Introduction to Digital Filters

Dras. Claudia Feregrino y Alicia Morales

Material original Dr. René Cumplido
CCC INAOE

Introduction to Digital Filters

Digital filters are used for two general purposes

1. Separation of signals that have been combined, and

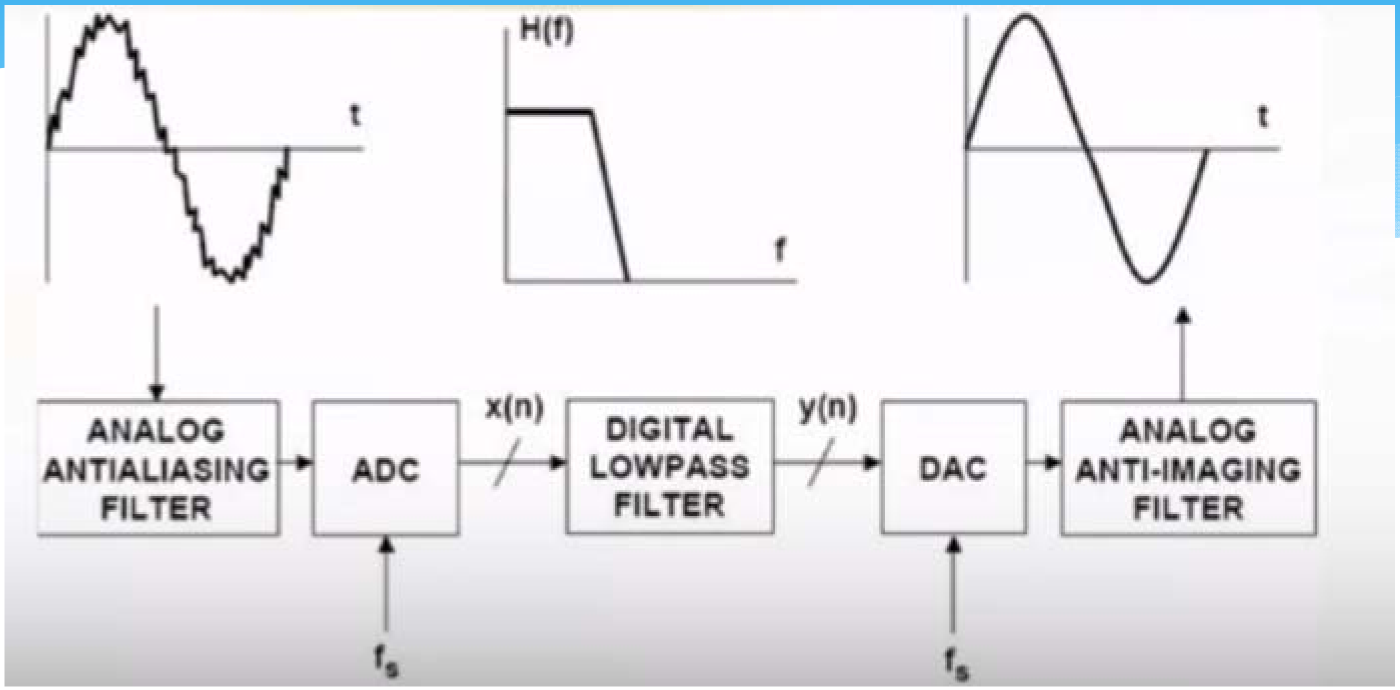
2. Restoration of signals that have been distorted in some way.

* Analog (electronic) filters can be used for these same tasks; however, digital filters can achieve far superior results.

Filter Basics

- Digital filters are a very important part of DSP.
- Their performance is one of the key reasons that DSP has become so popular.
- Filters have two uses:
  - Signal separation is needed when a signal has been contaminated with interference, noise, or other signals.
    - E.g. Measuring the electrical activity of a baby's heart (EKG) while still in the womb.
      - The raw signal will be corrupted by the breathing and eartbeat of the mother
      - A filter might be used to separate these signals.
  - Signal restoration is used when a signal has been distorted in some way.
    - E.g. An audio recording made with poor equipment may be filtered to better represent the sound as it actually occurred.
    - E.g. Another example is the deblurring of an image acquired with an improperly focused lens, or a shaky camera.

Example:



Effect of Lowpass Filter

- The caracteristics of the filter can be easily changed under software control

Digital vs. Analog Filtering

Digital

* High Accuracy
* Linear Phase (FIR Filters)
* Flexible, Adaptive Filtering Possible
* Easy to Simulate and Design
* Digital filters are superior in the level of performance that can be achieved. The entire transition occurs within only 1 hertz.
* Requires High Performance ADC, DAC & DSP
* Digital filters can achieve *thousands* of times better performance than analog filters.

