Course Notes 9 – FIR Filter Design Techniques

9.0 Design of FIR Filters by Windowing

The approach here is simple in concept:

1. Expand the frequency response of the desired filter using the Fourier Series. The Fourier Series coefficients will then be the impulse response of the desired filter.
2. Truncate the Fourier Series coefficients to the desired filter length.

Let $H_d(\omega)$ denote the desired frequency response

Since $H_d(\omega)$ is a periodic function of ω with a period 2π, it can be expressed as a Fourier series

$$H_d(\omega) = \sum_{n=-\infty}^{\infty} h_d(n) e^{-j\omega n}$$

where

$$h_d(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(\omega) e^{j\omega n} d\omega, \quad -\infty \le n \le \infty$$

- In general, $H_d(\omega)$ has infinitely sharp (i.e. abrupt) transitions between bands.

- The consequence of infinitely sharp transition bands is that $h_d(n)$ is of infinite length and noncausal.

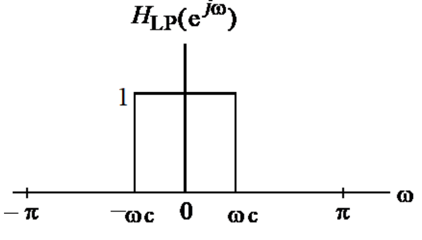
- The objective is to find a finite-duration $h(n)$ of finite length whose DTFT $H(\omega)$ approximates the desired DTFT $H_d(\omega)$ in some sense.

- A causal FIR filter with an impulse response $h(n)$ can be derived from $h_d(n)$ by delaying:

$$h(n) = h_d(n-\alpha), \quad 0 \le n \le N-1 \qquad \text{where} \qquad \alpha = \frac{N-1}{2}$$

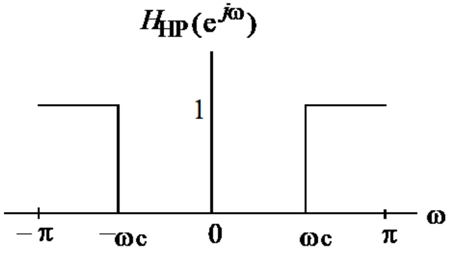
Some ideal filter models to serve as $h_d(n)$ are shown below:

Ideal Lowpass Filter



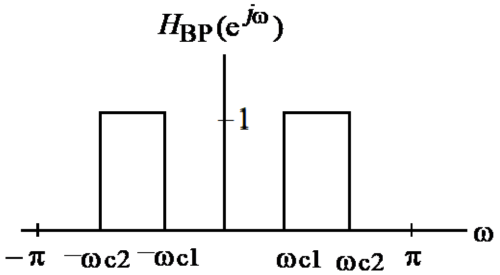
$$h_{LP}[n] = \begin{cases} \dfrac{\omega_c}{\pi}, & n = 0 \\ \dfrac{\sin \omega_c n}{\pi n}, & n \neq 0 \end{cases}$$

Ideal Highpass Filter



$$h_{HP}[n] = \begin{cases} 1 - \dfrac{\omega_c}{\pi}, & n = 0 \\ -\dfrac{\sin(\omega_c n)}{\pi n}, & n \neq 0 \end{cases}$$

Ideal Bandpass Filter



$$h_{BP}[n] = \begin{cases} \dfrac{\sin(\omega_{c2} n)}{\pi n} - \dfrac{\sin(\omega_{c1} n)}{\pi n}, & n \neq 0 \\ \dfrac{\omega_{c2}}{\pi} - \dfrac{\omega_{c1}}{\pi}, & n = 0 \end{cases}$$

Let's examine some of the principles involved in this technique. Let's use an ideal lowpass filter as our desired filter. Since an ideal LPF is not causal and infinite in extent, we want to delay it and truncate it as follows:

$$h_d(n) = \begin{cases} \dfrac{\omega_c}{\pi}, & n = \alpha \\ \dfrac{\sin[\omega_c(n-\alpha)]}{\pi(n-\alpha)}, & n \neq \alpha \end{cases}$$

This impulse response is infinite but symmetric about α.

Obtain the actual filter impulse response by truncating:

$$h(n) = \begin{cases} h_d(n), & 0 \le n \le N\text{-}1 \\ 0, & \text{otherwise} \end{cases}$$

