Digital Signal Processing

Prof: S. C. Dutta Roy

Department of Electrical Engineering

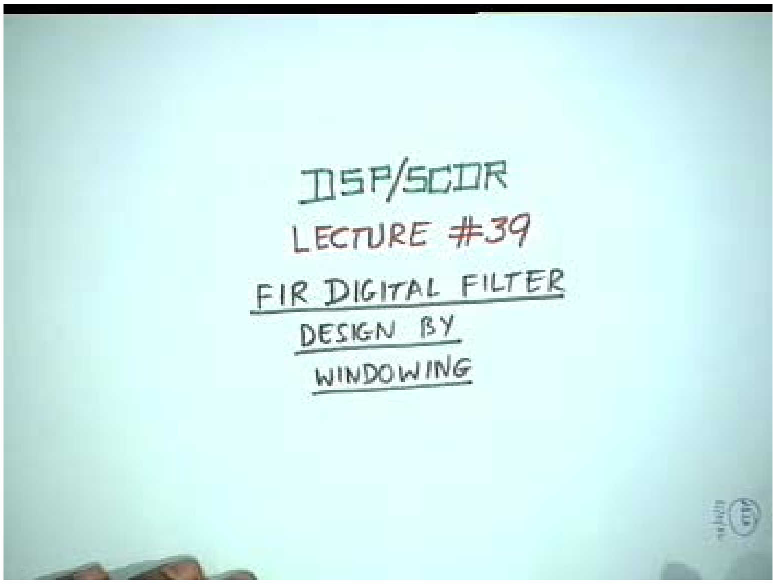
Indian Institute of Technology, Delhi

Lecture - 39

FIR Digital Filter Design by Windowing

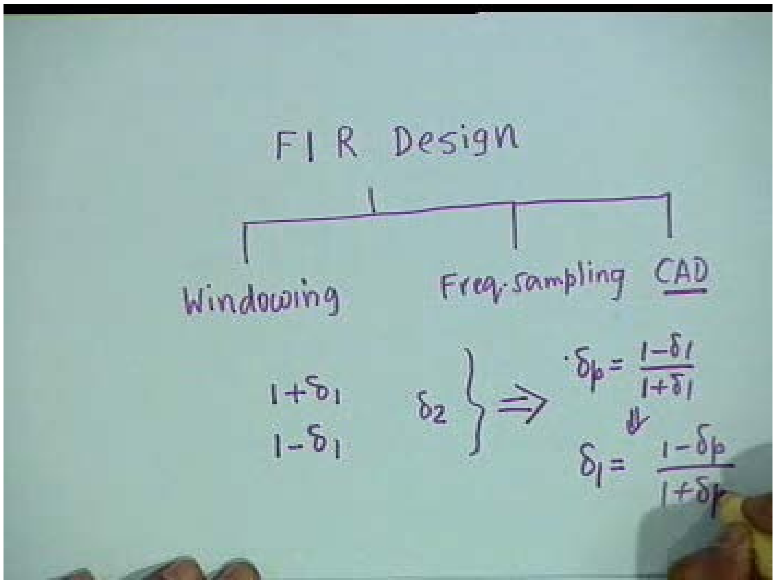
This is the 39th lecture and our topic for today is FIR Digital Filter Design by Windowing.

(Refer Slide Time: 01.11 - 01.14 min)



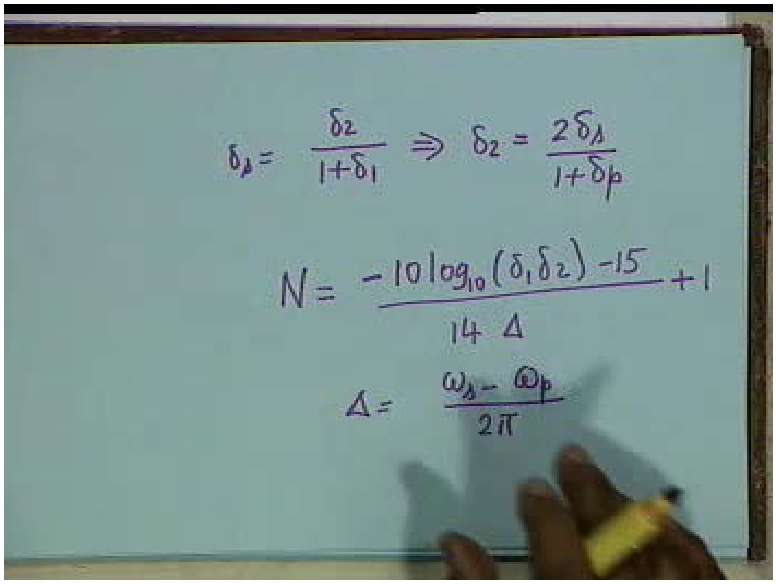
This is one of the most popular techniques for FIR filter design and we shall discuss this in some detail. In the last lecture, we had taken an example of Digital-to-Digital Transformation. We took a low pass filter and then transformed it to a high pass, band pass, and a band stop filter. And then we started the FIR design and we discussed the importance of FIR.

