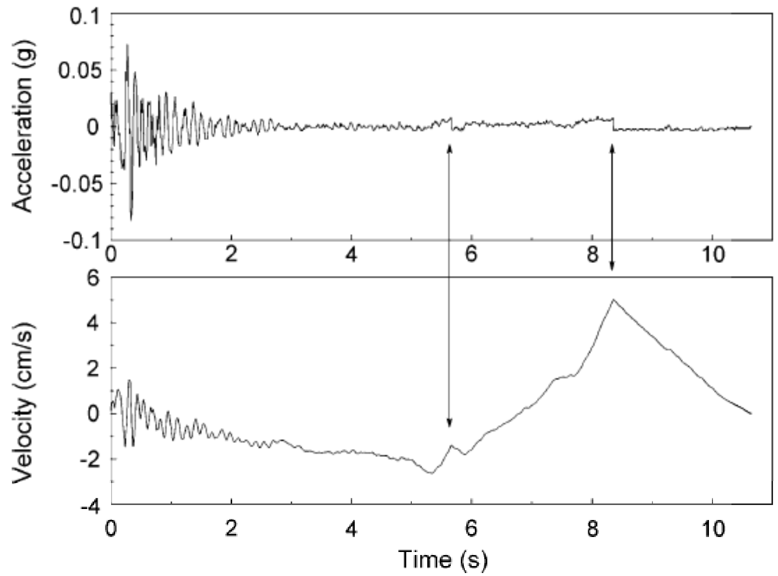
- The figure illustrates the application of a piece-wise sequential fitting of baselines to the velocity trace from a digital recording in which there are clearly identifiable offsets in the baseline.



Baseline adjustments

- A similar procedure could be applied directly to the acceleration time-history to correct for the type of baseline shifts shown in the figure.

- The figure shows NS component of the 21 May 1979 Italian earthquake (12:36:41 UTC) recorded at Nocera Umbra, showing shifts in the baseline at 5.6 and 8.3 s.

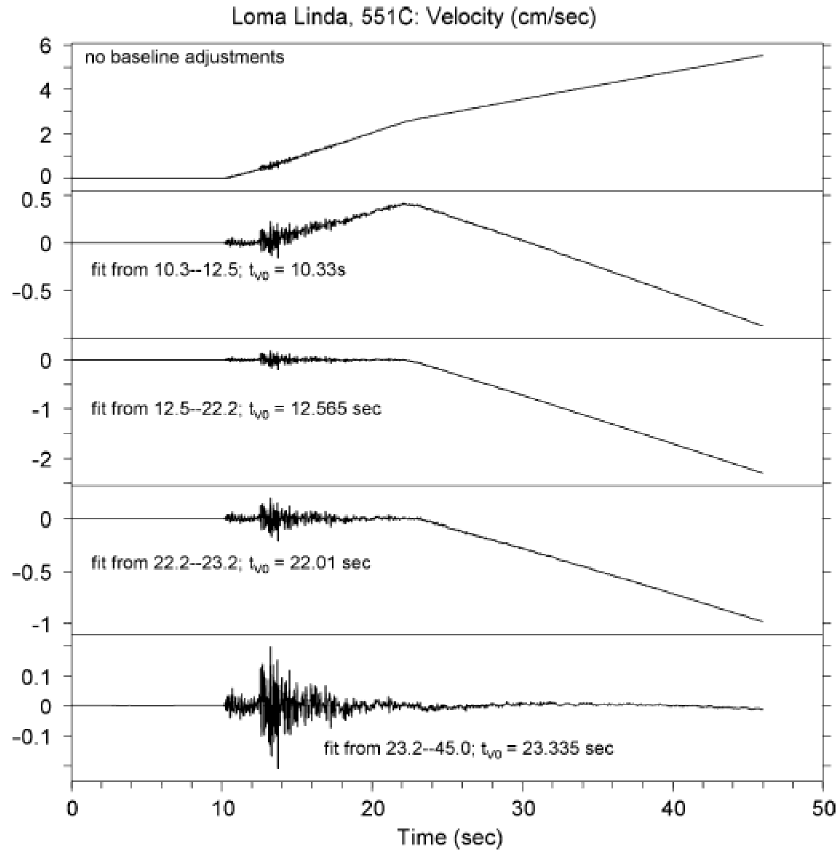


Baseline adjustments

- The procedure applied in the figure is to identify (by blowing up the image) sections of the velocity that appear to have a straight baseline, and then fitting a straight line to this interval.

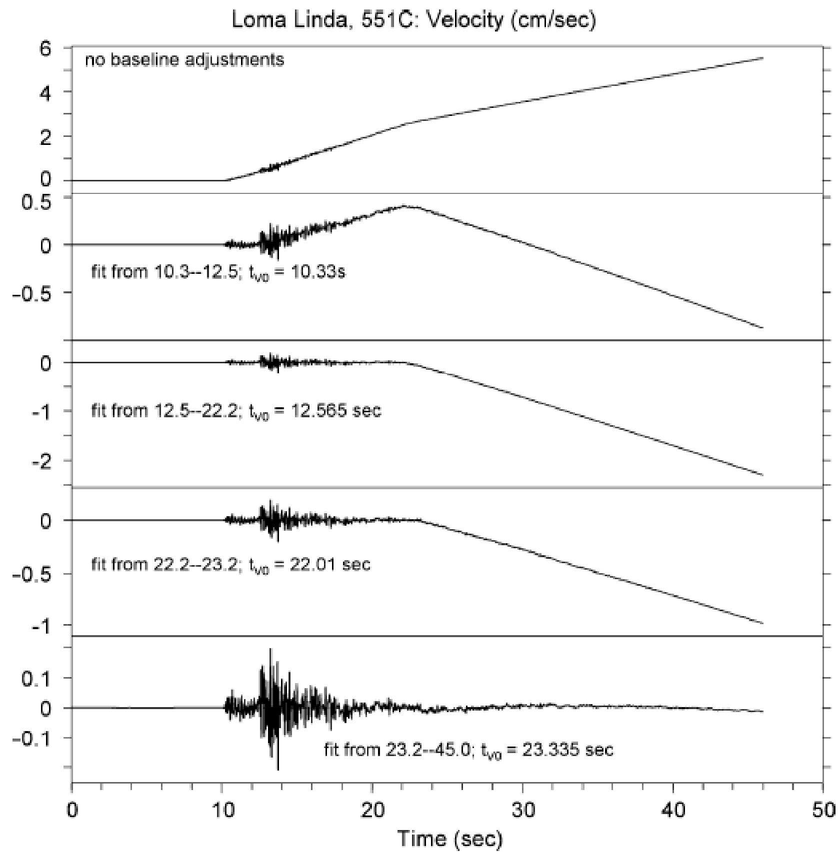
- This line in effect is then subtracted from the velocity trace, but in practice it is necessary to apply the adjustment to the accelerations.

- The adjustment to the acceleration is a simple shift equal to the **gradient (i.e. the derivative) of the baseline on the velocity**; this shift is applied at a time t_{v0}, which is the time at which the line fit to the velocity crosses the zero axis.



Baseline adjustments

- The adjusted velocity trace is then inspected to identify the next straight line segment, which is fit in the same way.
- In the particular case illustrated in the figure, a total of four line segments were required to remove the most severe distortions of the baseline visible in uppermost plot, although the baseline instabilities are not entirely removed, as evident in the residual long-period trends.



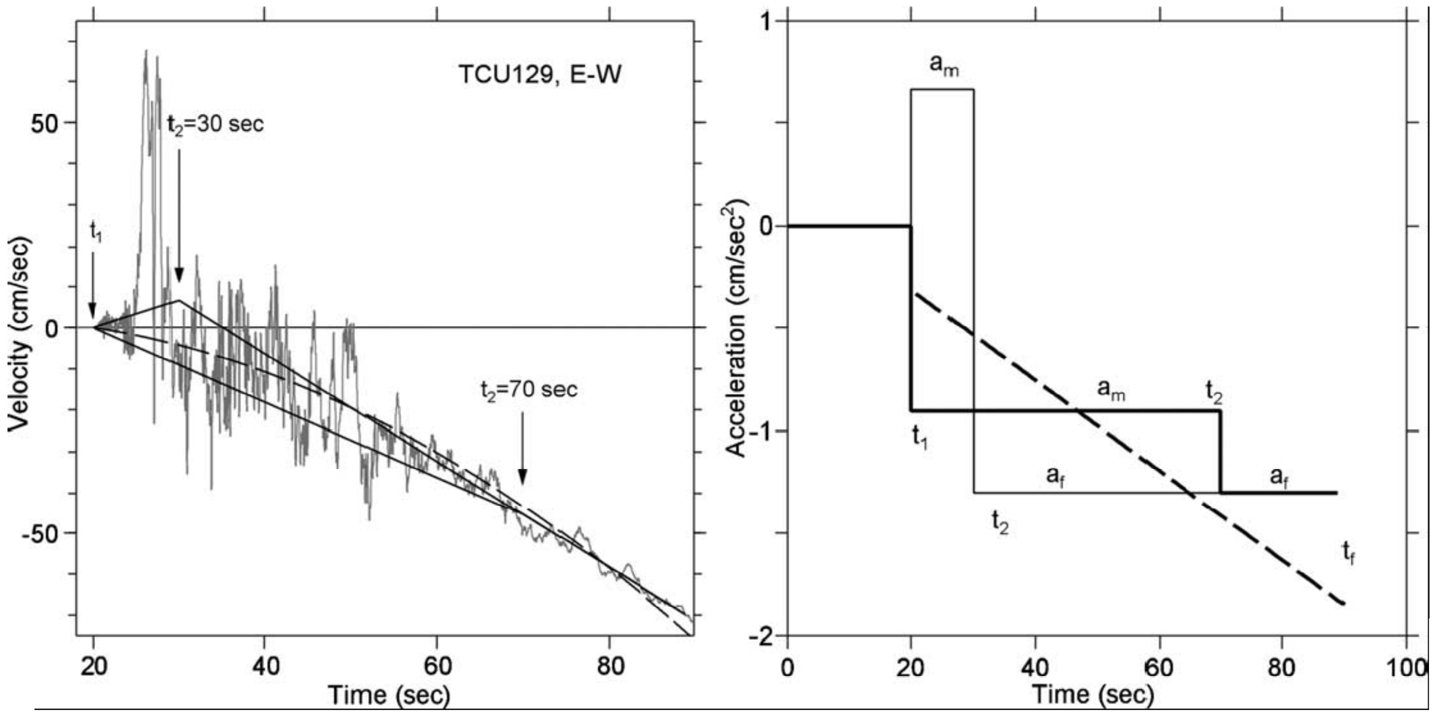
Baseline adjustments

- The distortion of the baseline encountered in digitized analog accelerograms is generally interpreted as being the result of long-period noise combined with the signal.