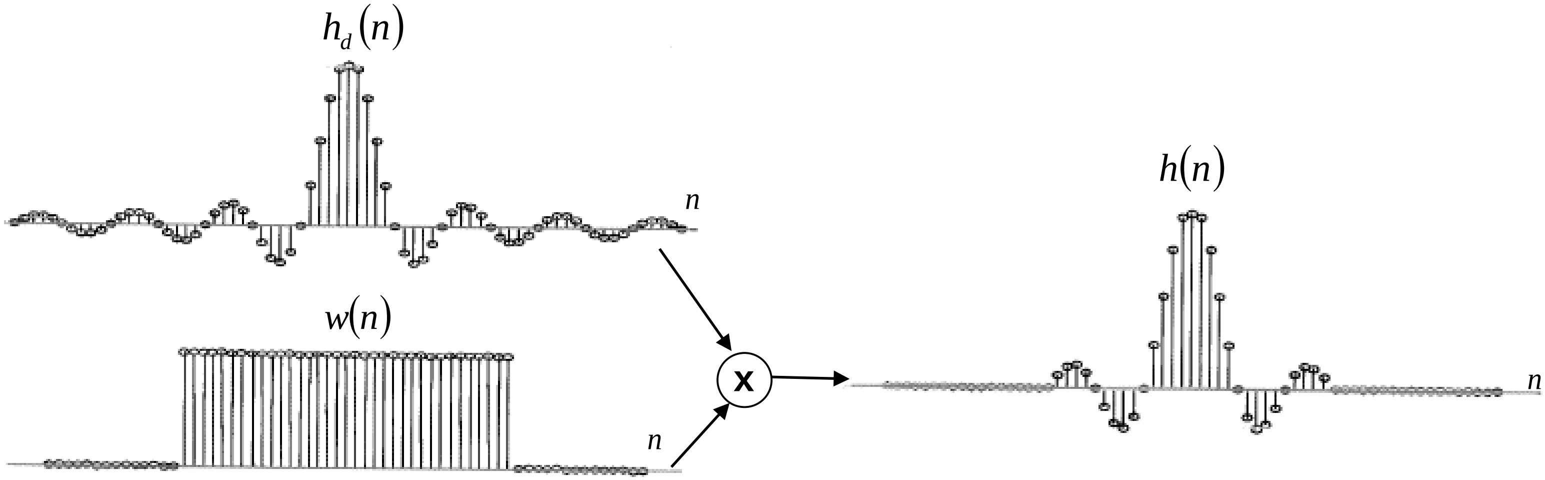
This yields a causal and linear phase FIR filter as long as symmetry conditions exist about

$$\alpha = \frac{N-1}{2}$$

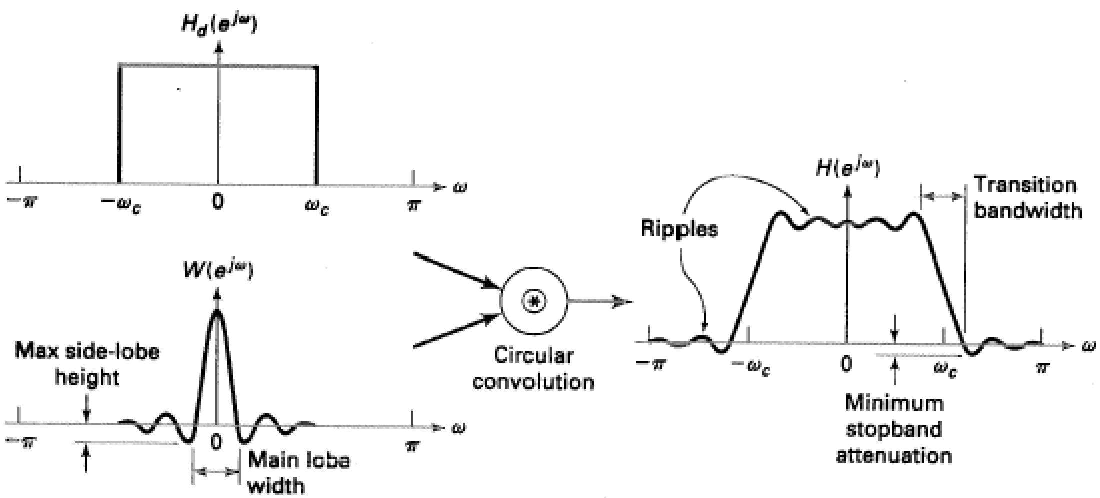
This process is called "windowing", and is implemented by multiplying the desired impulse response with a finite window function:

$$h(n) = h_d(n)w(n)$$

When viewed in the time domain:



… and in the frequency domain:



Some notes regarding this process:

- Since the window $w(n)$ has a finite length of N, its frequency response has a mainlobe proportional to $1/N$.
- The convolution produces a smeared version of the ideal response.
- The mainlobe of the window frequency response is responsible for the transition width. The wider the mainlobe, the wider the transition width.
- The sidelobes produce ripples that have similar shapes in both the passband and stopband.

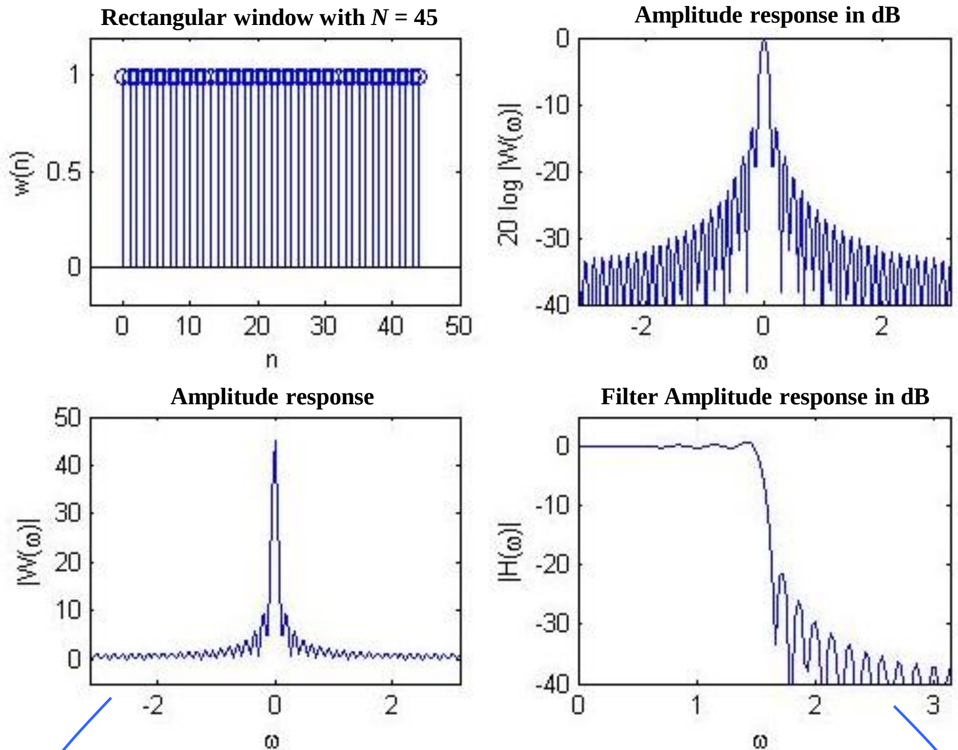
The window function that simply truncates the impulse is called the **Rectangular Window**:

- Mainlobe of this window function is characterized by its width $4\pi/(N+1)$ defined by first zero-crossings on both sides of $\omega = 0$ (otherwise known as the "null-to-null" width)
- As N increases, the mainlobe width decreases.
- The amplitude of each lobe remains constant while the width of each lobe decreases with an increase in N.
- Ripples in $H(\omega)$ around the "point of discontinuity" (the abrupt transition) occur more closely but with no decrease in amplitude as N increases.

Here is a closer look at the essential characteristics of the rectangular window:

The time domain window function

The amplitude response expressed in dB. Note the first sidelobe is at -13 dB.

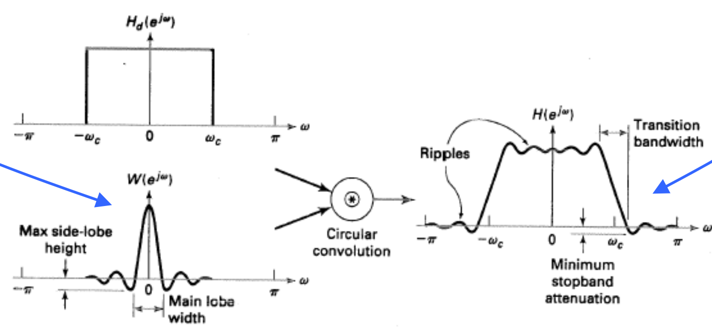


The amplitude response of the DTFT of the window function (absolute scale)

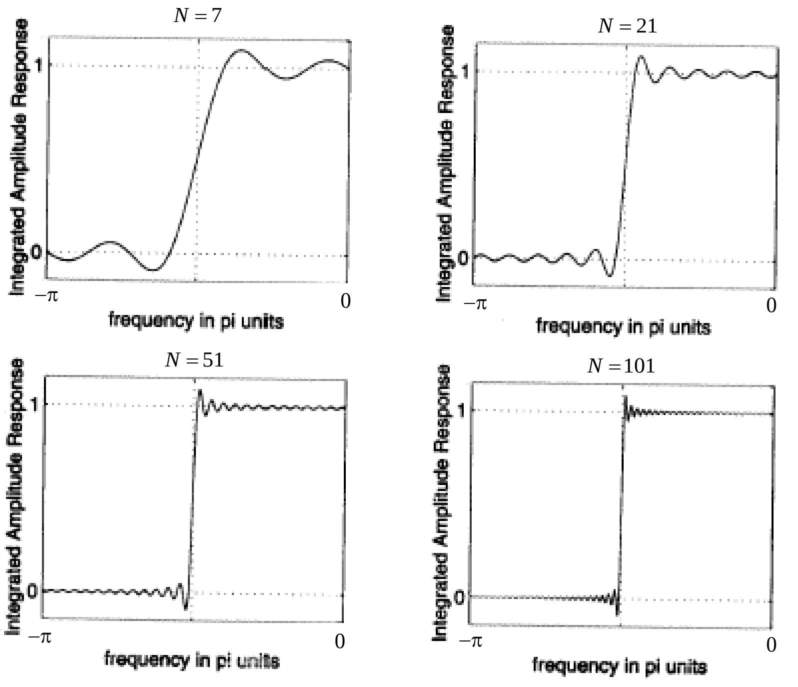
This is the amplitude response characteristic of the actual filter obtained by using this window. Note the first sidelobe is at -21 dB.

This is sometimes called the *peak approximation error*

This is the diagram describing windowing in the frequency domain from a previous page



The rectangular window involves a direct truncation, and therefore it is usually not a desirable window to use because of the presence of ringing due to the Gibbs effect.