(Refer Slide Time: 01.45 - 03.56 min)



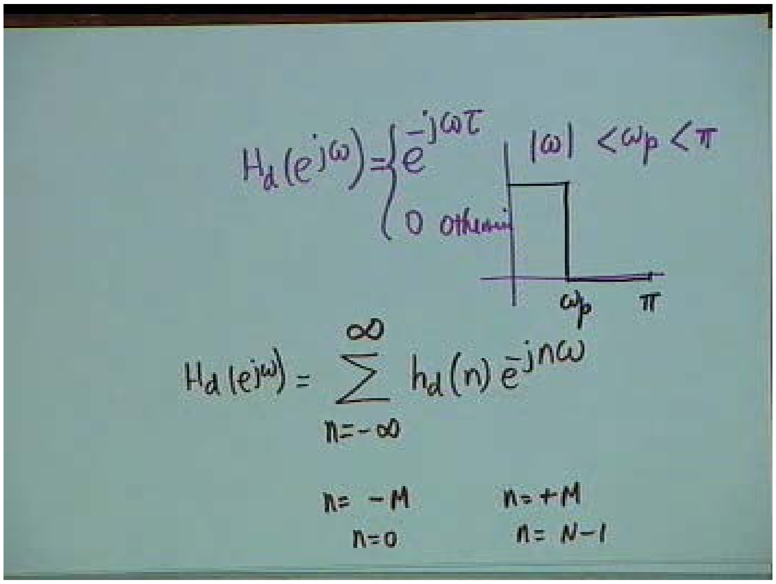
We also discussed the disadvantages of FIR design as compared to IIR. We said that only two analytical techniques are available, and that too, they are not completely analytical; they are semi-analytical. They only give you a rough idea of what it is going to be. You might have to undergo a number of iterations before you zero in on a particular design; one is windowing, the other is frequency sampling and the third, of course, in common with IIR design, is the computer-aided design techniques. The last one is very powerful; nevertheless, it is not analytical but is based on empirical procedures and algorithms. In FIR design, we also said that the terminology used is slightly different from what we have been used to. Here, the pass band varies between $1 + \delta_1$ and $1 - \delta_1$. The stop band transmission is δ_2 and we said that these can be easily transformed into our nomenclature δ_p and δ_s and vice versa. You normalize the highest response to 1; so $\delta_p = (1 - \delta_1)/(1 + \delta_1)$, which by compendendo and dividendo, gives $\delta_1 = (1 - \delta_p)/(1 + \delta_p)$.

(Refer Slide Time: 04.00 - 06.10 min)



The δs in our nomenclature shall be = $\delta_2/(1 + \delta_1)$ which gives δ_2 in terms of δ_p and δ_s, as $2\delta_s/(1 + \delta_p)$. The only formula available for estimating the order, unlike IIR design where N_B and N_C are given by analytical formulas, is emperical. There are complicated formulas also available, but there is no point in using them because, after all, you shall have to iterate a number of times if the number that comes by the formula does not suffice. The simplest formula, which is an approximation, is $N = \{[— 10 \log_{10} (\delta_1 \delta_2) – 15]/(14 \times \Delta)\} + 1$. And in terms of our terminology, $\Delta = (\omega_s – \omega_p)/(2\pi)$; this is for low pass. Obviously for high pass filters it has the same formula except that Δ will be $(\omega_p – \omega_s)/(2\pi)$. What the formula transforms into in the case of band pass and band stop shall be discussed later.

(Refer Slide Time: 06.28 - 10.49 min)



The philosophy of windowing is as follows. Given the required low-pass transfer function $H_d(e^{j\omega})$, since you cannot design the filter analytically, there is no point in starting from this. Therefore you aim at the ideal. In other words, you assume that the $H_d(e^{j\omega})$ magnitude is unity between 0 and ω_p and zero between ω_p and π. This freedom can be exercised because the design is not analytical; it is semi-analytical.

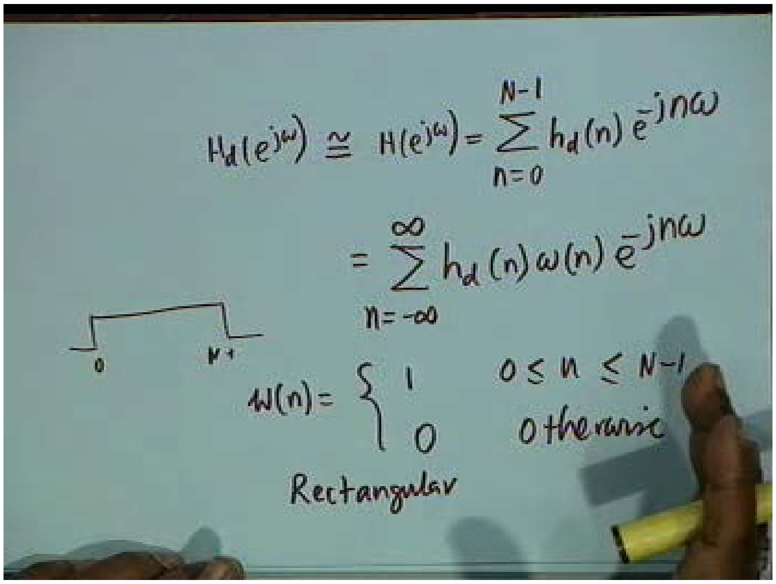
In the case of IIR design, we could start from δ_p and δ_s and then estimate the order and write the transfer function. Here we cannot do that, therefore we aim at the ideal transfer function. The FIR design is most useful when you have linear phase, and therefore what we assume is that $H_d(e^{j\omega}) = e^{-j\omega\tau}$, whose magnitude is 1, but the phase is linear, for $\omega \leq |\omega_p| < \pi$ and it is 0 otherwise; so this is our ideal response. We did not bring in δ_p, δ_s or δ_1, δ_2 because it is not useful to do so.

When we get the design we shall have to find the frequency response, check whether it satisfies the specs or not and if the answer is no, then go back. That is why the whole procedure is semi-analytical. So what we do is to expand $H_d(e^{j\omega})$ in Fourier series. And in general, the Fourier series is summation $h_d(n)\ e^{-jn\omega}$, $n = -\infty$ to $+\infty$. The Fourier series has infinite number of terms.