- Baselines can be used as a tool to remove at least part of this noise—and probably some of the signal with it—as a means of recovering more physically plausible velocities and displacements. There are many procedures that can be applied to fit the baselines, including polynomials of different orders. A point that is worth making clearly is that, in effect, **baseline adjustments** are **low-cut filters of unknown frequency** characteristics.

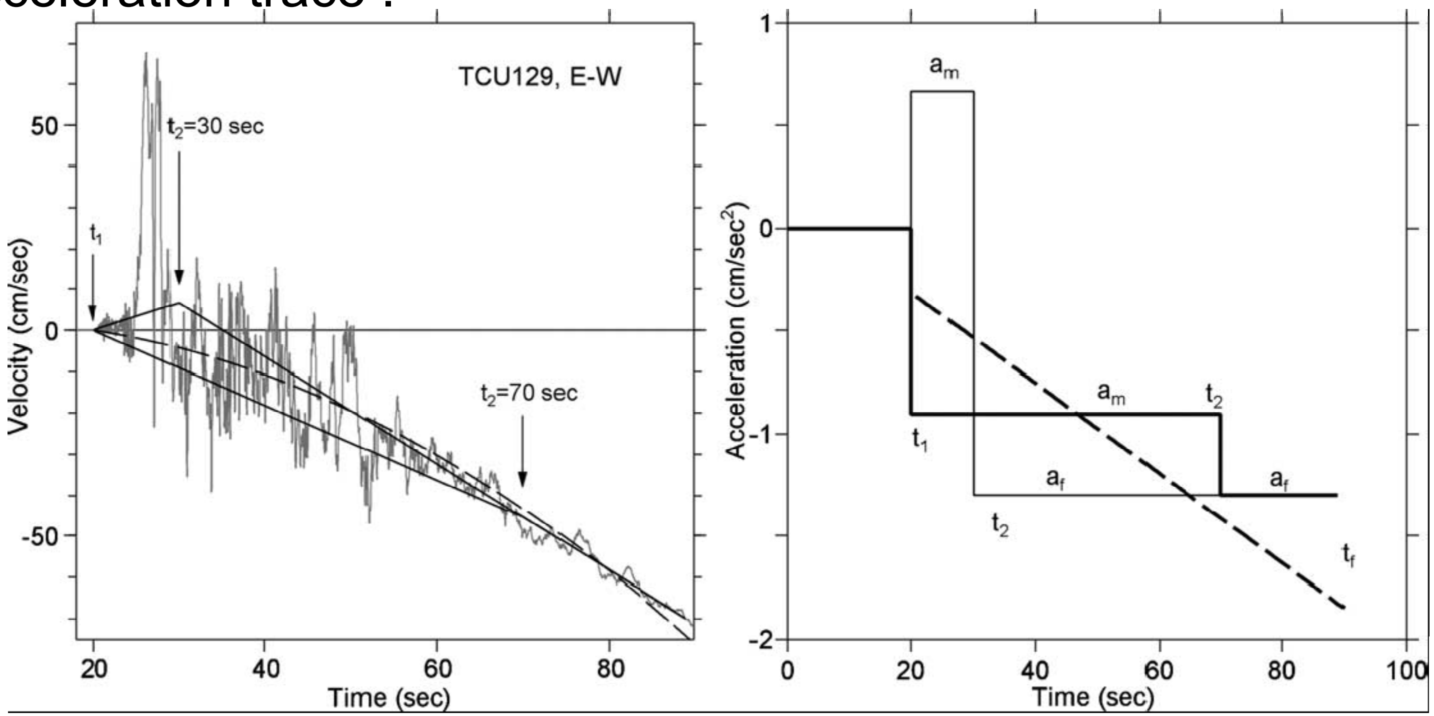
Baseline adjustments

- The figure on the left: Shaded line: velocity from integration of the east–west component of acceleration recorded at TCU129, 1.9 km from the surface trace of the fault, from the 1999 Chi-Chi earthquake, after removal of the pre-event mean from the whole record. A least-squares line is fit to the velocity from 65 s to the end of the record. Various baseline corrections using the Iwan et al. (1985) scheme are obtained by connecting the assumed time of **zero velocity t_1** to the fitted velocity line at time t_2. Two values of t_2 are shown: 30, and 70 s.



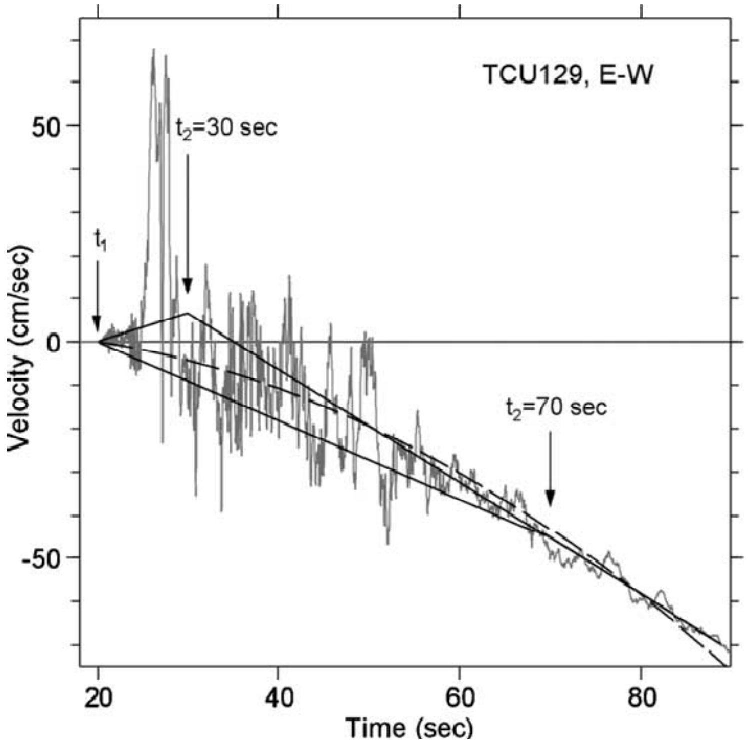
Baseline adjustments

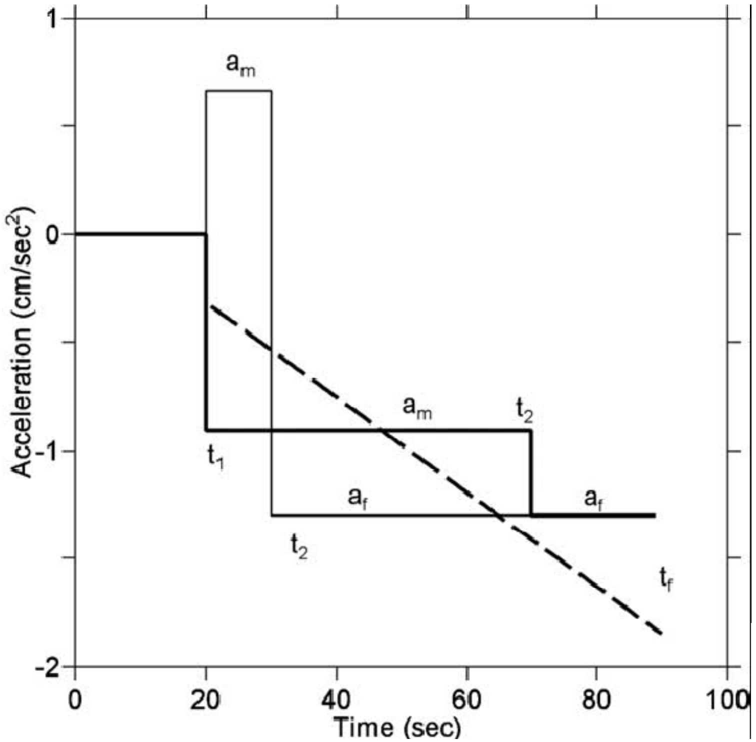
- The dashed line is the quadratic fit to the velocities, with the constraint that it is 0.0 at t=20 s.
- The acceleration time series are obtained from a force-balance transducer with natural frequency exceeding 50 Hz, digitized using 16.7 counts/cm/s^2 (16,384 counts/g). Right: The derivatives of the lines fit to the velocity are the baseline corrections applied to the acceleration trace .



Baseline adjustments

- The line fit approach is the more complex scheme proposed by Iwan et al. The method was motivated by studies of a specific instrument for which the baseline shifted during strong shaking due to hysteresis; the accumulation of these baseline shifts led to a velocity trace with a linear trend after cessation of the strong shaking. The correction procedure approximates the complex set of baseline shifts with two shifts, one between times of t_1 and t_2, and one after time t_2. The velocity will oscillate around zero (a physical constraint), but the scheme requires selection of the times t_1 and t_2.



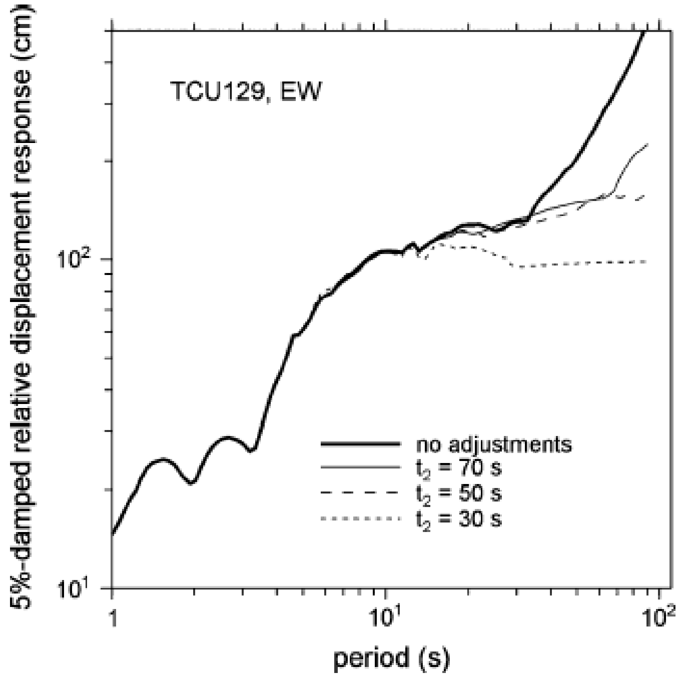


Baseline adjustments

- Figure shows the response spectra of the east–west component of acceleration recorded at TCU129 from the 1999 Chi-Chi, Taiwan, earthquake, modified using a variety of baseline corrections.