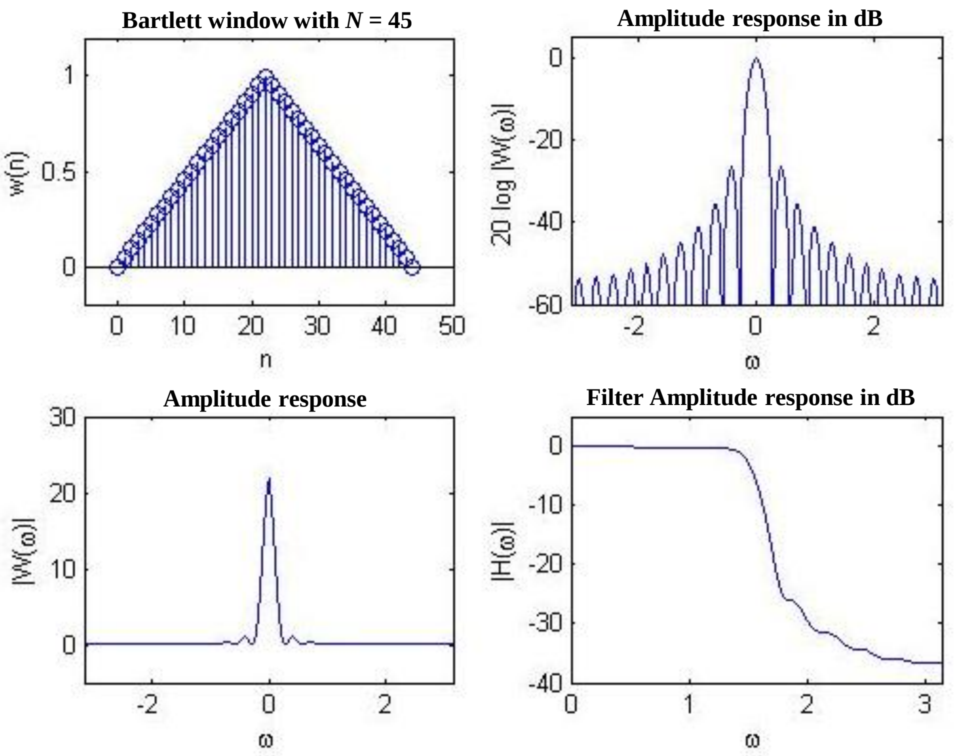


As the window length increases, the width of each sidelobe decreases but the amplitude of each lobe remains constant.