Since infinite number of terms gives rise to an IIR filter, i.e. $h_d(n)$ comes of infinite length $-\infty$ to $+\infty$, we arbitrarily truncate it at some point. Suppose we truncate it at n = – m on the left side and n = + m on the right hand side, then we get a filter of length 2m + 1, and it shall be a zero phase filter. It will be a zero phase filter because we have samples on the left and samples on the right. We cannot realize a zero phase filter because the filter is non-causal. h(n) must be = 0 for n < 0; otherwise we cannot realize.

Therefore after you obtain summation $h_d(n)\, e^{-jn\omega}$, n = –m to +m, we shall have to multiply the transfer function summation $h_d(n)\, z^{-n}$, n = –m to +m, by z^{-m} to make it causal. Why not start from n = 0 and go up to n = N – 1? Since we have assumed a linear phase, we start from the expansion of a linear phase transfer function; it is guaranteed that the impulse response $h_d(n)$ that we get shall be symmetric or anti-symmetric depending on what you want. You might want phase = $-j\omega\tau + \pi/2$; then it will be anti-symmetric. If $\pi/2$ is not present, then $h_d(n)$ will be symmetric. So there is no realizibility problem if you start from n = 0.

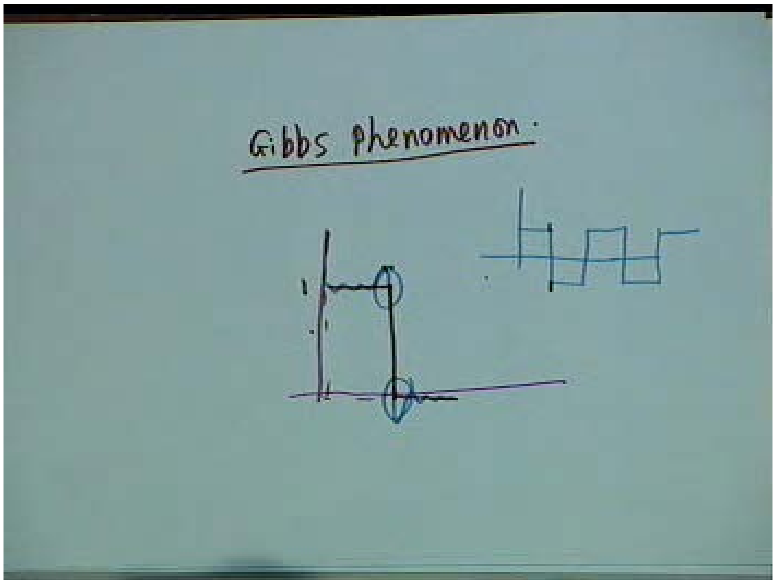
(Refer Slide Time: 11.01 - 13.30 min)



Therefore my $H_d(e^{j\omega})$ is approximated by $H(e^{j\omega})$ which is = $\sum_{n=0}^{N-1} h_d(n)\, e^{-jn\omega}$; it is as if the infinite length impulse response is allowed to pass through a window of rectangular shape. You can write this as $\sum h_d(n)\, w(n)\, e^{-jn\omega}$; n = – ∞ to + ∞ where w(n) is unity only between n = 0 and N – 1. That is w(n) = 1; $0 \leq n \leq N - 1$ or 0 otherwise. So it looks like a window of length N through which the impulse response passes. And this window is rectangular in shape, its amplitude is 1 for all values of n between 0 and N – 1 and 0 otherwise. Therefore only the impulse response within this region n = 0 to N – 1 passes and all the rest are made to vanish.

Now there is no guarantee that you will get a good approximation. The only thing you know is an estimate of N; this is the starting point and now you need to iterate. This is why this design is called Windowing Technique. In other words you expand in Fourier series and then starting from n = 0, truncate it at some point. In this particular case this window is rectangular in shape. Rectangular windows have a number of disadvantages. One of them is the so called Gibbs Phenomenon.

(Refer Slide Time: 14.00 - 17.20 min)

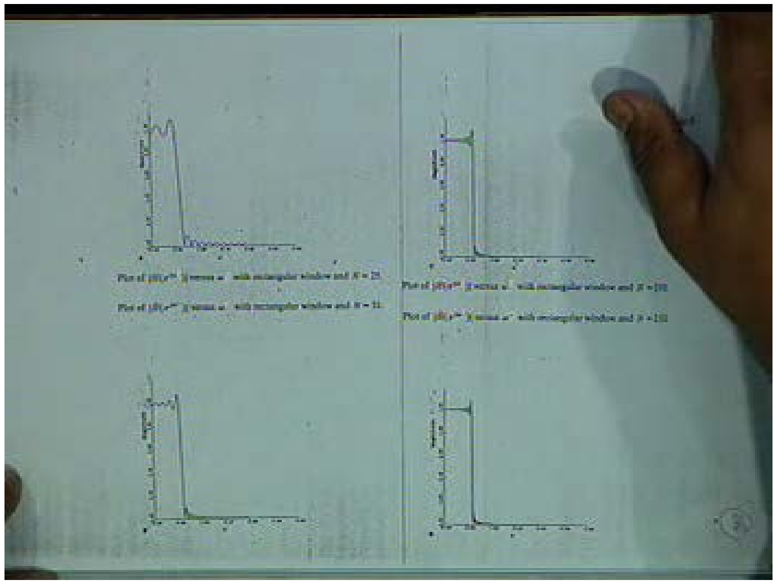


The discovery of this phenomenon is due to A. J. Gibbs. And Gibbs Phenomenon occurs whenever you want to approximate a discontinuity, like you have in any ideal filter. If you want to approximate a discontinuity by a finite number of terms, Gibbs phenomenon occurs. And the phenomenon is the following: There are oscillations around the points of discontinuity. In other words, the approximation is not smooth, but there are oscillations. And the last peak is important. The oscillations start building and shows the highest peak close to the point of discontinuity. The response comes down there and then shows an undershoot. The first minimum in the undershoot is the highest. This is called Gibbs Phenomenon. It occurs even when the number of terms tends to infinity and it is a very peculiar phenomenon.