- Without a physical reason for choosing these times (for example, based on a knowledge of a specific instrument), the choices of t_1 and t_2 become subjective.

- Figure shows that the long-period response spectrum ordinates are sensitive to the choice of t_2 (t_1 was not varied in this illustration).

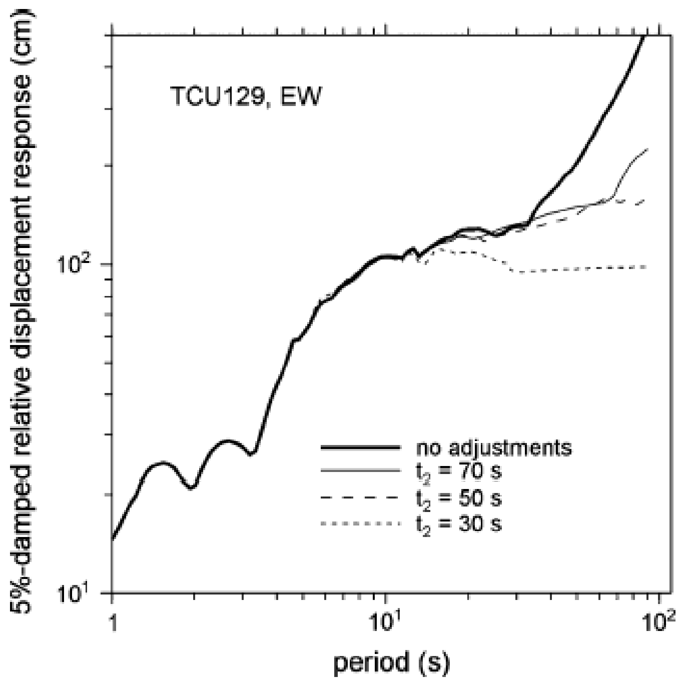


- It is important to note that for this particular accelerogram the differences in the response spectrum are not significant until beyond 10 s oscillator period).

Baseline adjustments

- A commonly used simplification of the generalized Iwan et al. method is to assume that $t_1=t_2$, with the time given by the zero intercept of a line fit to the later part of the velocity trace.

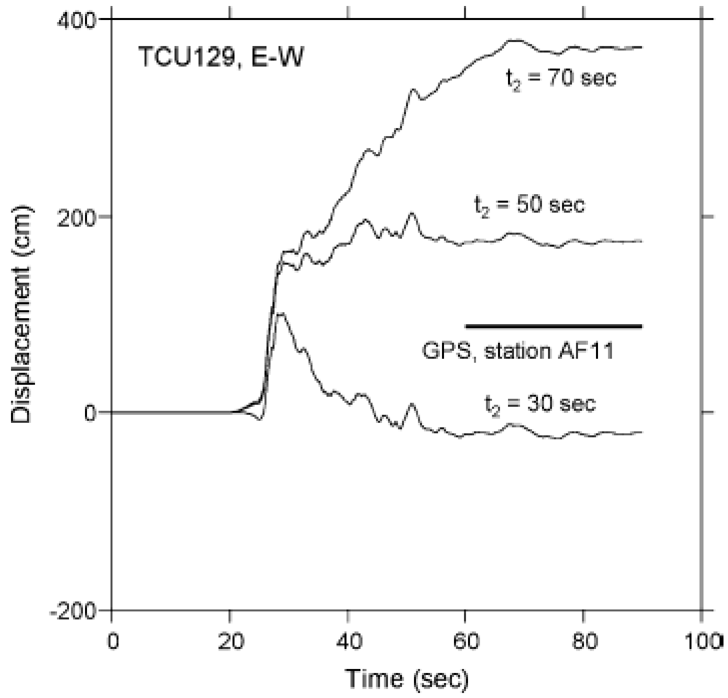
- This corresponds to the assumption that there was only one baseline offset and that it occurred at a single time (for many records this seems to be a reasonable assumption). We call this simplification the v_0 correction.



Residual displacements

- One of the possible advantages of baseline fitting techniques just discussed is that the displacement trace can obtain a constant level at the end of the motion and can have the appearance of the residual displacement expected in the vicinity of faults as shown in the figure.

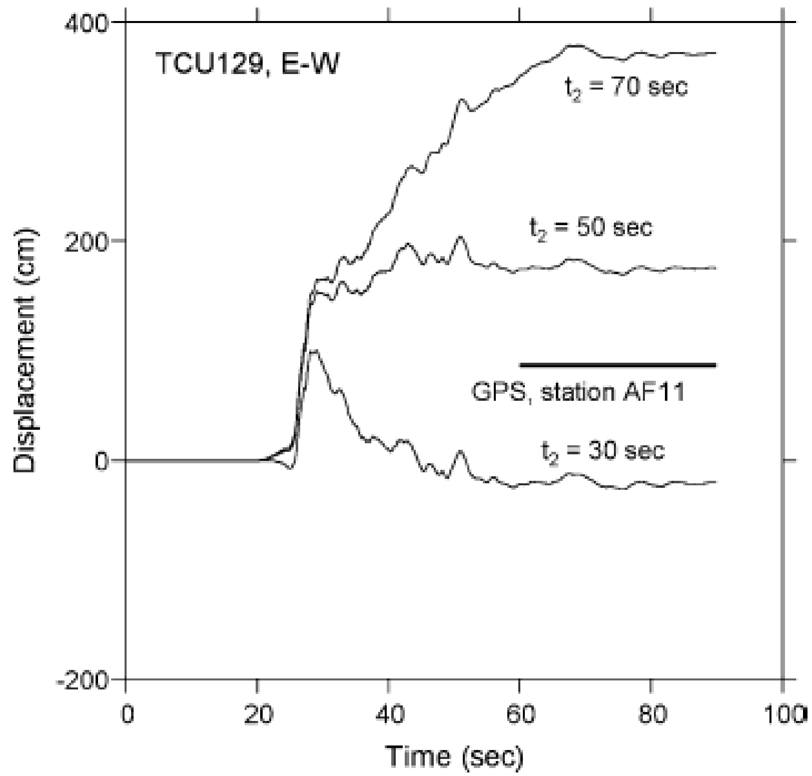
- This character of the displacement record cannot be achieved using low-cut filters.



Residual displacements

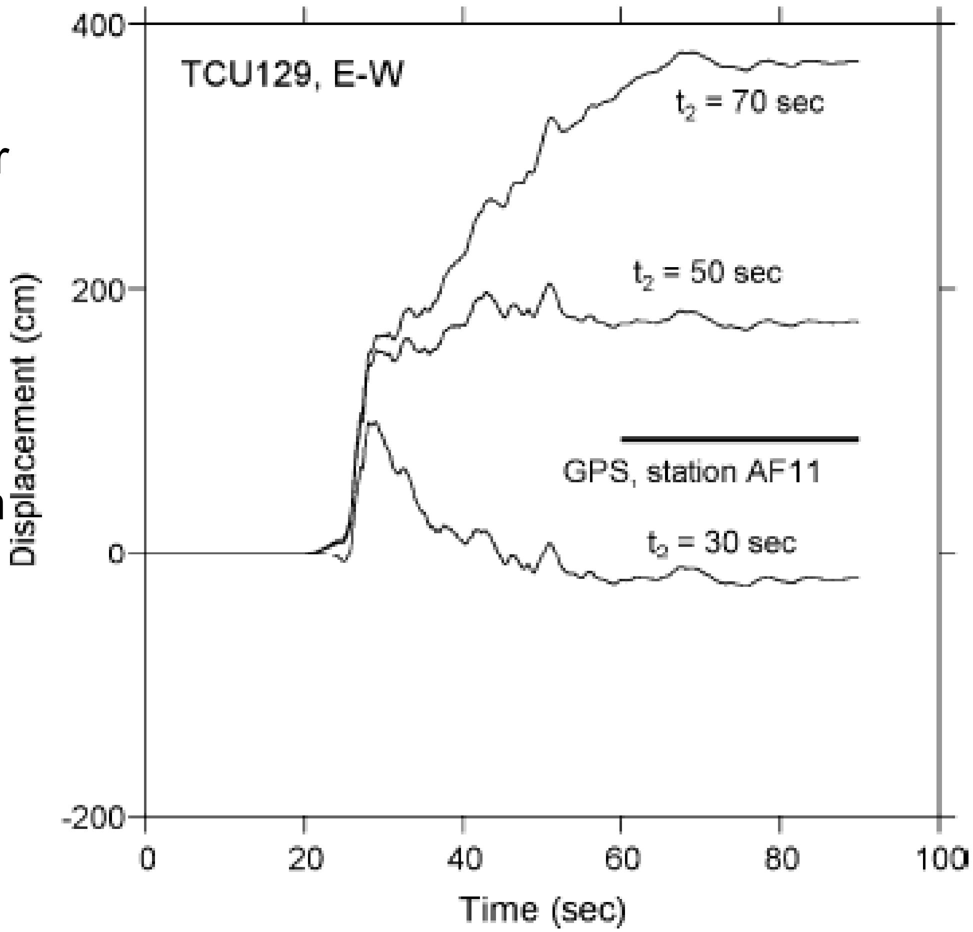
- At the end of the ground shaking caused by an earthquake, the ground velocity must return to zero, and this is indeed a criterion by which to judge the efficacy of the record processing.

- The final displacement, however, need not be zero since the ground can undergo permanent deformation either through the plastic response of near-surface materials or through the elastic deformation of the earth due to co-seismic slip on the fault.