Other common window functions

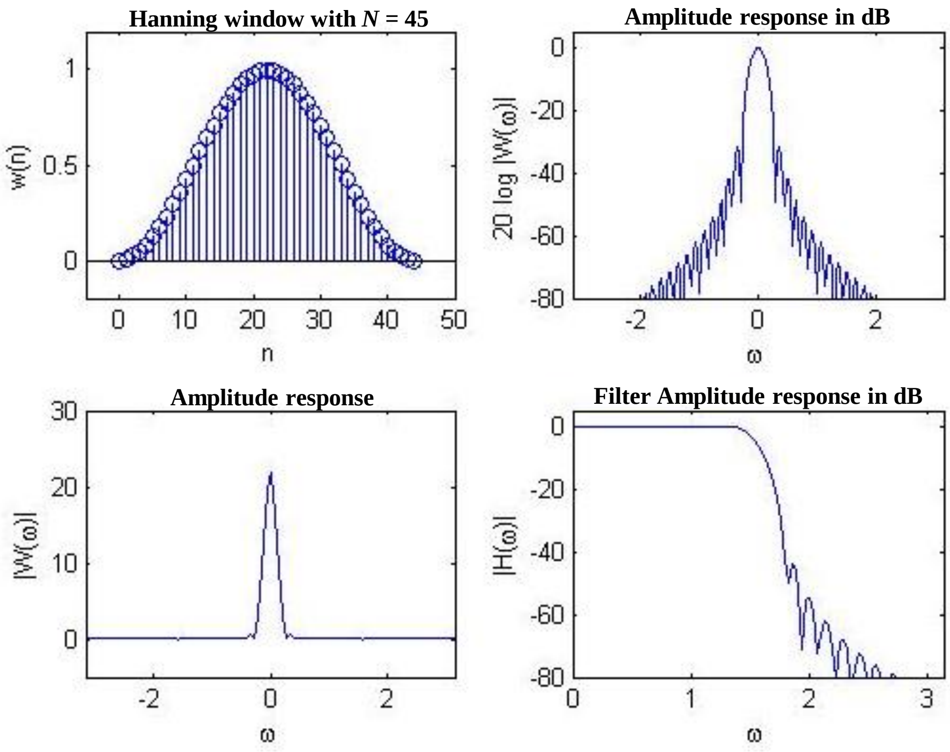
Bartlett (triangular) Window

$$w(n) = 1 - \frac{2\left|n - \frac{N-1}{2}\right|}{N-1}$$



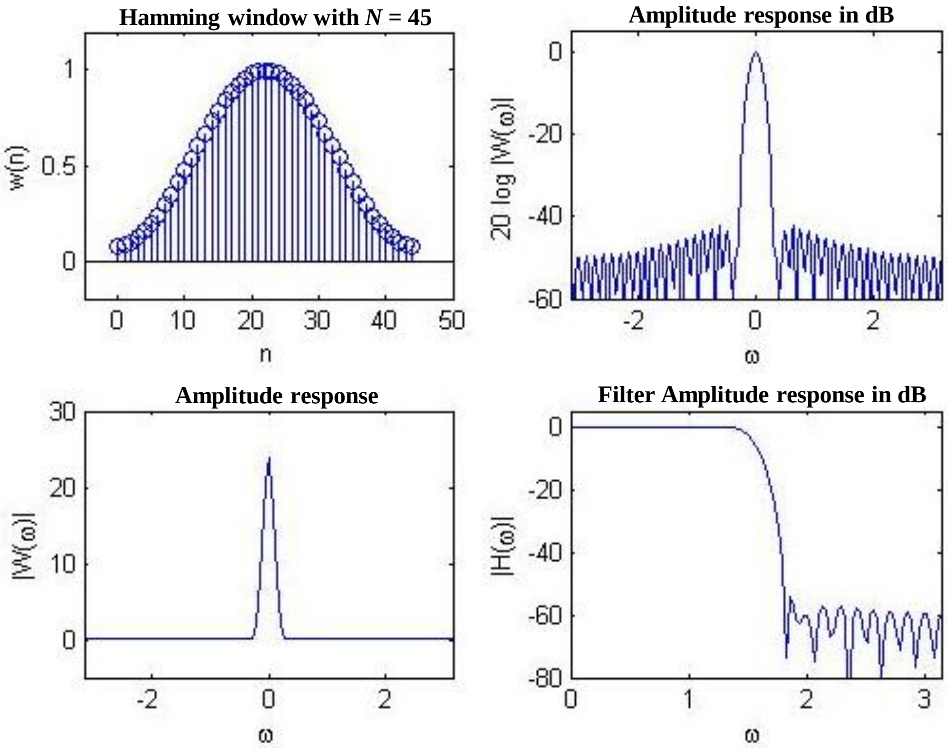
Hanning Window

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



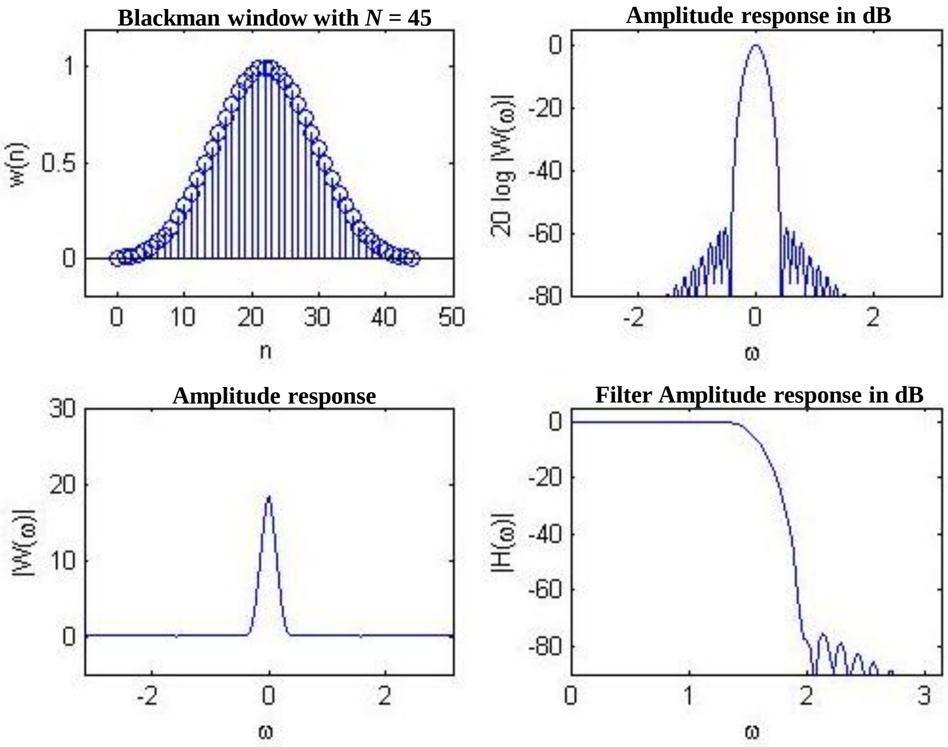
Hamming Window

$$w(n) = 0.54 - 0.46\cos\frac{2\pi n}{N-1}$$



Blackman Window

$$w(n) = 0.42 - 0.5\cos\frac{2\pi n}{N-1} + 0.08\cos\frac{4\pi n}{N-1}$$



Comparing some of the properties of these window functions:

Window Type	Relative SL Amplitude	ML Width	Peak Approx. Error
Rectangular	-13 dB	$4\pi/N$	-21 dB
Bartlett	-25 dB	$8\pi/N$	-25 dB
Hanning	-31 dB	$8\pi/N$	-44 dB
Hamming	-41 dB	$8\pi/N$	-53 dB
Blackman	-57 dB	$12\pi/N$	-74 dB

Fundamental trade-off: the lower the sidelobes are suppressed, the wider the mainlobe becomes.

Steps in design of an FIR low pass filter:

1. Estimate the cutoff frequency as:

$$\omega_c = \frac{\omega_s + \omega_p}{2}$$

ω_p – edge of passband frequency
ω_s – edge of stopband frequency

2. Choose window type based on specified **stop band** attenuation requirements.

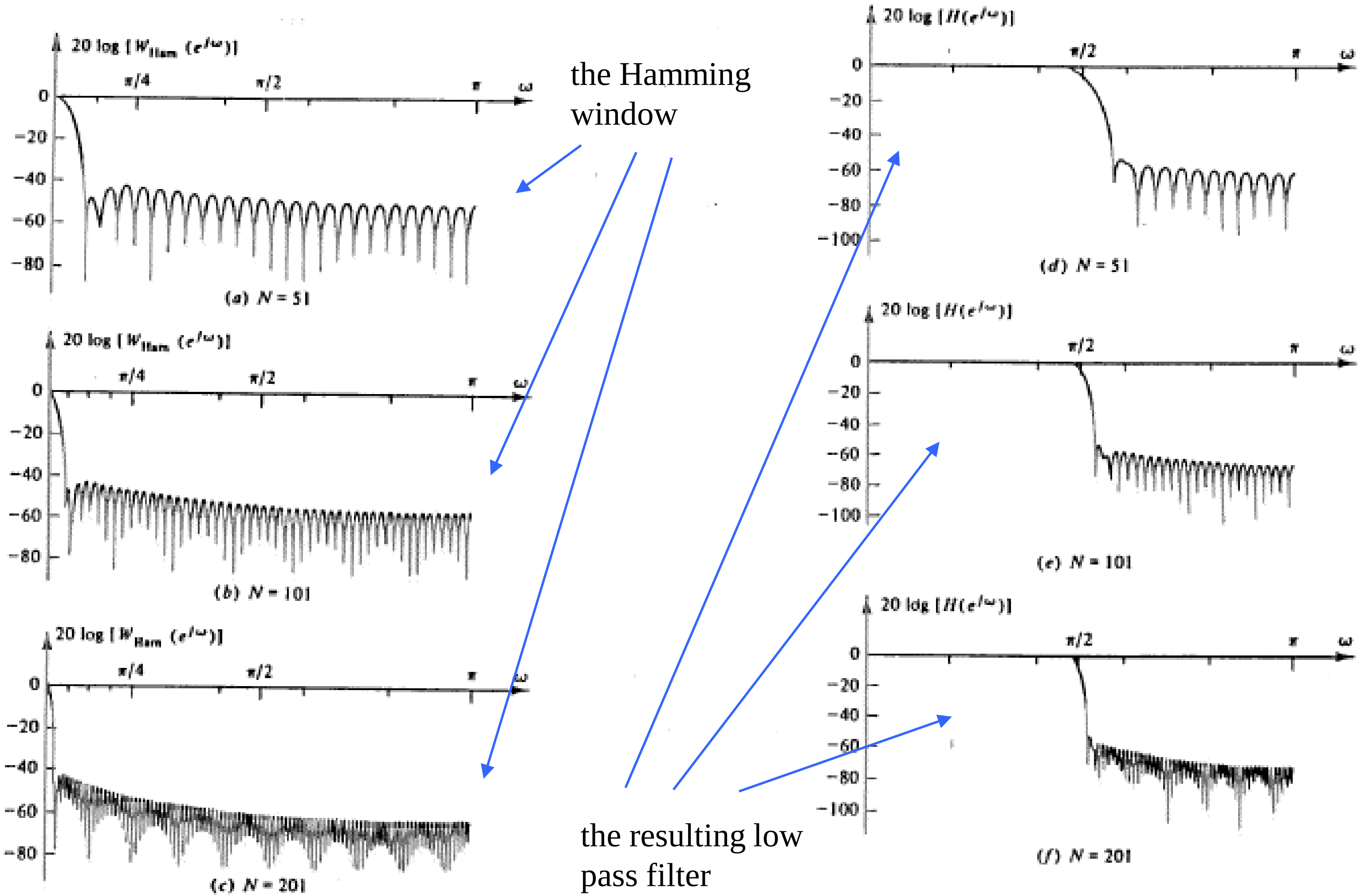
3. Estimate the filter length N by using the transition width requirement as:

$$\Delta\omega = \omega_s - \omega_p \geq \Delta_{ML}$$

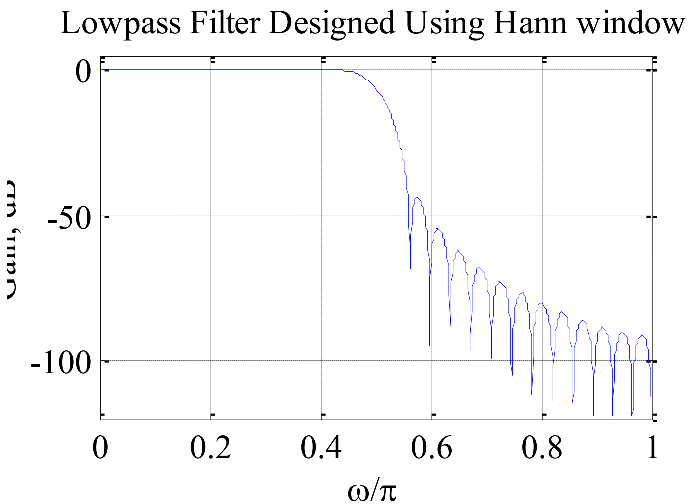
$\Delta\omega$ - transition band width
Δ_{ML} - main lobe width of window

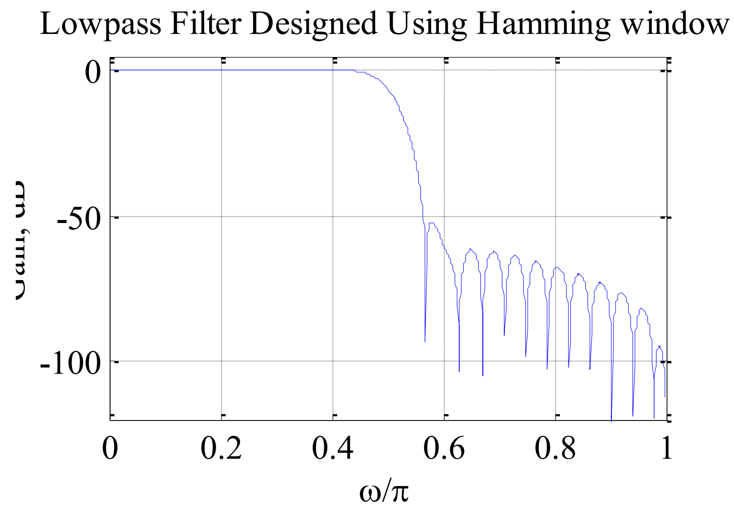
4. Compute ω_c and α for the given filter and then test to ensure specs are met. If not, iterate steps 1 and 3 (and perhaps even 2).