In other words, the Fourier series is not an exact representation of a periodic function if the function has discontinuities. For a smooth function like a sinusoid corrupted by third harmonic or fifth harmonic, you will get a perfect approximation. But suppose the Fourier series is that of a function like a rectangular wave, then even if you take infinite number of terms it is still an approximation. And if you take an infinite number of terms you get a rod at this point of discontinuity jutting out in both directions. The amplitude of the rise or fall at this point does not depend on the number of terms.

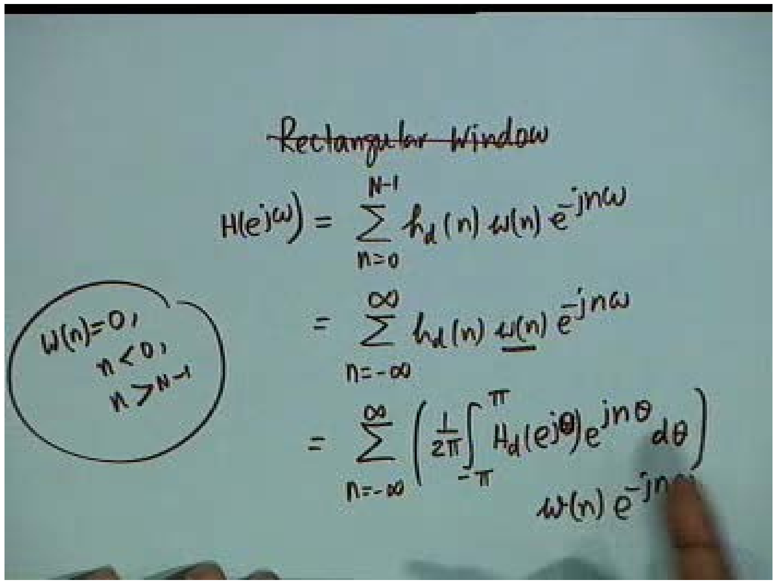
Gibbs phenomenon has many peculiarities and one of them is the amplitude. That is, if you take 5 terms, 25 terms, 35 terms, 100 terms and 10,000 terms the amplitude of the last oscillation before the discontinuity and the first oscillation after the discontinuity remain almost constant. For a sharp finite discontinuity, it is about 18%; actually the amplitude rises to 1.1785 just before the discontinuity, and this we take as 18% approximately. This is the reason why we do not use the rectangular window unless the specifications are so relaxed that 18% overshoot or undershoot can be tolerated. I shall project a diagram here which shows how the Gibbs phenomenon shows itself with increasing number of terms.

(Refer Slide Time: 18.05 - 19.30 min)



This is an approximation of a low pass filter. That is, we take $e^{-j\omega\tau}$ as $H_d(e^{j\omega})$ and then plot the magnitude for so many values of N. We have plotted the magnitude so you see undershoots also as overshoots. You notice that the value above unity in the last peak and the first peak after the discontinuity are almost the same and this is 0.178. The lengths chosen are 25, 51, 101 and 151. There are more oscillations as N increases because we are using higher and higher harmonics. But the amplitude before and after the discontinuity almost remain constant. When you go to infinite number of terms, oscillations will hardly be detectable but there would be a rod at this point of discontinuity, as mentioned earlier. This is what Gibbs oscillation is and therefore we require to do something about the window to reduce Gibbs oscillations.

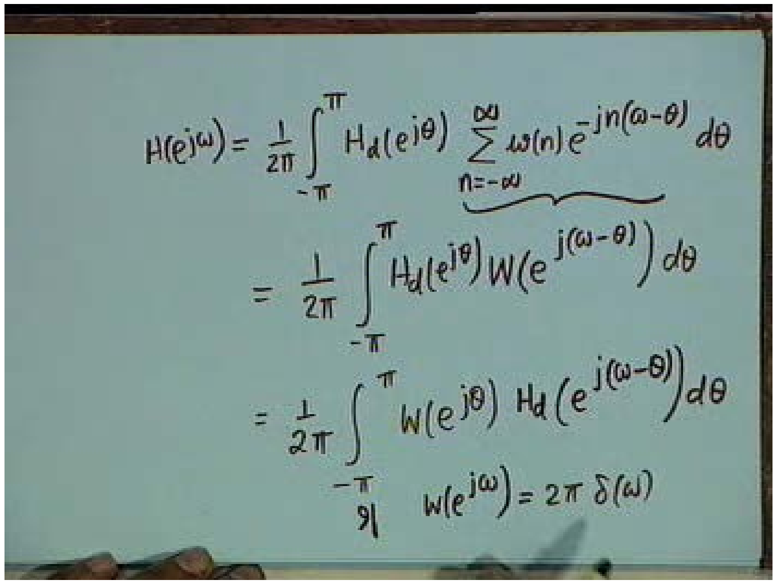
(Refer Slide Time: 19.40 - 23.56 min)



The rectangular window itself has sharp discontinuities; it suddenly rises at n = 0 to 1 and then suddenly falls. If the rise and the fall of the window are smooth, then perhaps we shall get better results, by reducing the Gibbs phenomenon. In other words, window requires to be tapered at n = 0 and at n = N – 1. But before we look at specific windows, let us find out what should be the ideal or the optimum window. We have $H(e^{j\omega})$, that is the approximation of $H_d(e^{j\omega})$, which is summation n = 0 to N – 1, $h_d(n)\, w(n)\, e^{-jn\omega}$.