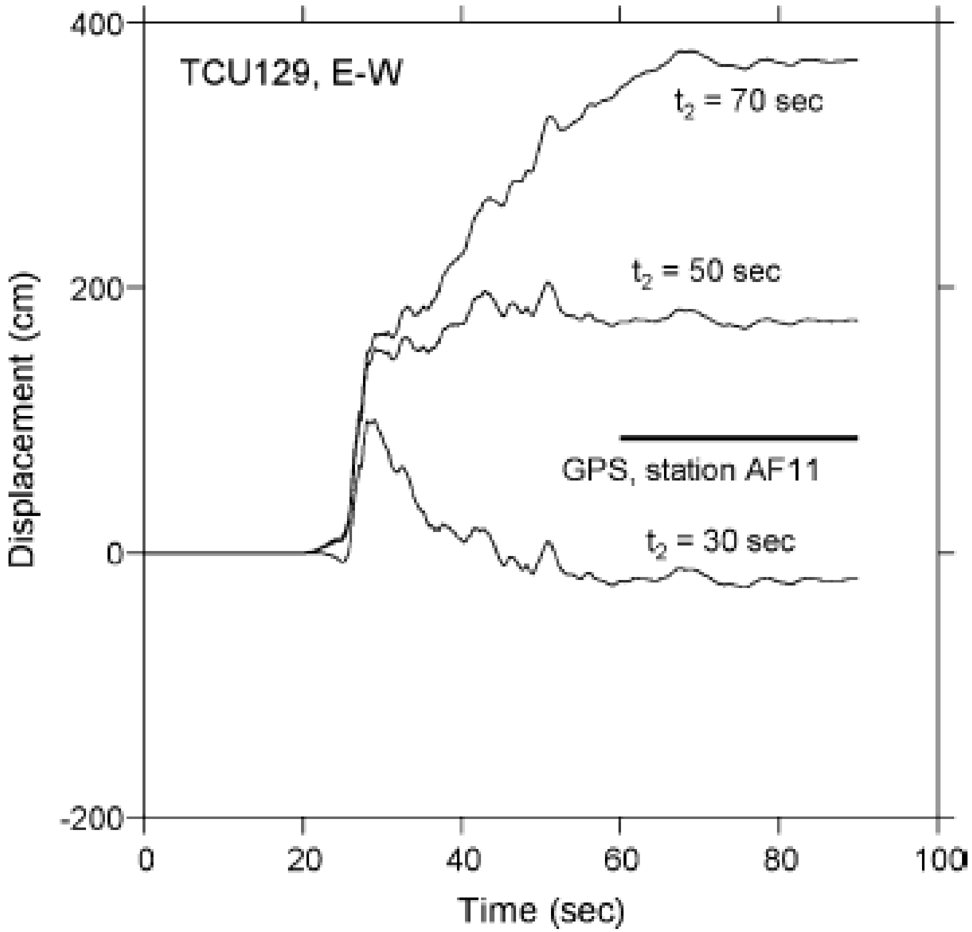
Residual displacements

- Close to the fault rupture of large magnitude earthquakes (~M_w = 6.5 and above) this residual displacement can be on the order of tens or hundreds of centimeters.
- This can become an important design consideration for engineered structures that cross the trace of active faults, cases in point being the Trans Alaskan Pipeline System and the Bolu viaduct in Turkey, the former being traversed by the fault rupture of the November 2002 Denali earthquake, the latter by the rupture associated with the November 1999 Duzce earthquake.



Residual displacements

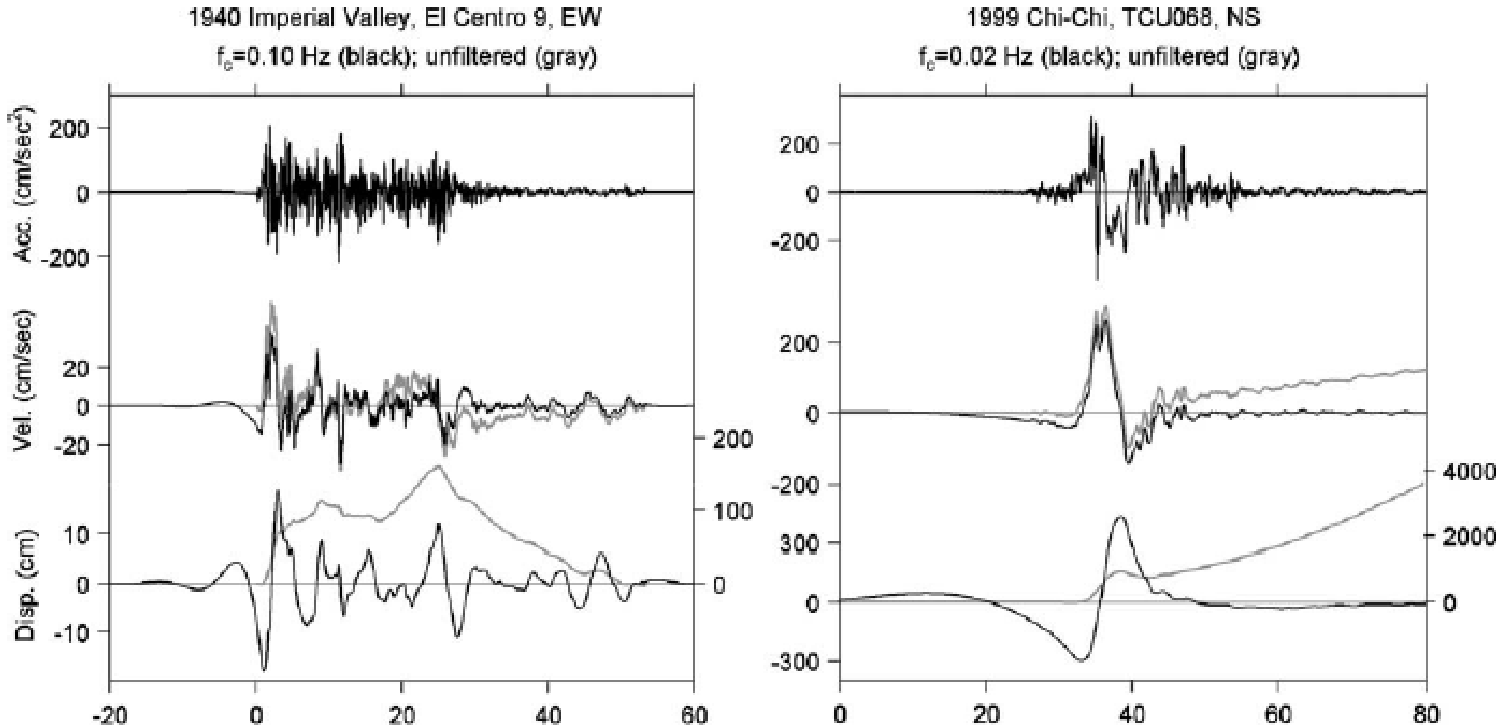
- The problem presented by trying to recover the residual placement through baseline fitting is that the resulting offset can be highly sensitive to the choice of parameters as shown in the figure.
- Furthermore there are few data with independently measured offsets exactly at the location of strong-motion instruments.
- The lack of independently-measured offsets is beginning to be overcome with the installation of continuous GPS stations sampling at sufficiently high rates colocated with accelerographs.



High pass filters

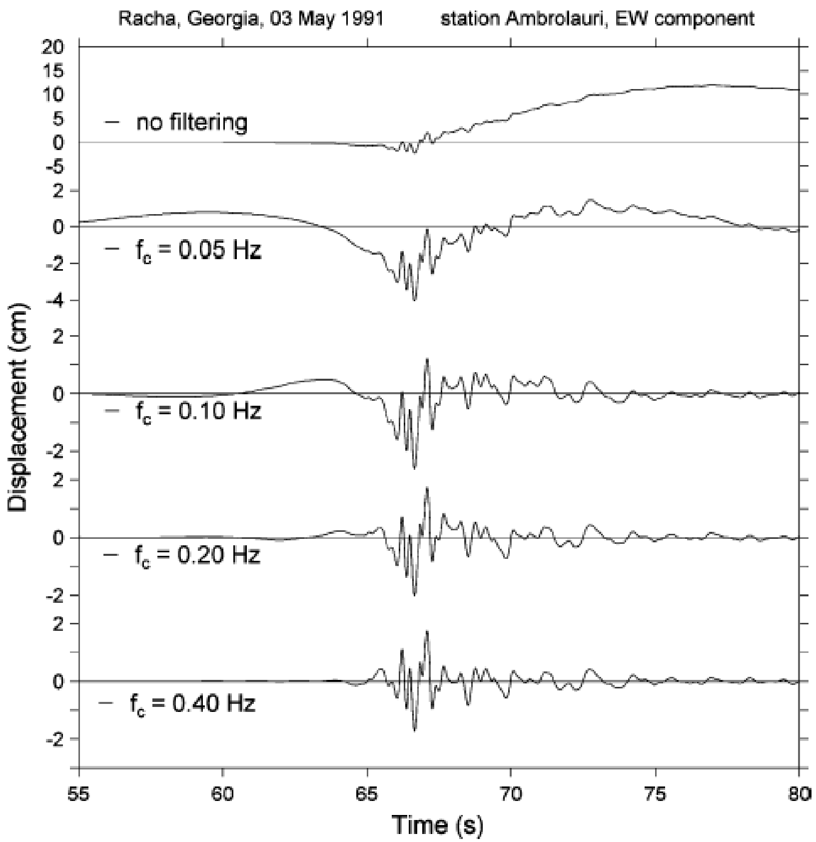
- The most widely used—and also the most effective and least subjective—tool for reducing the long-period noise in accelerograms is the low-cut filter. The figure shows the accelerograms after the application of filters to the acceleration time-history, and the improvement in the appearance of velocity and displacement time histories is obvious.

- It should also be noted that there is little discernable difference between the filtered and unfiltered accelerations.



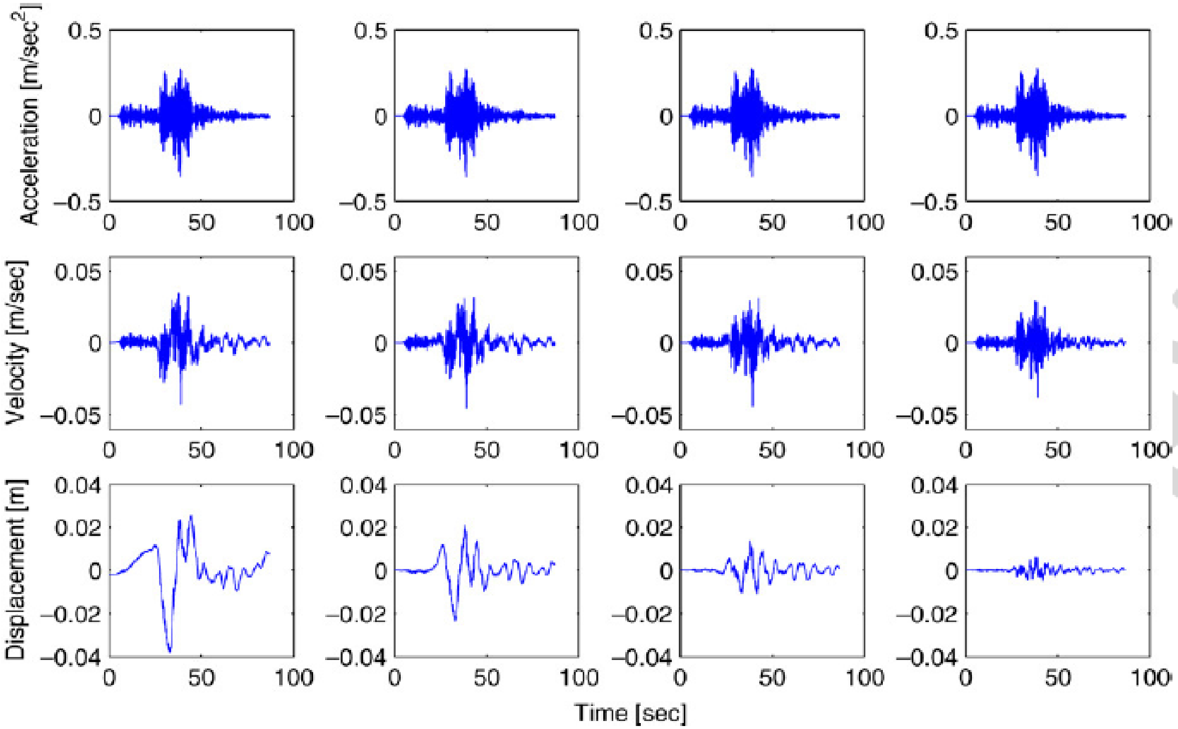
High pass filters

- Although the benefits of applying filters are clear, it is important to be aware of the sensitivity of the results obtained to the actual parameters selected for the filter.