Analog

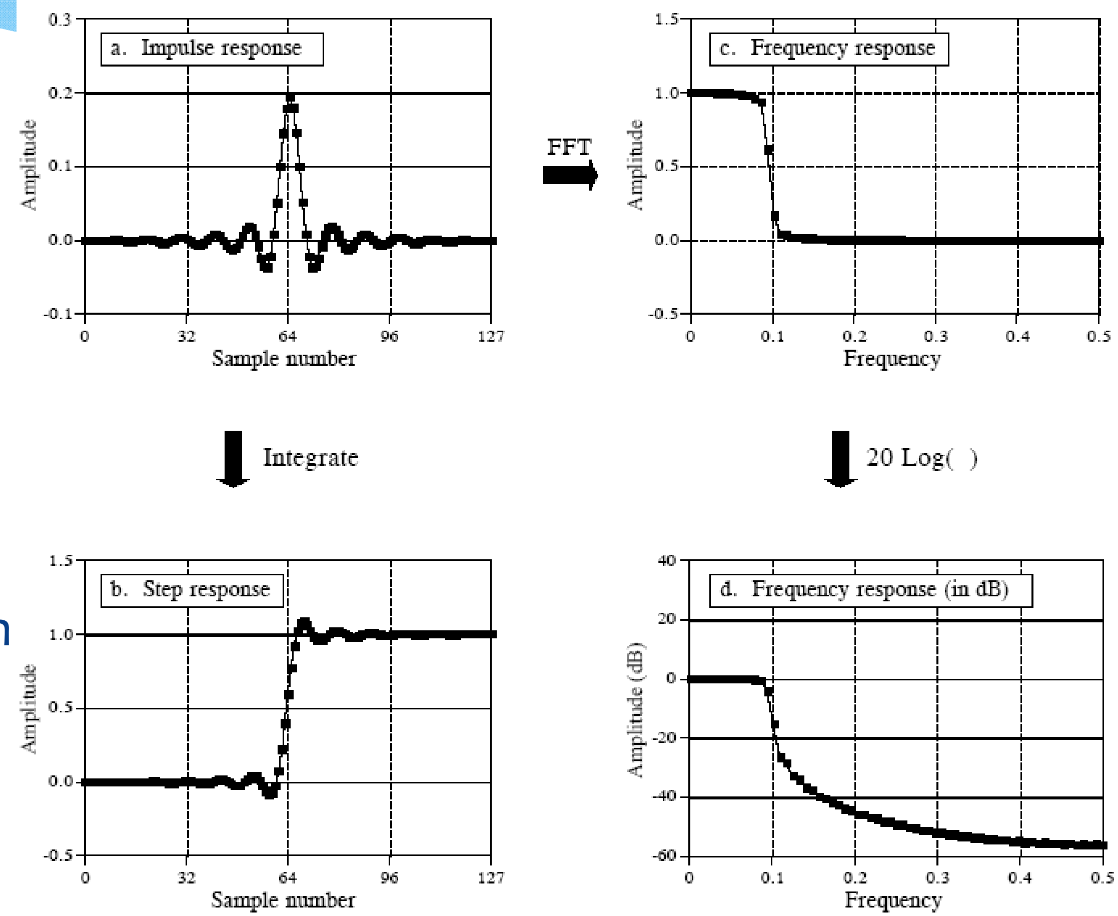
* Less Accuracy – Component Tolerances
* Non-Linear Phase
* Adaptive Filters Difficult
* Difficult to Simulate and Design
* Analog Filters required at High Frequencies and for Anti-Aliasing Filters
* No ADC, DAC, or DSP required
* Analog filters are cheap, fast, and have a large dynamic range in both amplitude and frequency.

* Widely used in communications such as:
  * Echo cancellation
  * Noise cancellation
  * Speech recognition

  * They are NOT the answer to ALL filtering requirements

Filter reponse

- Every linear filter has an **impulse response**, a **step response** and a **frequency response.**
- Each of these responses contains complete information about the filter, but in a different form.
- If one of the three is specified, the other two are fixed and can be directly calculated.
- All three of these representations are important, because they describe how the filter will react under different circumstances.



Convolution and recursion

- The most straightforward way to implement a digital filter is by convolving the input signal with the digital filter's impulse response.
- All possible linear filters can be made in this manner.
- When the impulse response is used in this way, it is called a **filter kernel**.
- There is also another way to make digital filters, called **recursion**.
  - When a filter is implemented by convolution, each sample in the output is calculated by *weighting* the samples in the input, and adding them together.
  - Recursive filters are an extension of this, using previously calculated values from the *output*, besides points from the *input*.
  - Instead of using a filter kernel, recursive filters are defined by a set of **recursion coefficients**.

FIR & IIR Filters

* To find the impulse response of a recursive filter, simply feed in an impulse, and see what comes out.
  * The impulse responses of recursive filters are composed of sinusoids that exponentially decay in amplitude.
  * In principle, this makes their impulse responses *infinitely long*.
  * However, the amplitude eventually drops below the round-off noise of the system, and the remaining samples can be ignored.
  * Because of this characteristic, recursive filters are also called **Infinite Impulse Response or IIR** filters.
* Filters carried out by convolution are called **Finite Impulse Response or FIR** filters.

Step and frequency response

* The *impulse response* is the output of a system when the input is an *impulse*.
* In this same manner, the *step response* is the output when the input is a *step*
* Since the step is the integral of the impulse, the step response is the integral of the impulse response.
* Two ways to find the step response:
  * 1) feed a step waveform into the filter and see what comes out, or
  * 2) integrate the impulse response.
* The frequency response can be found by taking the DFT (using FFT) of the impulse response.

Bell

* **Bel** in honor of Alexander Graham Bell
* Means the power changed by a factor of ten
  * Example: An electronic circuit that has 3 bels of amplification produces an output signal with 1000 times the power of the input.
  * A **decibel** (dB) is one-tenth of a bel.
    * Every ten decibels mean that the power has changed by a factor of ten.

20 dB → 100 times	-20 dB → ?
10 dB → 10 times	-10 dB → ?
0 dB → 1 time	

* Decibel originates from methods used to quantify signal loss in telegraph and telephone circuits.

* The unit for loss was originally Miles of Standrard Cable (MSC). 1 MSC corresponded to the loss of power over 1 mile lenght of standard telephone cable of 5000 radians per second (795.8 Hz), and matched closely to the smallest attenuation detectable to the average listener.