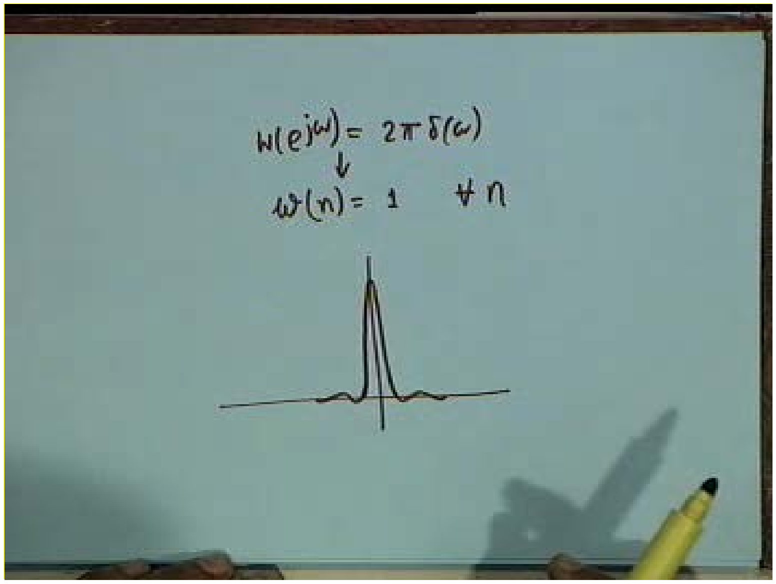
Let us not specialize w(n) to rectangular window to start with but we want to find what kind of window shall be the optimum. What are the characteristics of the window function? We need to have a finite impulse response h(n), which is $h_d(n) \times w(n)$; $h_d(n)$ is of infinite length, w(n) must be finite. I can write this as $\sum h_d(n)\, w(n)\, e^{-jn\omega}$, I take from n = – ∞ to + ∞. But I choose my w(n) such that I get an FIR. I must choose w(n) = 0 for n < 0 and n > N – 1 that is why these two are identical. But the second formulation now will help us to find out the optimum w(n). I can write this as $\sum_{n=-\infty}^{n=\infty}$ (I replace $h_d(n)$ by the inverse Fourier transform relationship) $[1/(2\pi)] \int_{-\pi}^{\pi} H_d(e^{j\theta})$ (Let us change the variable, because I have another $e^{-jn\omega}$; to θ. It does not matter because we are going to integrate with respect to θ) $e^{jn\theta}\, d\theta \times w(n)\, e^{-jn\omega}$. The ∑ is over n and the integration is over θ. Integration is over a continuous variable, $\sum$ is over a discrete variable; they are not related to each other so interchange the $\sum$ and integration.

(Refer Slide Time: 24.01 - 27.22 min)



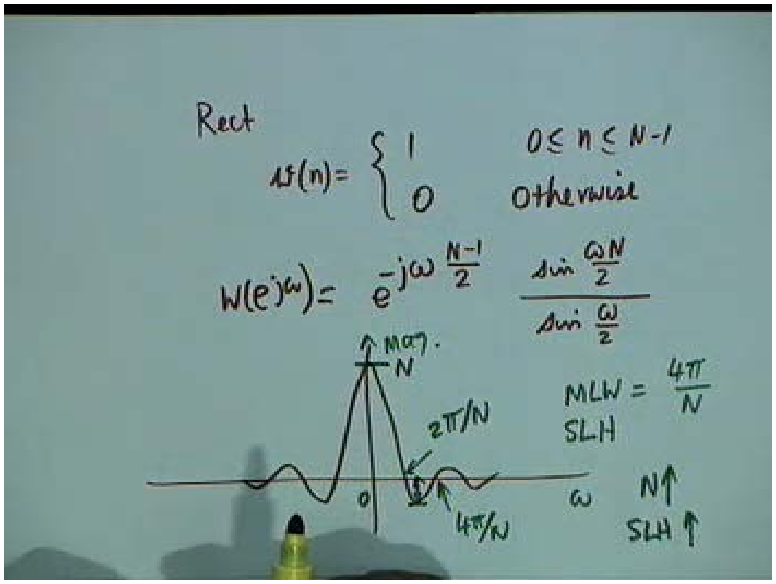
In other words, we write; $H(e^{j\omega}) = [1/(2\pi) \int^{\pi}_{-\pi} H_d(e^{j\theta}) \sum w(n)\ e^{-jn(\omega-\theta)}\ d\theta$, n goes from $-\infty$ to $+\infty$. If there was only $e^{-jn\omega}$, the summation would result in $W(e^{j\omega})$. Since ω is replaced by $\omega - \theta$, we shall have W $(e^{j(\omega-\theta)})$. So we get $H(e^{j\omega}) = [1/(2\pi) \int^{\pi}_{-\pi} H_d(e^{j\theta})\ W(e^{j(\omega-\theta)})\ d\theta$. If you look at this integration, is it not complex convolution? Therefore the realized frequency response is the complex convolution of the desired frequency response and the window frequency response. The integrand can also be written as $W(e^{j\theta})\ H_d(e^{j(\omega-\theta)})$ because convolution operation is commutative. Now this gives us a clue as to what the window function should be. What you want is $H(^{j\omega}) = H_d\ (e^{j\omega})$. Obviously, this is obtained when $W(e^{j\omega}) = 2\pi\ \delta(\omega)$; 2π is brought here because there is $1/(2\pi)$ before the integral. Suppose $W(e^{j\omega})$, the spectrum of the window function is $2\pi\ \delta(\omega)$, which exists only at $\omega = 0$. What does the integral become? It becomes $H_d(e^{j\omega})$. Therefore what we should aim for is a window function whose spectrum is a δ function.