High pass filters

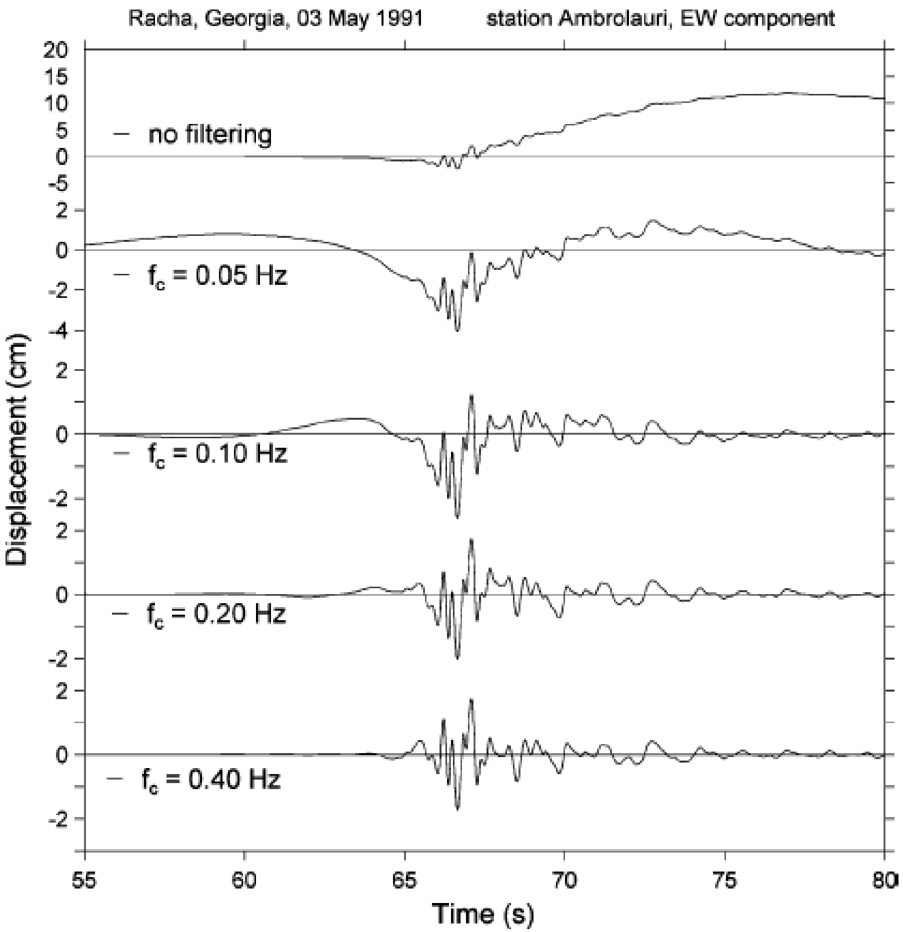
- NS component of the FAT record of the Duzce earthquake: displacement and velocity time history obtained by using acausal high pass Butterworth filters with different corner frequencies (from the left column to the right: fc = 0.02 Hz, fc = 0.05 Hz, fc = 0.1 Hz, and fc = 0.2 Hz).



High pass filters

- The reason that the filters are described as acausal is that to achieve the zero phase shift they need to start to act prior to the beginning of the record, which can be accomplished by adding lines of data points of zero amplitude, known as pads, before the start of the record and after the end of the record.

- The length of the pads depends on the filter frequency and the filter order.



High pass filters

- Pad length proposed by Brady:

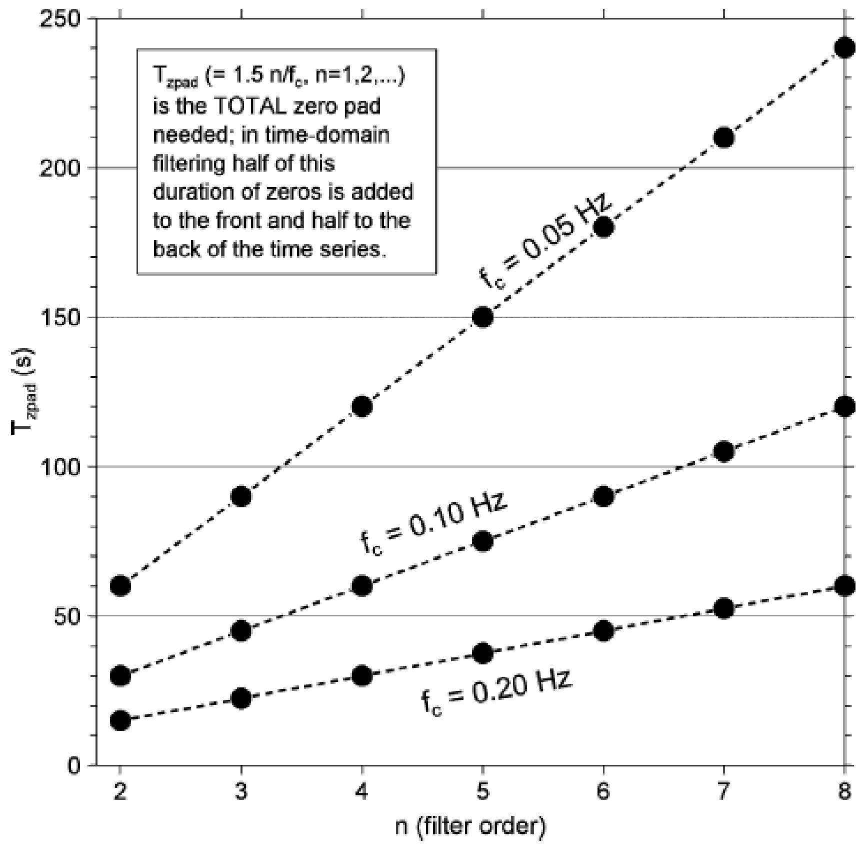
$$t_p = 1.5\frac{n_{roll}}{f_c}$$

- The required length of the filter pads will often exceed the usual lengths of pre and post-event memory on digital recordings, hence it is not sufficient to rely on the memory to act as the pads.

Zero pads

- The figure shows the total length of the time-domain zero pad recommended by Converse and Brady to allow for the filter response in 2-pass (acausal), nth-order Butterworth filters (these pads are needed regardless of whether the filtering is done in the time- or frequency-domain).

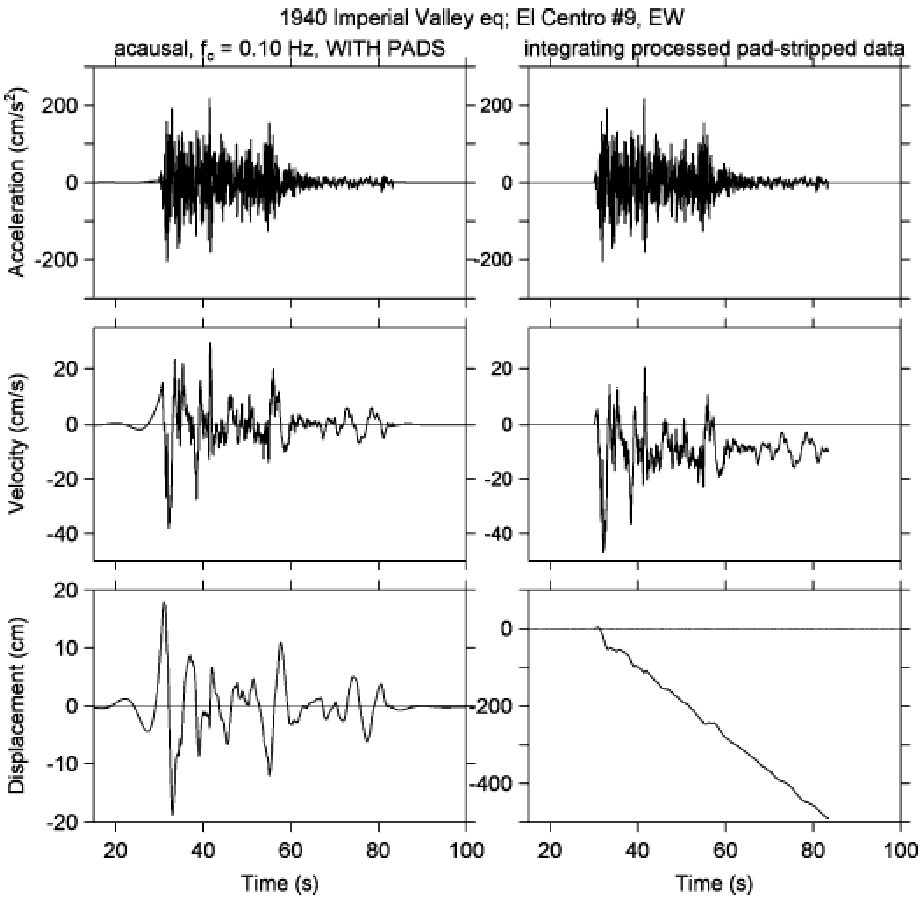
- Pre- or post-event data count as part of the required pad length. Shown are the pad lengths for three values of the filter corner frequency, as a function of filter order.