* In 1924, Bell Telephone Labs received favorable response to a new unit definition among members of the International Advisory Committee on Long Distance Telephony in Europe and replaced the MSC with the *Transmission Unit* (TU).
* 1 TU = 10 times the base-10 logarithm of the ratio of measured power to a reference power level.
* Definition conveniently chosen such that 1 Tu $\approx$ 1MSC (1 MSC = 1.056 TU)
* In 1928, TU was renamed decibel.

Examples of sound pressure in air at standard atmospheric pressure

Source of sound	Sound pressure* (Pa)	Sound pressure level (dB_{SPL})
Risk of instantaneous noise-induced hearing loss	20	120
Jet engine at 100 m	6.32–200	110–140
Non-electric chainsaw at 1 m[17]	6.32	110
Jack hammer at 1 m	2	100
Traffic on a busy roadway at 10 m	0.2–0.632	80–90
Hearing damage (over long-term exposure, need not be continuous)[18]	0.356	85
Passenger car at 10 m	$(2–20)\times10^{-2}$	60–80
EPA-identified maximum to protect against hearing loss and other disruptive effects from noise, such as sleep disturbance, stress, learning detriment, etc.[19]	6.32×10^{-2}	70
Handheld electric mixer		65
TV (set at home level) at 1 m	2×10^{-2}	60
Washing machine, dishwasher[20]		42–53
Normal conversation at 1 m	$(2–20)\times10^{-3}$	40–60
Very calm room	$(2–6.32)\times10^{-4}$	20–30
Light leaf rustling, calm breathing	6.32×10^{-5}	10
Auditory threshold at 1 kHz[18]	2×10^{-5}	0

Examples of sound pressure in air at standard atmospheric pressure

Source of sound	Sound pressure* (Pa)	Sound pressure level (dB_{SPL})
Shockwave (distorted sound waves > 1 atm; waveform valleys are clipped at zero pressure)	>101,325	>194
Theoretical limit for undistorted sound at 1 atmosphere environmental pressure	101,325	194
Stun grenade	6,000–20,000	170–180
Simple open-ended thermoacoustic device[10]	12,619	176
.30-06 rifle being fired 1 m to shooter's side	7,265	171
Rocket launch equipment acoustic tests	4000	165
LRAD 1000Xi Long Range Acoustic Device at 1 m[11]	893	153
Jet engine at 1 m	632	150
Threshold of pain[12][13][14]	63-200	130-140
Loudest human voice at 1 inch[14]	110	135
Trumpet at 0.5 m[15]	63.2	130
Vuvuzela horn at 1 m[16]	20	120

* The commonly used reference sound pressure in air is often considered as the *threshold of human hearing* (roughly the sound of a mosquito flying 3 m away)

$$\text{dB} = 10 \log_{10} \frac{P_2}{P_1}$$

$$\text{dB} = 20 \log_{10} \frac{A_2}{A_1}$$

How Information is Represented in Signals

* Information represented in the time domain describes when something occurs and what the amplitude of the occurrence is.
* In contrast, information represented in the frequency domain is more indirect.
  * Many things in our universe show periodic motion.
  * By measuring the frequency, phase, and amplitude of this periodic motion, information can often be obtained about the system producing the motion.
* A single sample, in itself, contains no information about the periodic motion, and therefore no information about the source
  * The information is contained in the *relationship* between many points in the signal.

* The *step response* describes how information represented in the *time domain* is being modified by the system.

* In contrast, the *frequency response* shows how information represented in the *frequency domain* is being changed.

* This distinction is absolutely critical in filter design because it is not possible to optimize a filter for both applications.
  * Good performance in the time domain results in poor performance in the frequency domain, and vice versa.

* If a filter is designed to remove noise from an EKG signal (time domain), the step response is the important parameter, and the frequency response is of little concern.

* If your task is to design a digital filter for a hearing aid (frequency domain), the frequency response is all important, while the step response doesn't matter.