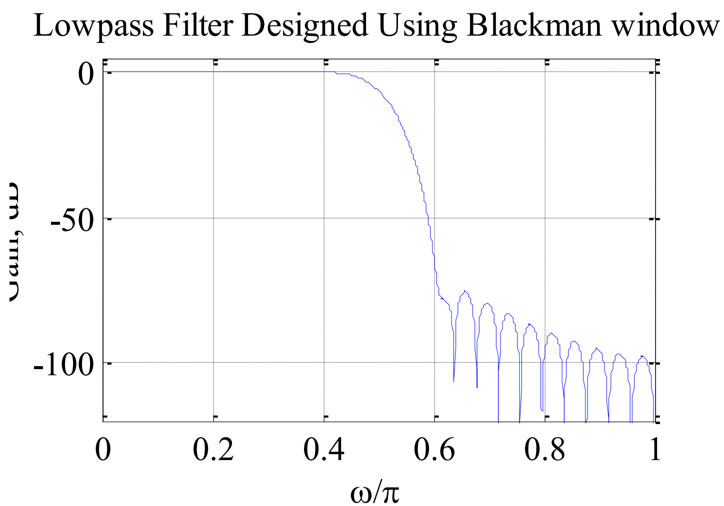
If we were to use the Hamming window of several different lengths (N = 51, 101, and 201) to construct a low pass filter, here are the results.



Another example – a lowpass filter of length 51 with cutoff frequency $\omega_c = \pi / 2$







- An increase in the mainlobe width is associated with an increase in the width of the transition band

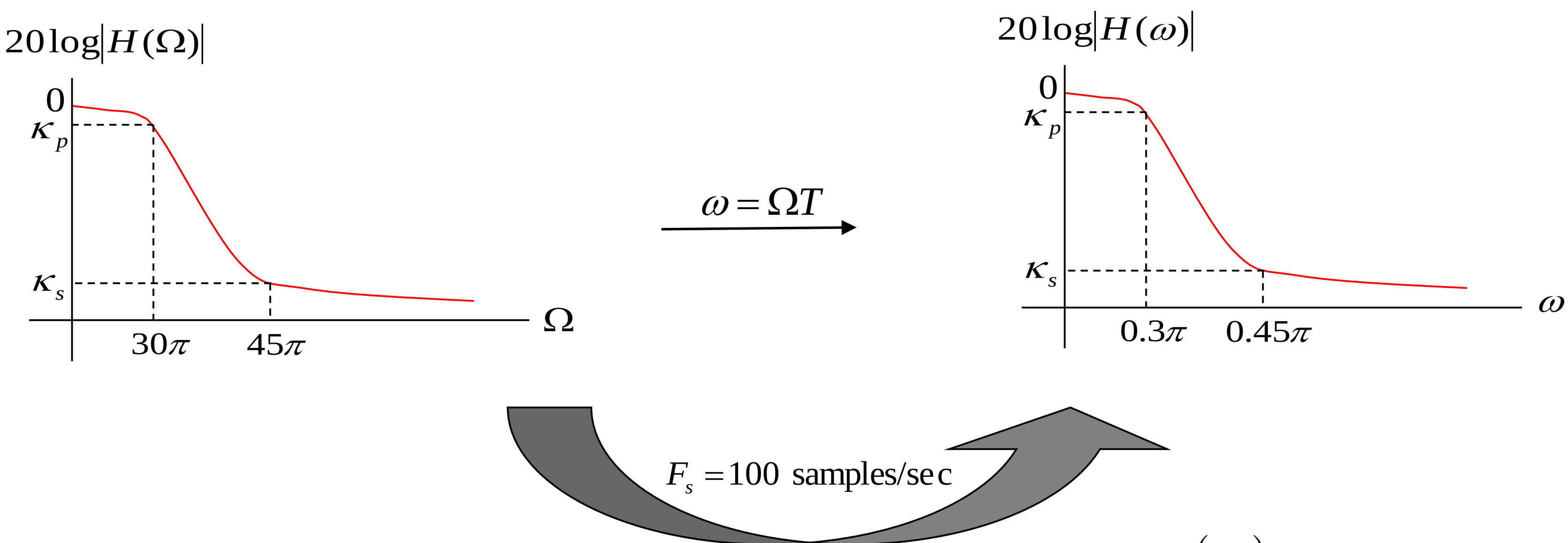
- A decrease in the sidelobe amplitude results in better stopband attenuation

To insure a sharp transition from passband to stopband, the window should have a very small mainlobe width – this is controlled by the window length.

To reduce the passband and stopband ripple, the sidelobes need to be small – this is controlled by the choice of window function.

We will illustrate the design procedure with an example:

1) Specify the filter using either an analog or digital frequency domain specification



$$\omega_p = 30\pi(.01) = 0.3\pi \text{ @ -3 dB}$$
$$\omega_s = 45\pi(.01) = 0.45\pi \text{ @ -40 dB}$$

2) Select a window type so that the stopband requirement is exceeded.