Zero pads

- One of the causes for data incompatibility for the records disseminated by the Strong-motion processing centers is the removal of the pads that are added for the application of the filter.

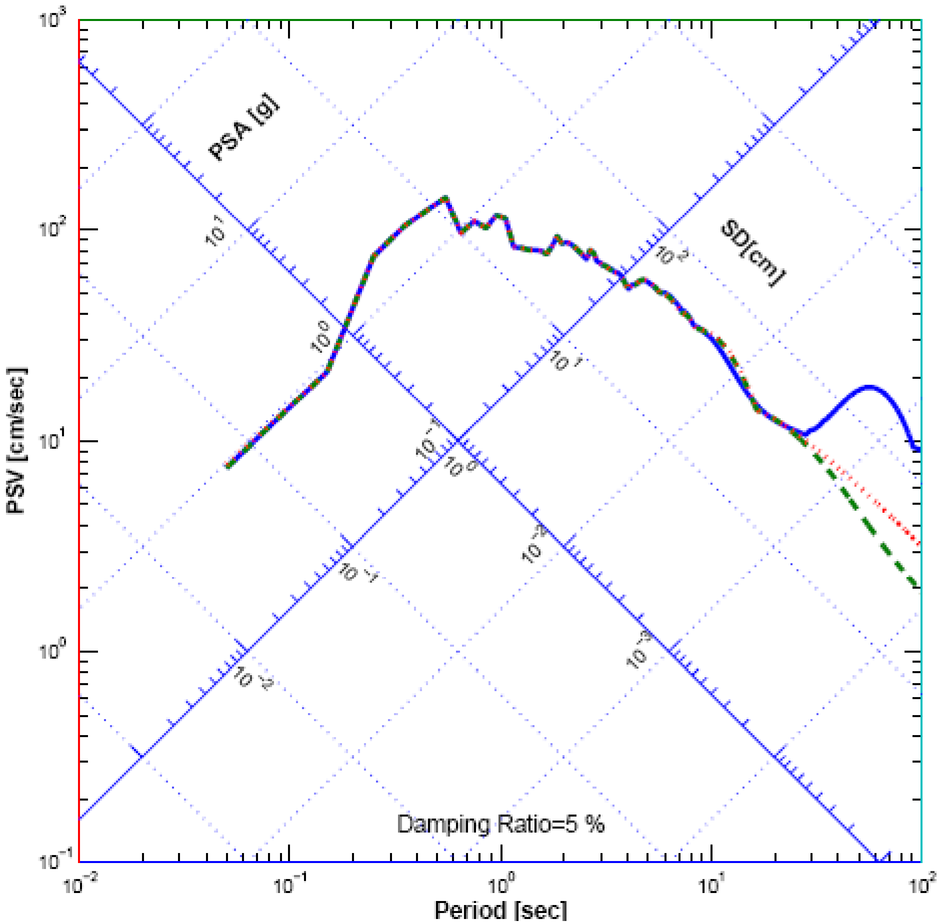
- This is an issue that creates some controversy because some argue that the pads are artificial and therefore do not constitute part of the data and hence should be removed. The consequence of their removal, however, is to undermine the effect of the filter and this can result in offsets and trends in the baselines of the velocity and displacements obtained by integration.



Zero pads

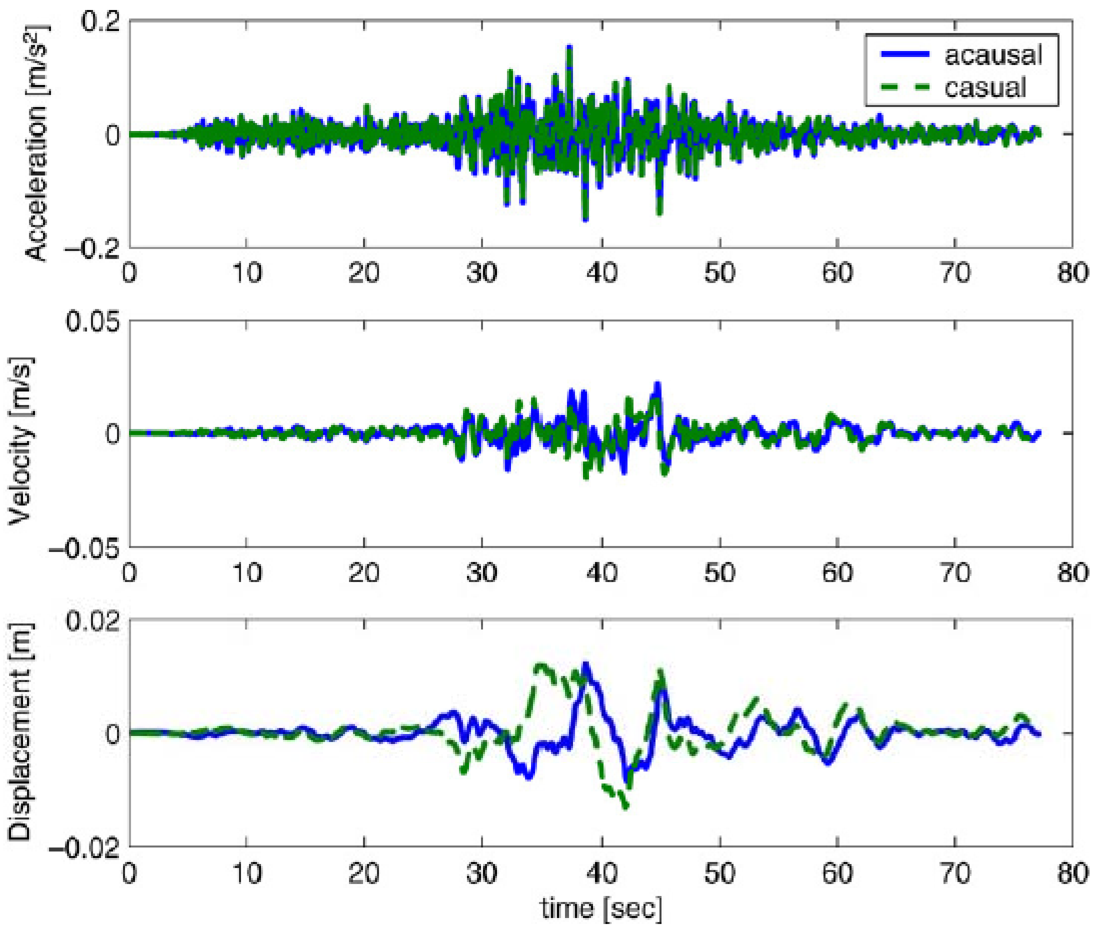
- The removal of the pads also has an influence on the long period response spectral ordinates as shown in the figure (with pads (dashed line), without pads (solid line)).

- For this reason, it is recommended that when acausal filters are used, sufficient lengths of zero pads should be added to the records and these pads should not be stripped out from the filtered data.



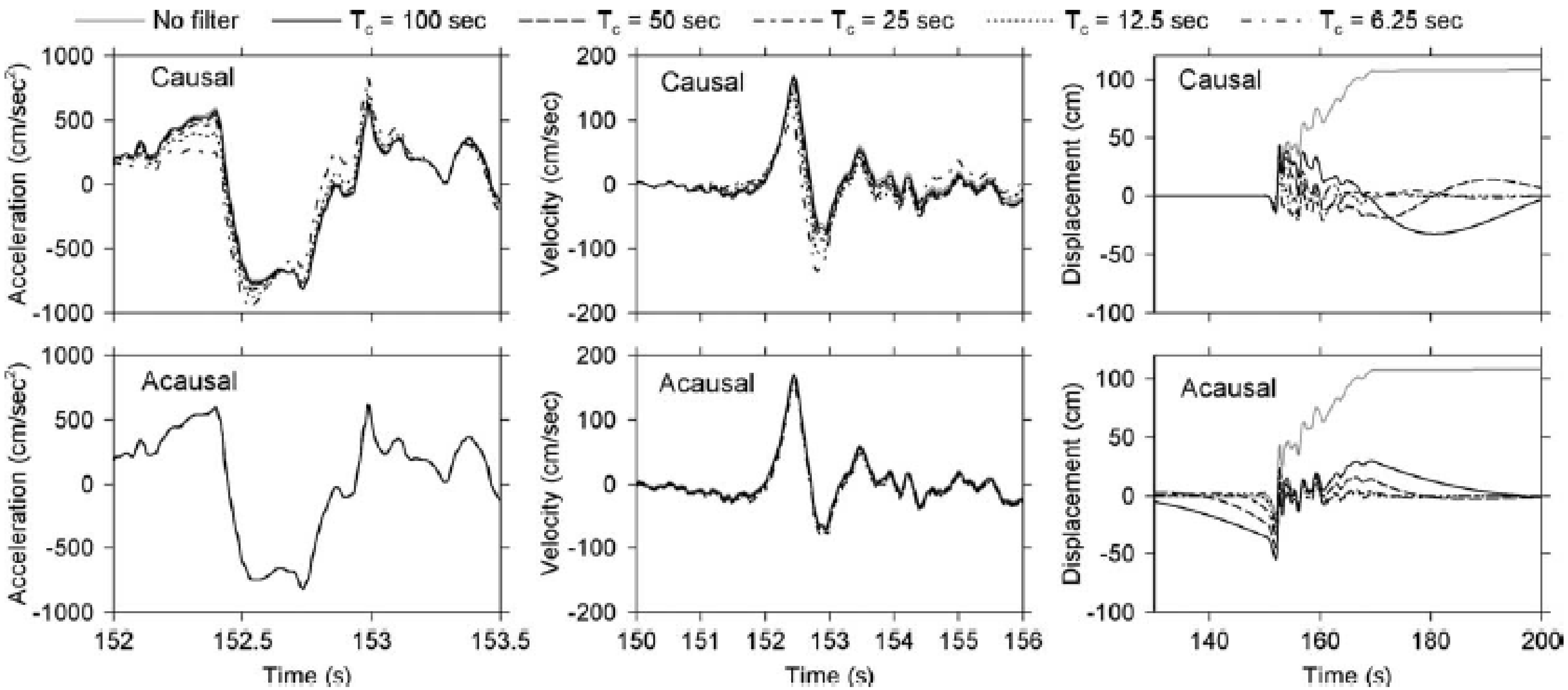
Causal-acausal Butterworth filters

- NS component of the CNA record of the Duzce earthquake processed by SPIDER using causal (dashed line) and acausal (solid line) filters.