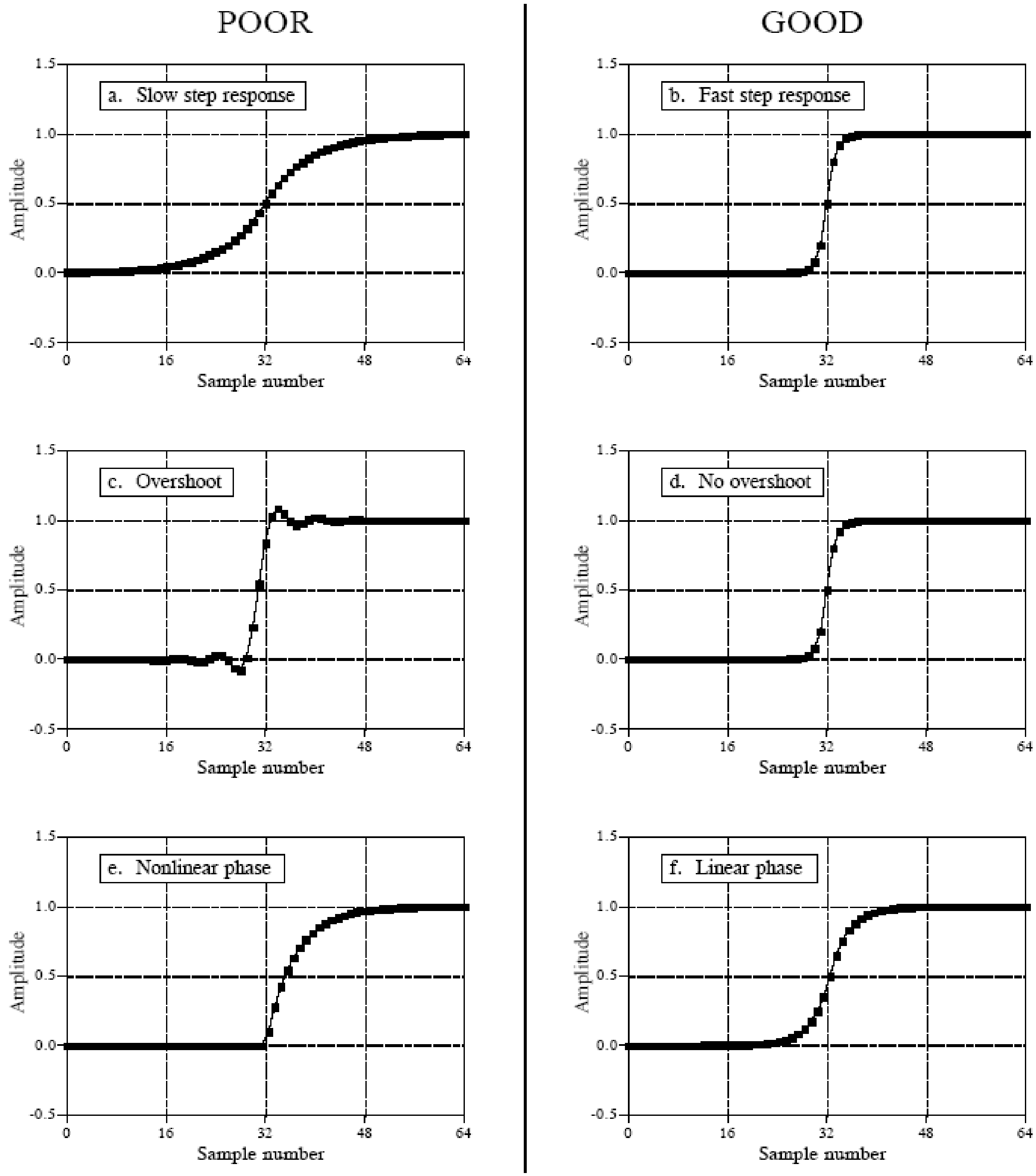
Time Domain Parameters

* Why step response instead of impulse response?
  * The step response is useful in time domain analysis because it matches the way humans view the information contained in the signals.
  * E.g. you are given a signal of some unknown origin and asked to analyze it. Then you:
    * divide the signal into regions of similar characteristics.
      * Some of the regions may be smooth; others may have large amplitude peaks; others may be noisy.
      * This segmentation is accomplished by identifying the points that separate the regions.
      * The step function is the purest way of representing a division between two dissimilar regions.
      * It can mark when an event starts, or when an event ends.
    * The step response, in turn, is important because it describes how the dividing lines are being modified by the filter.

* To distinguish events in a signal, the duration of the step response must be shorter than the spacing of the events.

* This dictates that the step response should be as *fast* as possible.

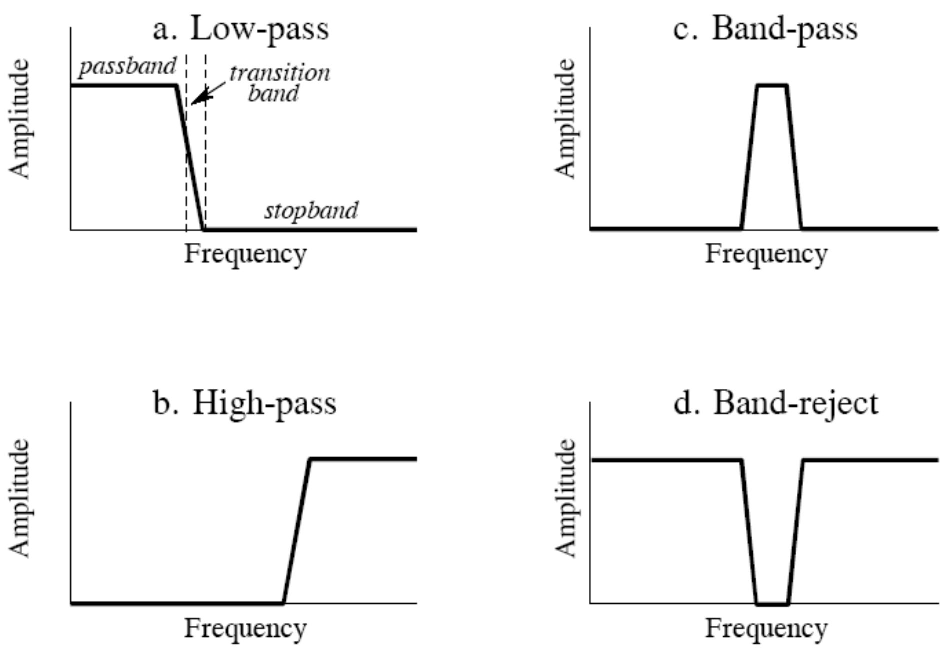
* The most common way to specify the **risetime** is to quote the number of samples between the 10% and 90% amplitude levels.



Frequency Domain Parameters

The purpose of these filters is to allow some frequencies to pass unaltered, while completely blocking other frequencies.