Remember these points:

- The stopband attenuation is relatively insensitive to the size of the window *N*, and the selection of ω_c. It depends primarily on the type of window used.
- The transition width of the filter is approximately equal to the mainlobe width of the window.

For this example we can select either the Hanning, Hamming, or Blackman window. The Hamming window (-53 dB) has lower sidelobes than Hanning (-44 dB) for the same transition width while the Blackman window (-74 dB) may be overkill which we would pay for with higher *N* to achieve the desired transition width. Therefore, let's use the Hamming window (-53 dB).

$$w(n) = .54 - .46\cos\left(\frac{2\pi n}{N-1}\right) \qquad 0 \le n \le N-1$$

3) Select *N* to satisfy the transition width

$$\Delta\omega = \omega_s - \omega_p \ge \Delta_{ML} \longrightarrow \Delta\omega \ge \frac{8\pi}{N}$$

We shall use $N = 54$.

with equality

$$.45\pi - .3\pi = \frac{8\pi}{N}, \quad \text{so} \quad N = 53.33$$

So select $N \ge 54$
(want *N* even or odd? Why?)

4) Compute ω_c and α:

$$\omega_c = \frac{\omega_s + \omega_p}{2} = .38\pi \qquad \alpha = \frac{N-1}{2} = 26.5$$

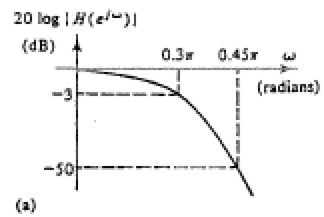
Therefore, the impulse response of the filter is:

$$h(n) = \frac{\sin[.38\pi(n-26.5)]}{\pi(n-26.5)} \cdot \left[.54 - .46\cos\left(\frac{2\pi n}{53}\right)\right] \qquad 0 \le n \le 53$$

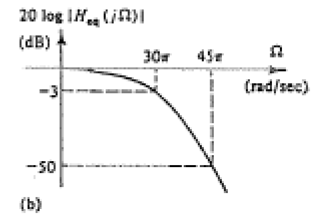
Verify design by plotting the frequency response (see next page)

If specifications are not met, iterate the design procedure by trial and error

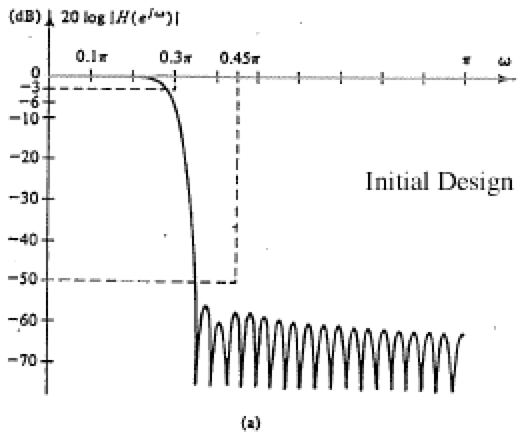
- First change ω_c (usually increase it) until the passband requirement is met.
- Next change *N*. Always try to keep *N* as small as needed to satisfy the specs.

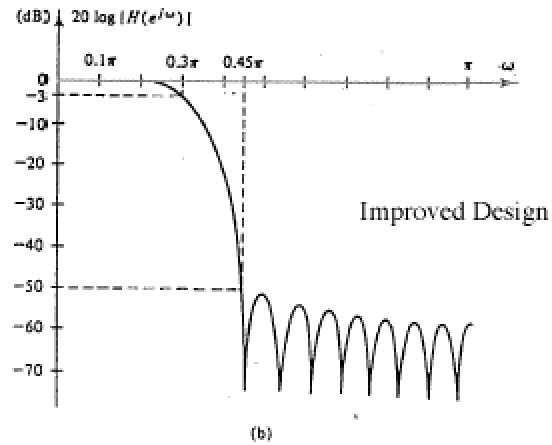


Digital Frequency Response



Analog Frequency Response





FIR Filter Design: Not based on traditional filter design techniques

- Window method
  - Simple and flexible, but does not allow adequate control over filter parameters
  - Good for tutorial and academic purposes
- Frequency Sampling method (we did not examine this technique)
  - Coefficients are obtained by taking inverse DFT of frequency response
  - Good for specialized frequency response design
- Optimal methods (we did not examine these techniques)
  - Method of choice for most practical designs
  - Attempts to achieve a best fit for the desired frequency response, within certain constraints
  - Often employ Computer-Aided Design (CAD).

Remember → **Filter Design = Coefficient Determination**

9.1 Choosing Between FIR and IIR Filters

- FIR filters can have exactly linear phase
- FIR filters are always stable. IIR filters could be unstable.
- Effects of finite wordlength are less severe in FIR than in IIR.
- FIR filter requires more coefficients for sharp cutoff filters than IIR.
- FIR filters are more difficult to synthesize if CAD support is not available.
- Analog filters can be readily transformed into equivalent IIR filters meeting similar specs. FIR filters have no analog counterparts; however, with FIR it is easier to synthesize filters with arbitrary frequency responses.

Broad Guidelines on when to use FIR and IIR Filters

- Use IIR when the only important requirements are sharp cutoff filters and high throughput. IIR filters generally require fewer coefficients than FIR.
- Use FIR if the number of filter coefficients is not too large and <u>if little or no phase distortion is required</u>.