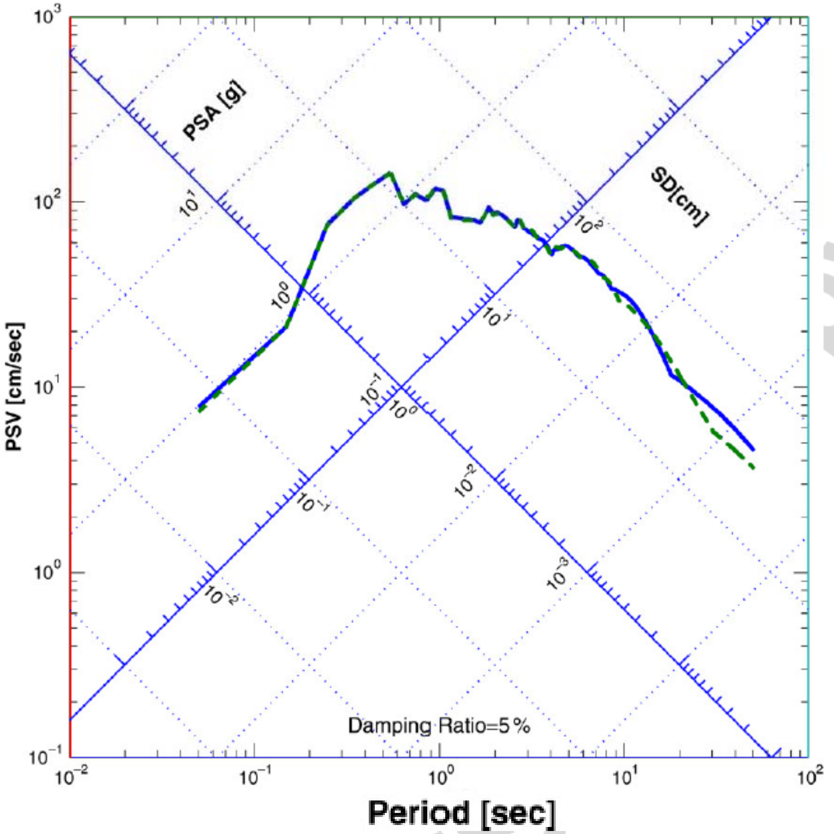
Causal-acausal Butterworth filters

- The application of causal and acausal filters, even with very similar filter parameters (the transfer functions will not be identical if time-domain filtering is used, since the causal filter will have a value of $1/\sqrt{2}$ at the filter corner frequency, f_c, whereas the acausal filter will have a value of 0.5, regardless of the filter order), have been shown to produce very different results in terms of the integrated displacements



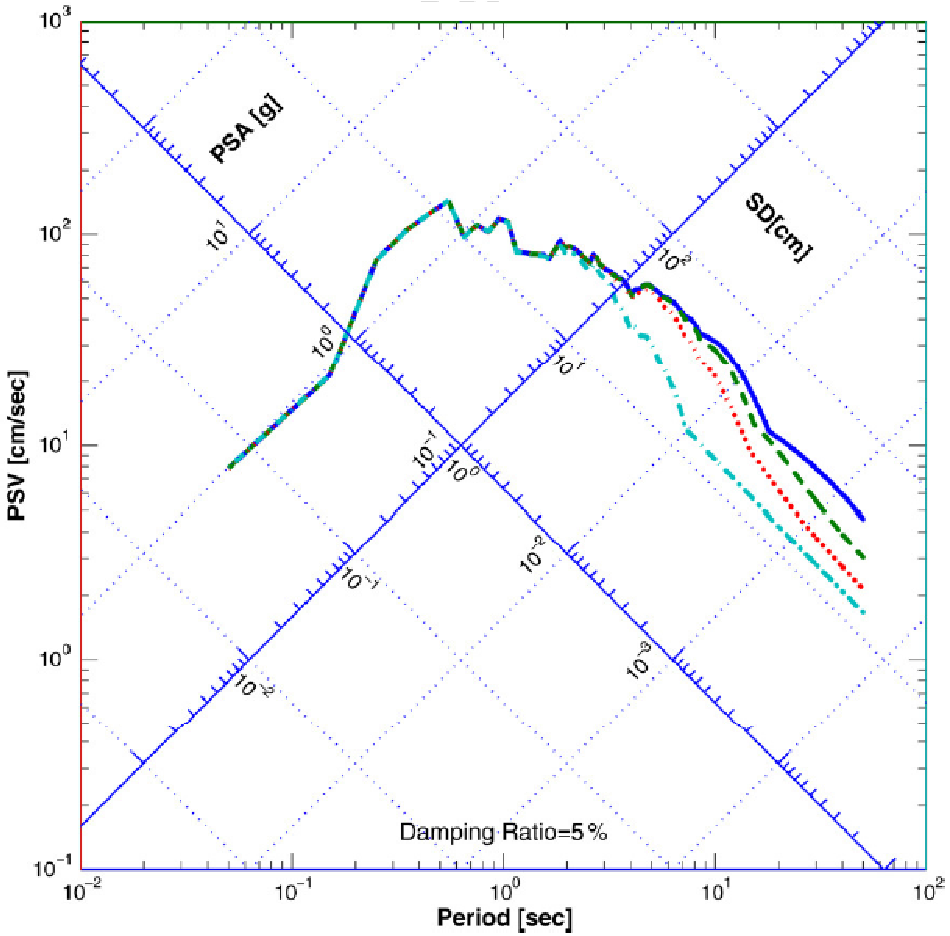
Causal vs Acausal Filters

- Response spectra (damping ratio = 5%) of the FP component of the Bolu record of the **Duzce earthquake** processed by acausal (solid line) and causal filters (dashed line).



Causal vs Acausal Filters

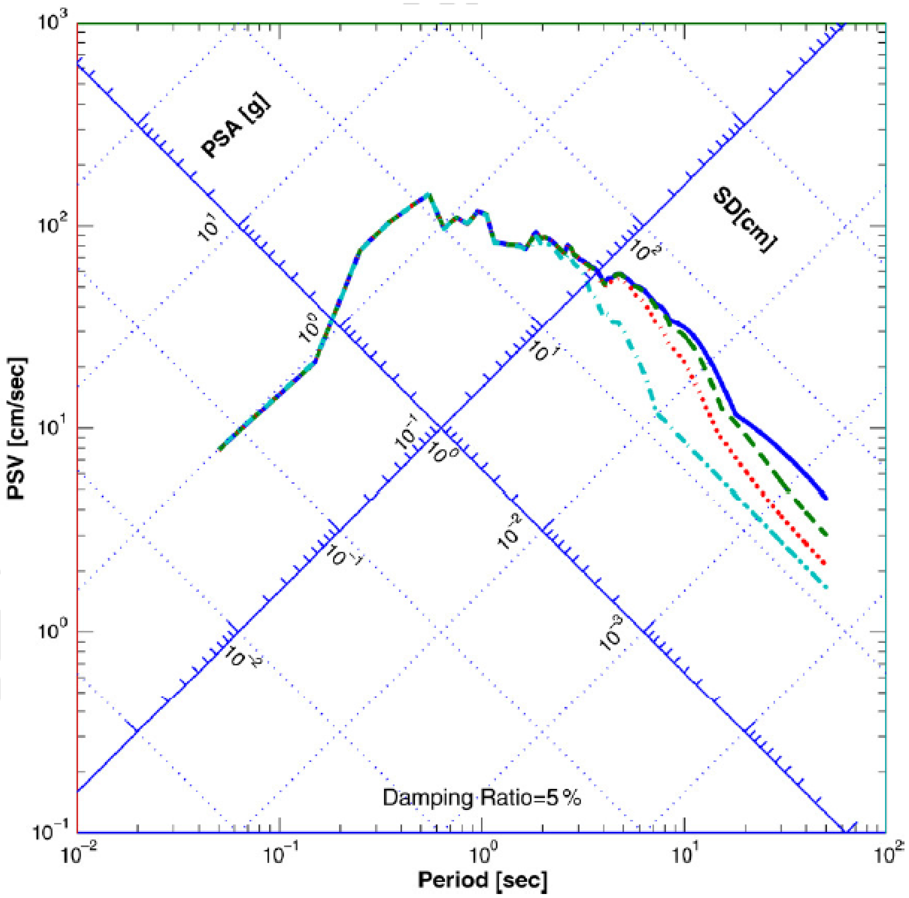
- Response spectra (damping ratio = 5%) for the FP component of the Bolu record of the 1999 Duzce, Turkey earthquake processed by acausal filters with different corner frequencies: f_c = 0.025 Hz (solid line), f_c = 0.06 Hz (dashed line), f_c = 0.1 Hz (dotted line), and f_c = 0.2 Hz (dashed–dotted line).



Causal vs Acausal Filters

- The influence of causal and acausal filters on both elastic and inelastic response spectra has been investigated.

- It is found that both elastic response spectra and inelastic response spectra computed from causally-filtered accelerations can be sensitive to the choice of filter corner periods even for oscillator periods much shorter than the filter corner periods.



Tapers

- When adding zero pads to accelerograms prior to filtering, a potential undesired consequence is to create abrupt jumps where the pads abut the record, which can introduce ringing in the filtered record.
- There are two different ways to avoid this, one being to use tapers such as a **half-cosine function** for the transition from the motion to the zero pad.
- A simpler procedure is to start the pad from the **first zero crossing within the record**, provided that this does not result in the loss of a significant portion of record, as can happen if the beginning or end of the acceleration time series is completely above or below zero.