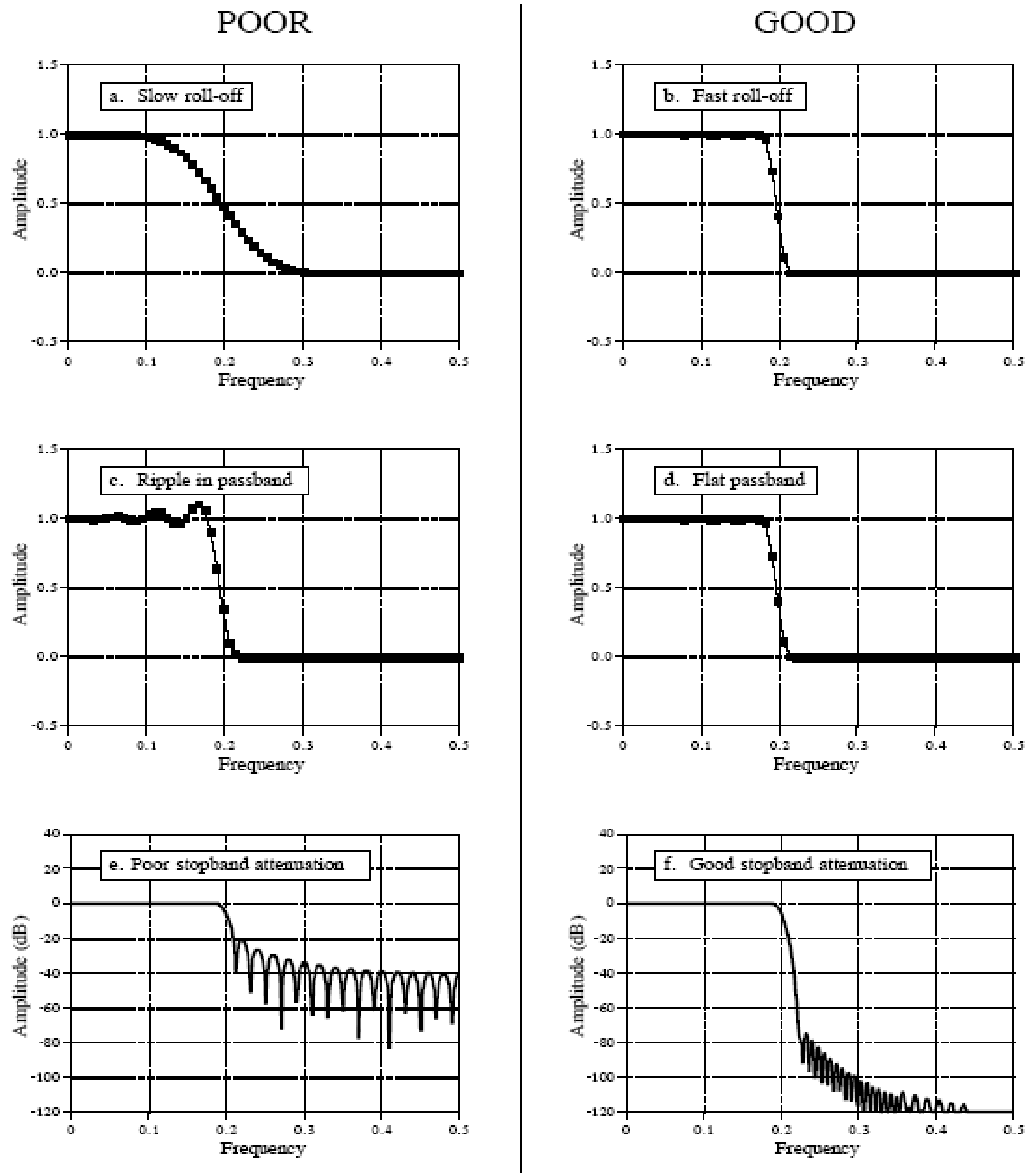
* The **passband** refers to those frequencies that are passed
* The **stopband** contains those frequencies that are blocked.
* The **transition band** is between.
* A **fast roll-off** means that the transition band is very narrow.
* The division between the passband and transition band is called the **cutoff frequency.** In DSP 99%, 90%, 70.7%, and 50% amplitude levels are defined to be the cutoff frequency.



Frequency Domain Parameters

There are three parameters that measure how well a filter performs in the frequency domain.

* To separate closely spaced frequencies, the filter must have a **fast roll-off,** (a) and (b).
* For the passband frequencies to move through the filter unaltered, there must be no **passband ripple,** (c) and (d).
* To adequately block the stopband frequencies, it is necessary to have good **stopband attenuation,** (e) and (f).