(Refer Slide Time: 27.30 - 28.50 min)



That is not surprising because if $W(e^{j\omega}) = 2\pi\delta(\omega)$, then w(n) is 1 for all n. In other words, there is no truncation; therefore it becomes an IIR filter and this result is to be expected. But this way of deriving the result points out that what we want for the window function is one whose spectrum is an approximation to an impulse of strength 2π. All such approximations shall have side lobes and therefore we shall have something like that shown in the figure. In fact, a rectangular window has a spectrum like this except that we do not want it because there is a lot of Gibbs phenomenon, and we want to smooth it out. So what is an optimum window function? There is no optimum, we need to have a window function which approximates an impulse in the frequency domain, this is our aim. With that end in view various windows have been tried and we shall go through this list and their performance one by one.

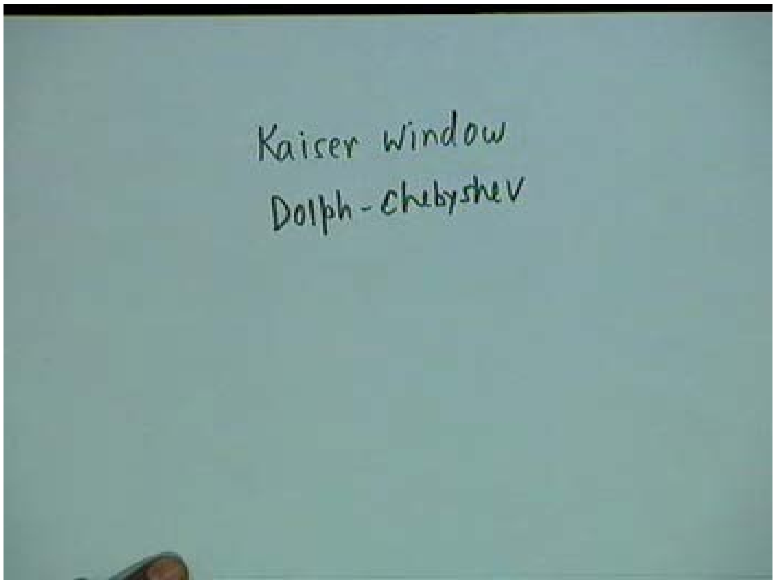
(Refer Slide Time: 29:25 - 32:30 min)



First let us look at the rectangular window, that is w(n) = 1 for $0 \leq 1 \leq N - 1$ and 0 otherwise. The spectrum is $W(e^{j\omega}) = e^{-j\omega(N-1)/2}\sin(\omega N/2)/\sin(\omega/2)$ and the spectrum is something like that shown in the slide. The first 0 shall occur at $\omega = 2\pi/N$ and not at $\omega = 0$ because at $\omega = 0$, the value is N. The next 0 shall occur at $4\pi/N$ and so on. So in terms of this spectrum what you want is that the Main Lobe Width (MLW) which is $4\pi/N$ should be as small as possible. Since we want an approximation to the impulse function, the Side Lobe Height (SLH) should be as small as possible.

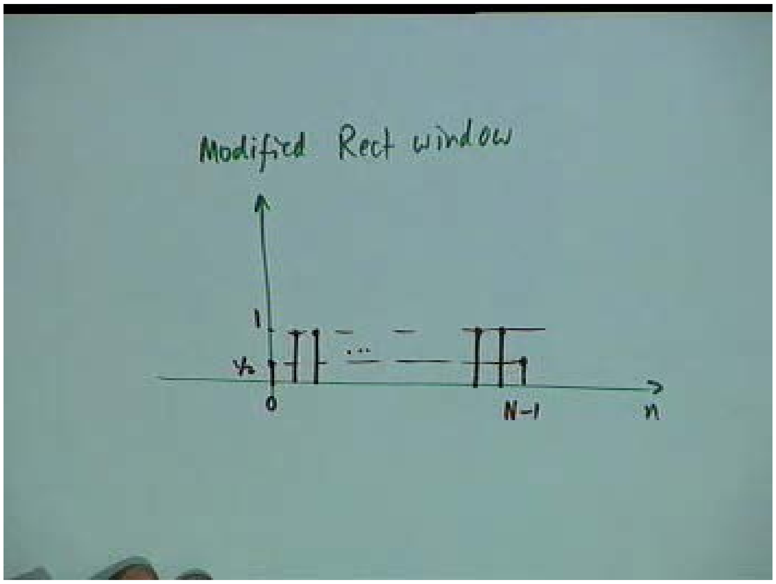
Unfortunately, these two requirements are contradictory. That is, if you want to decrease the main lobe width, then N should increase; as N increases it shrinks, but at the same time the side lobe height increases. The ratio MLW/SLH is approximately a constant. This is the problem in FIR filter design. Whatever window you choose, it would be a compromise between main lobe width and the side lobe height and there is hardly much of a choice except two windows which we shall not discuss in detail in the class; one is the Kaiser window and the other is Dolph Chebyshev Window, the idea of the latter being taken from Antenna Array Design.