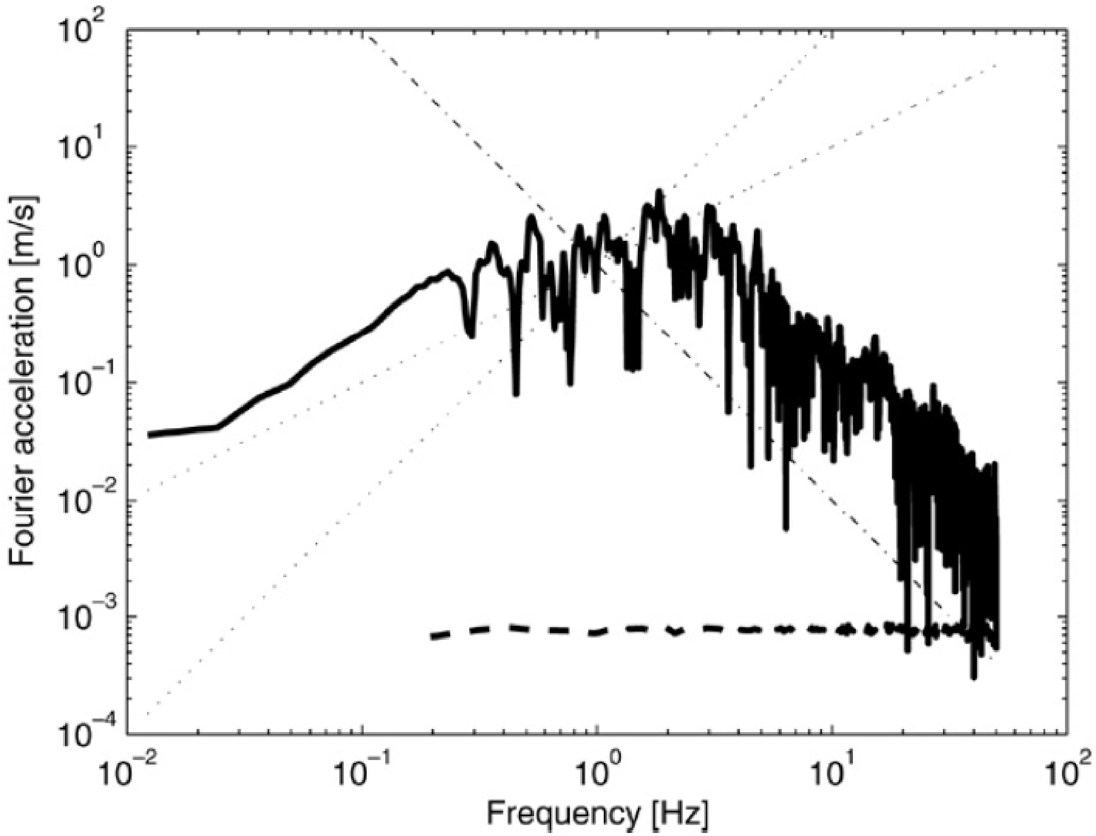
Selection of the long-period cut-offs

- As noted previously, the most important issue in processing strong-motion accelerograms is the choice of the long-period cut-off, or rather the longest response period for which the data are judged to be reliable in terms of signal-to-noise ratio. A number of broad criteria can be employed by the analyst to infer the period beyond which it is desirable to apply the filter cut-off, including:

- Comparison of the FAS of the record with that of a model of the noise, obtained from the pre-event memory for digital records, the fixed trace from analog records or from studies of the instrument and digitizing apparatus.A point of clarification is appropriate here regarding signal-to-noise ratios: the comparison of the record FAS with the FAS of the noise indicates the ratio of signal-plus-noise to noise, hence if the desired target is a signal-to-noise ratio of 2, the ratio of the record FAS to that of the noise model should be 3.

Selection of the long-period cut-offs

- Fourier amplitude spectrum of the FN horizontal component of the Bolu record of Duzce earthquake (thick solid line) and the noise spectrum (thick dashed line). Superimposed on this graph are the functions f^2 (dotted line), f (dotted line), and f^{-2} (dashed–dotted line).

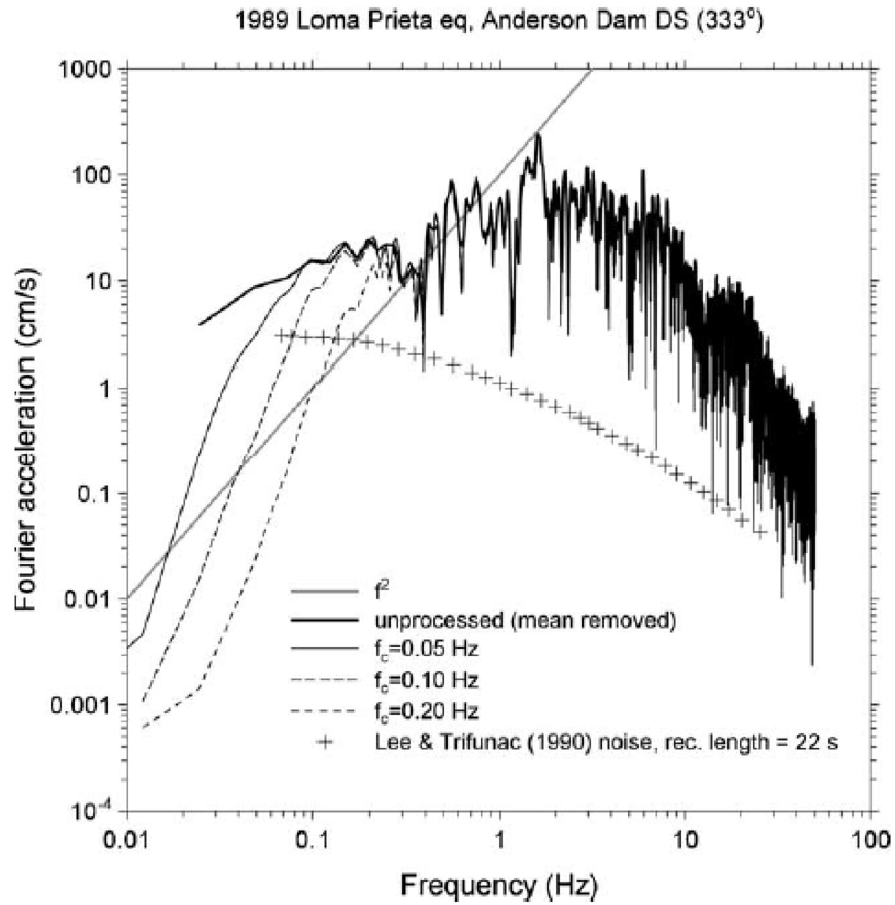


- Seismological theory dictates that at low frequencies, the FAS of acceleration decays according to f^2.

Selection of the long-period cut-offs

- The figure shows the selection of filter parameters for a component of the Anderson Dam (analog) recording of the 1989 Loma Prieta earthquake. The FAS of the record is compared with the model for the digitization noise proposed by Lee and Trifunac.

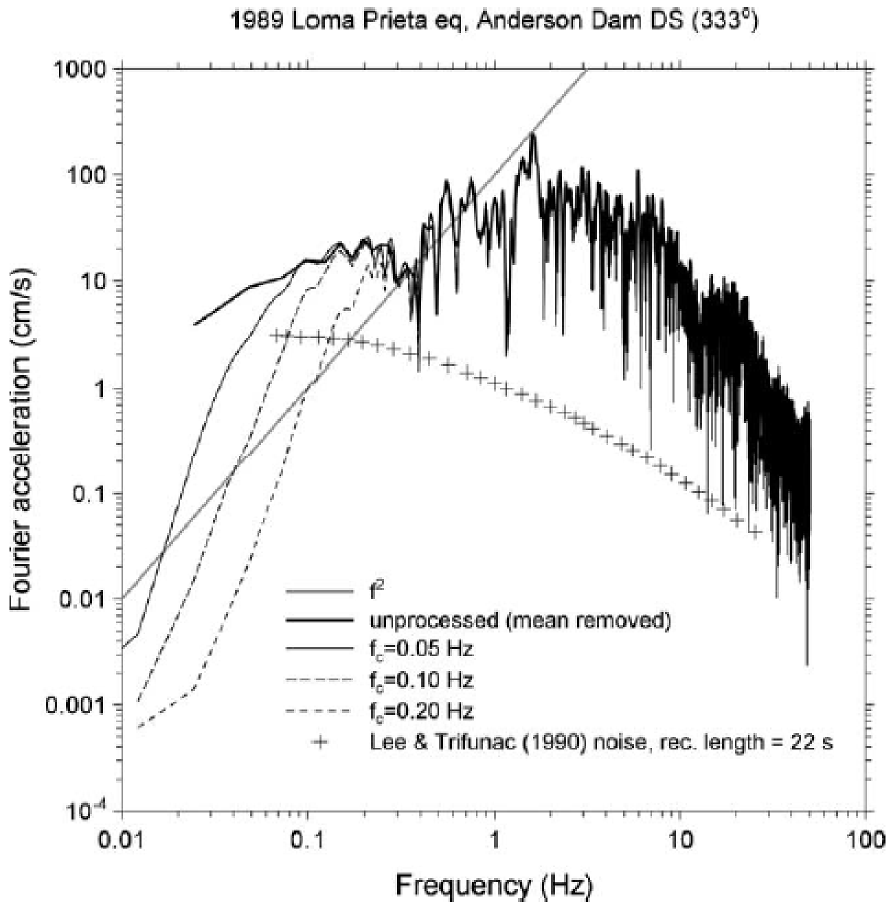
- Also shown is the gradient of the f^2 line, superimposed as a best fit (by eye) on the section of the FAS where the decay at low frequencies commences. Also shown in the graph are the FAS of the record after applying filters with three different low-frequency cut-offs.



Selection of the long-period cut-offs

- These decay more rapidly than indicated by the f^2 model, which is the expected result of effectively trying to remove all of the record—both signal and noise—at periods greater than the cut-off.

- Designing a filter with a gradual roll-off that will produce an FAS that approximates to the f^2 model is not advisable since the agreement with the theoretical seismological model would not mean that the real earthquake signal has been recovered, but only that an unknown mixture of signal and noise has been manipulated to produce the appearance of a genuine seismic motion.



Strong-motion processing

- Processing should be accomplished on a component by component basis.

- Analog recordings have limited usefullness at periods shorter than about 2 or 3 s.

- An issue to be considered in record processing is whether the same filter parameters should be used for all three components or whether optimal processing should be used to obtain the maximum information possible from each of the three components. If the same processing is applied to all three components, the filter cut-off will generally be controlled by the vertical component since this will usually have a lower signal-to-noise ratio than the horizontal components, particularly in the long-period range. Therefore, unless there is a compelling reason for the vertical and horizontal components to be processed with the same filter, this practice is not recommended.

Usable range of response periods

- The amplitude of long-period response spectral ordinates are highly sensitive to the parameters of low-cut filters, and this is most clearly visible when looking at the spectra of relative displacement. Care must be taken in deciding the range of periods for which the spectral ordinates can be reliably used, which depends on both the filter frequency and the order of the filter.

- For a low-order filter applied at 20 s, the spectral ordinates should probably not be used much beyond 10 s. The studies by Abrahamson and Silva and Spudich et al. to derive predictive equations for response spectral ordinates only used each record for periods up to 0.7 times the cut-off period.

- Bommer and Elnashai, in deriving predictions for displacement spectral ordinates, used each record up to 0.1 s less than its cut-off period, which will have inevitably resulted in underestimation of the spectral displacements at longer periods.