High-Pass, Band-Pass and Band-Reject Filters

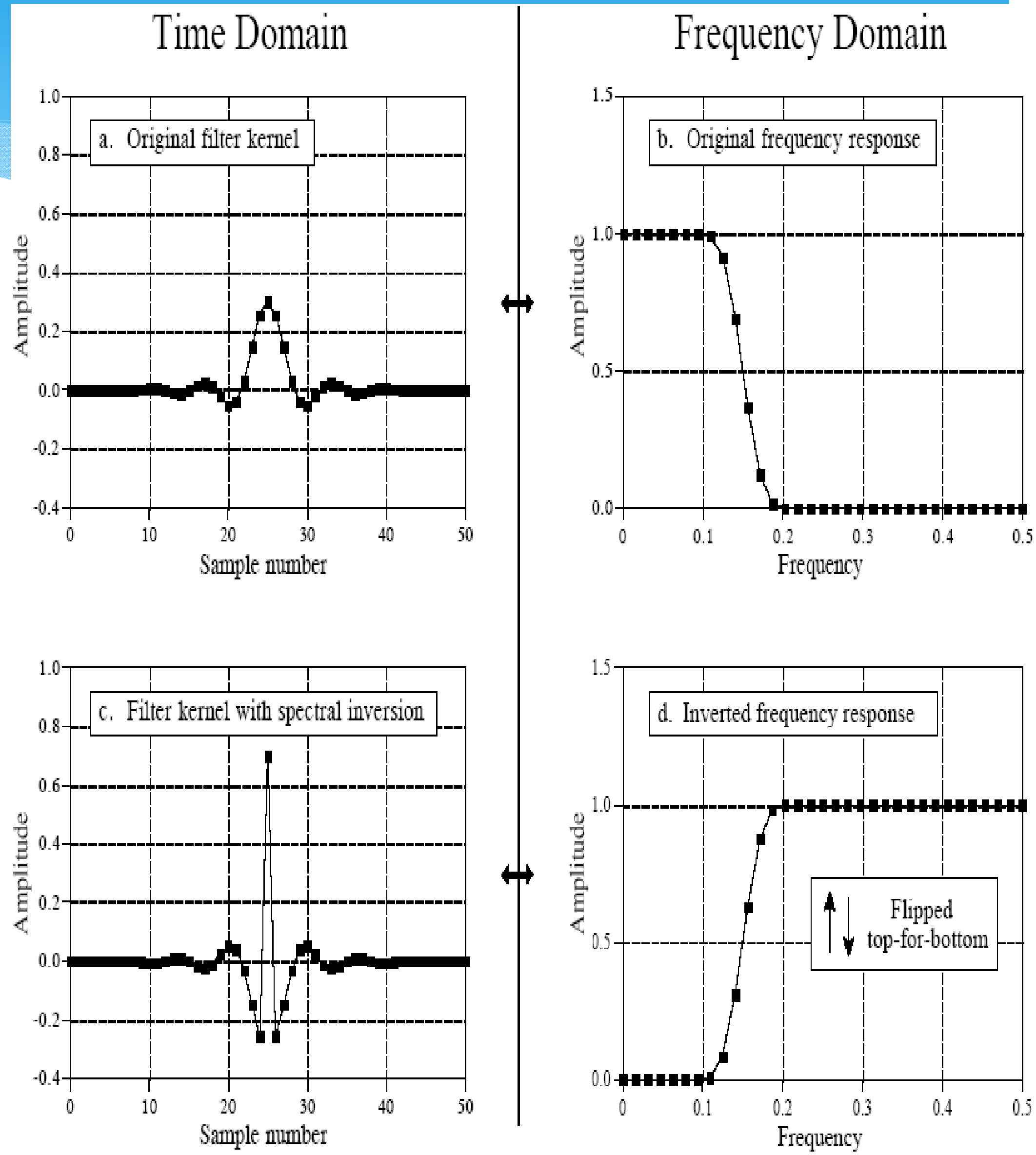
* High-pass, band-pass and band-reject filters are designed by starting with a low-pass filter, and then converting it into the desired response.
* For this reason, most discussions on filter design only give examples of low-pass filters.
* There are two methods for the low-pass to high-pass conversion:
  * **spectral inversion**
  * **spectral reversal**.

Spectral inversion

* To change the low-pass filter kernel into a high-pass filter kernel.
  1. Change the sign of each sample in the filter kernel.
  2. Add *one* to the sample at the center of symmetry.

* Spectral inversion *flips* the frequency response *top-for-bottom, i.e.*

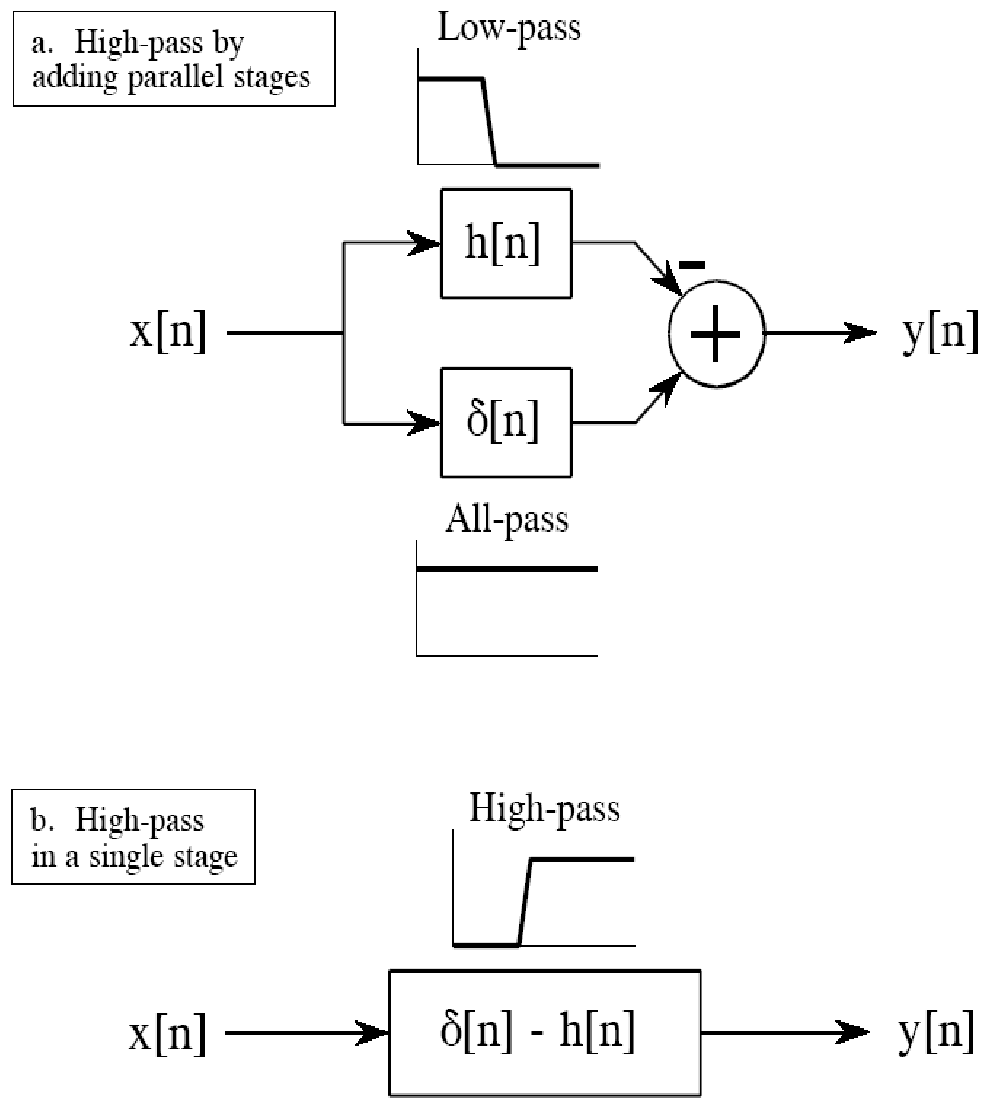
  passbands -> stopbands, stopbands -> passbands.