(Refer Slide Time: 32:35 - 33:32 min)



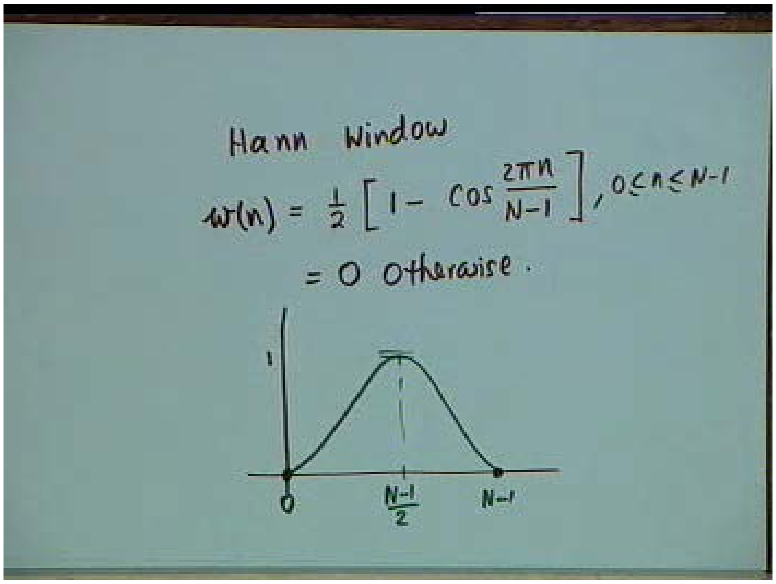
Kaiser window involves Bessel functions of order zero. And normally, if it is for sophisticated designs, we shall use Kaiser Window but for ordinary applications other simpler windows which are easy to calculate and incorporate in design are used. Let us look at some of the simpler windows. One of them is the Modified Rectangular Window which says that instead of starting from 1 why do we not start from 1/2? It is an attempt to taper the window.

(Refer Slide Time: 33:37 - 34:43 min)



We start from ½, then all other samples are 1, except the (N – 1)th or the last one which is also 1/2. This is a modified rectangular window, in which a taper has been introduced. This, as expected, reduces the side lobe but increases the main lobe width.

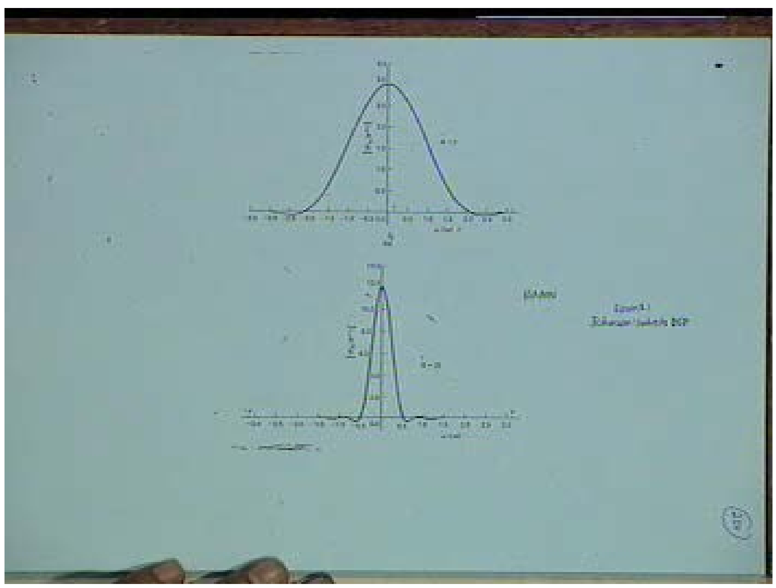
(Refer Slide Time: 34:45 - 36:46 min)



The next window for our discussion is the Hann Window, which is a smooth window; there are no abrupt discontinuities. It is $\frac{1}{2}\,[1 - \cos 2\pi n/(N - 1)]$ for $0 \leq n \leq N - 1$ and 0 otherwise. If you plot it, it looks like a cosine wave and is shown in the slide. For $n = 0$, the value is 0, and for $n = N - 1$, again the value is 0. The maximum occurs when the angle $2\pi n/(N - 1)$ is $= \pi$, so that the maximum value is 1. It occurs at $n = (N - 1)/2$. Obviously an odd N is to be preferred.

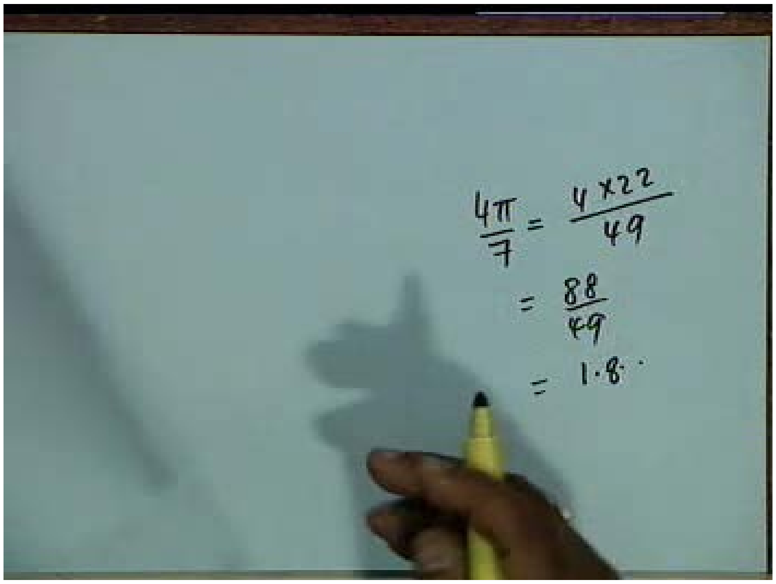
There is also another reason as to why N odd should be preferred. It is because the delay is an integer, and a half delay is not very easy to accommodate in a DSP. What happens if we apply such a window? The result is something like the one shown in the next slide.