Spectral inversion

- Why this two step modification to the time domain results in an inverted frequency spectrum?

- Since the low frequency components are subtracted from the original signal, only the high frequency components appear in the output. Thus, a high-pass filter is formed.



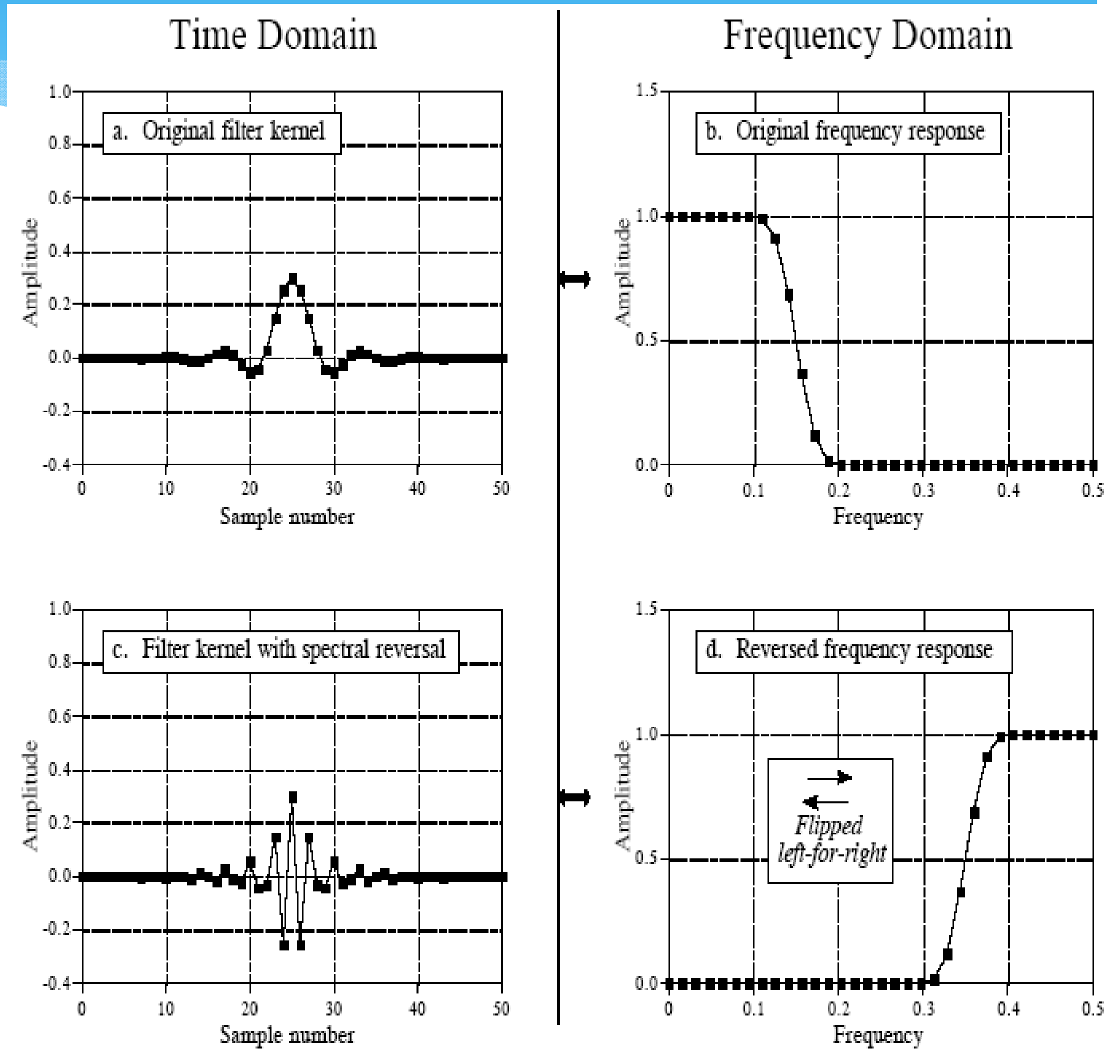
Spectral inversion

- This could be performed as a two step operation in a computer program:
  - run the signal through a low-pass filter, and then
  - subtract the filtered signal from the original.

- However, the entire operation can be performed in a signal stage by combining the two filter kernels.
  - Parallel systems with added outputs can be combined into a single stage by adding their impulse responses.
  - The filter kernel for the highpass filter is given by: $\delta[n] - h[n]$

Spectral reversal

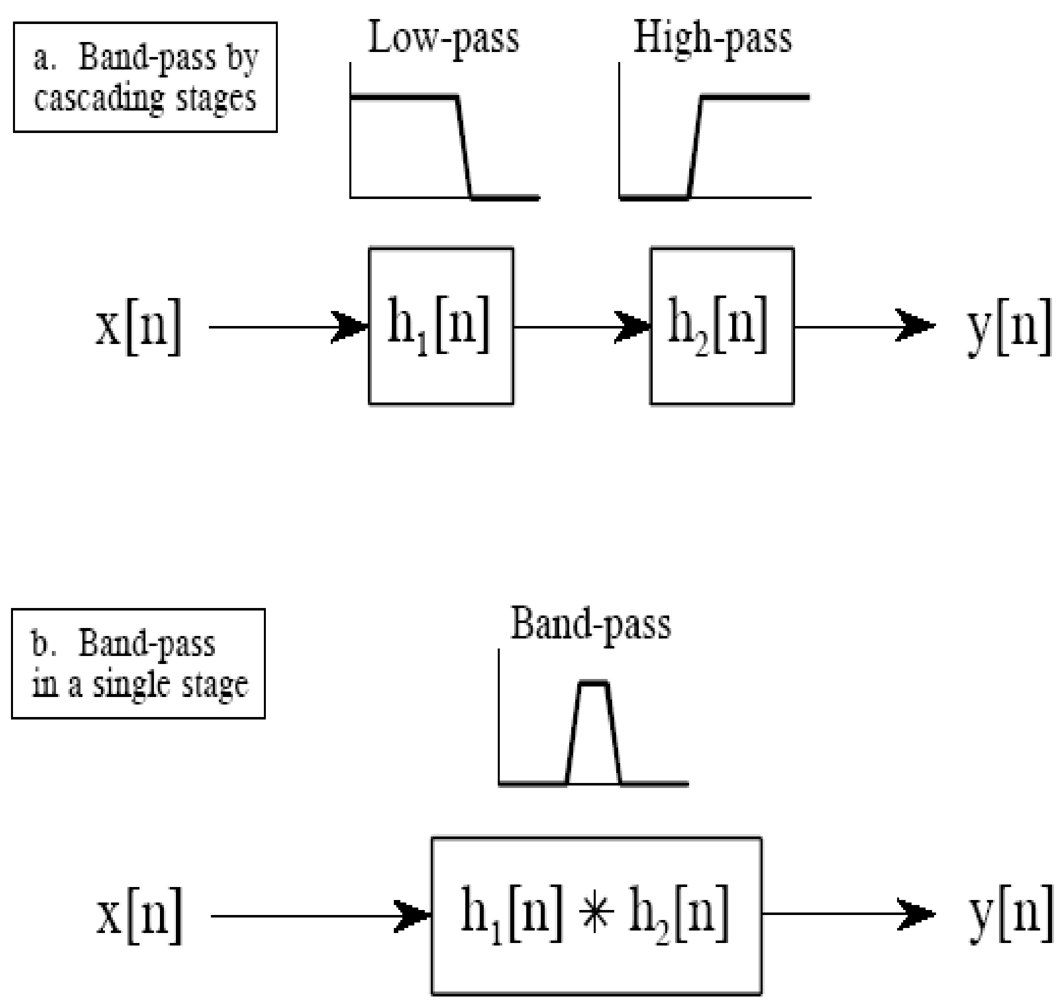
* The high-pass filter kernel, (c), is formed by *changing the sign of every other sample* in (a).
* This flips the frequency domain *left-for-right*: 0 becomes 0.5 and 0.5 becomes 0.
* The cutoff frequency of the example low-pass filter is 0.15, resulting in the cutoff frequency of the high-pass filter being 0.35.



Spectral reversal

* Changing the sign of every other sample is equivalent to multiplying the filter kernel by a sinusoid with a frequency of 0.5.
* This has the effect of *shifting* the frequency domain by 0.5.
* Look at (b) and imagine the negative frequencies between -0.5 and 0 that are of mirror image of the frequencies between 0 and 0.5.
* The frequencies that appear in (d) are the negative frequencies from (b) shifted by 0.5.

* Low-pass and high-pass filter kernels can be combined to form band-pass and band-reject filters.
* *Adding* the filter kernels produces a *band-reject* filter,
* *Convolving* the filter kernels produces a *band-pass* filter.
* These are based on the way cascaded and parallel systems are be combined
* Multiple combination of these techniques can also be used.



Filter Classification

- Digital filters are classified by their *use* and by their *implementation.*
- The use of a digital filter can be broken into three categories:
  - *time domain*
    - Used when the information is encoded in the
      - Smoothing, DC removal, waveform shaping shape of the signal's waveform
  - *frequency domain*
    - Used when information is contained in the amplitude, frequency, and phase of the component sinusoids
      - The goal of these filters is to separate one band of frequencies from another
  - *custom.*
    - Used when a special action is required by the filter
      - something more elaborate than the four basic responses (high-pass, low-pass, band-pass and band-reject).

Filter Classification

FILTER USED FOR:	FILTER IMPLEMENTED BY: Convolution *Finite Impulse Response (FIR)*	FILTER IMPLEMENTED BY: Recursion *Infinite Impulse Response (IIR)*
Time Domain *(smoothing, DC removal)*	Moving average (Ch. 15)	Single pole (Ch. 19)
Frequency Domain *(separating frequencies)*	Windowed-sinc (Ch. 16)	Chebyshev (Ch. 20)
Custom *(Deconvolution)*	FIR custom (Ch. 17)	Iterative design (Ch. 26)

Filters carried out by convolution (FIR) can have far better performance than filters using recursion (IIR), but execute much more slowly.