(Refer Slide Time: 36:48)



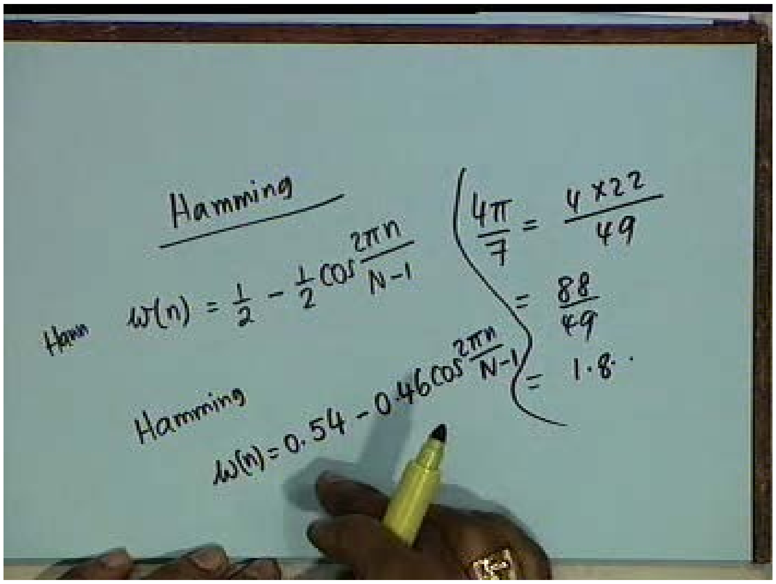
We have plotted for $N = 7$ and for length = 25. When we increase the length then the main lobe width shrinks but the side lobe height increases.

(Refer Slide Time: 37:29 - 37:52 min)



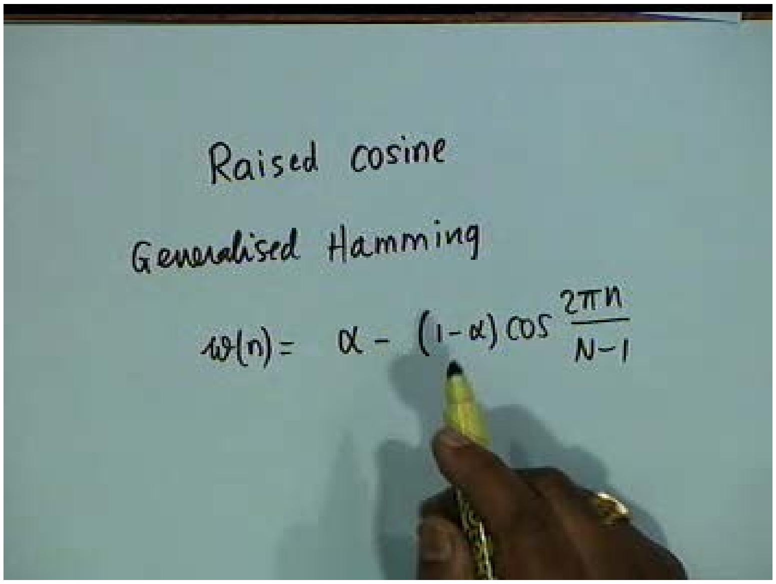
A point to notice about Hann window is that effectively the length is N – 2 because two of the samples are 0. You have not been able to utilize the efforts you have put in aiming for the length N; the effective length becomes N – 2. The next window that we consider is the Hamming window.

(Refer Slide Time: 39:09 - 40:54 min)



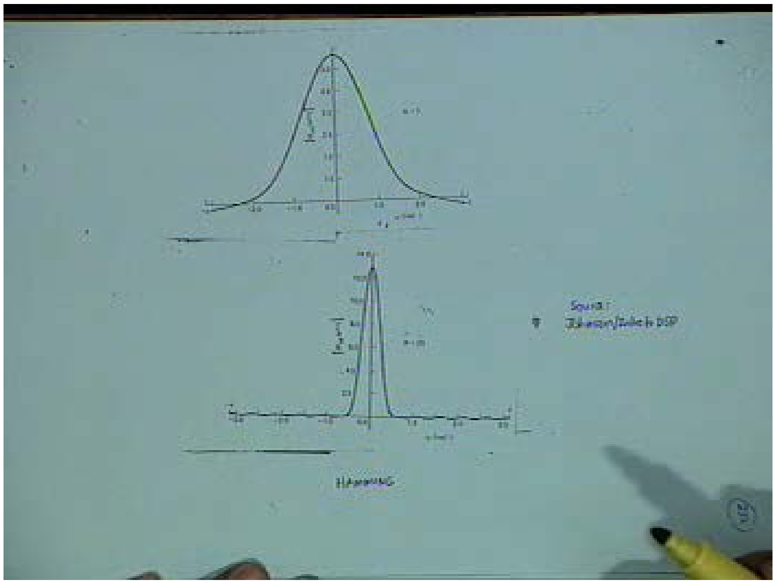
Hann window is (1/2) – (1/2) cos 2πn/(N – 1). Instead of the first 1/2, Hamming window uses 0.54 for the first term. Naturally for the second 1/2, you have to use 0.46; then only the maximum value becomes 1. The advantage of this is that you are utilizing the full length window. At n = 0, w(0) = 0.08 and this is also same as the w(N – 1). So instead of raising from the base of zero it is a cosine shaped wave form, but it has been raised by the amount point 0.08. So, Hamming window is also called Raised Cosine Window.

(Refer Slide Time: 40:56 - 43:27 min)



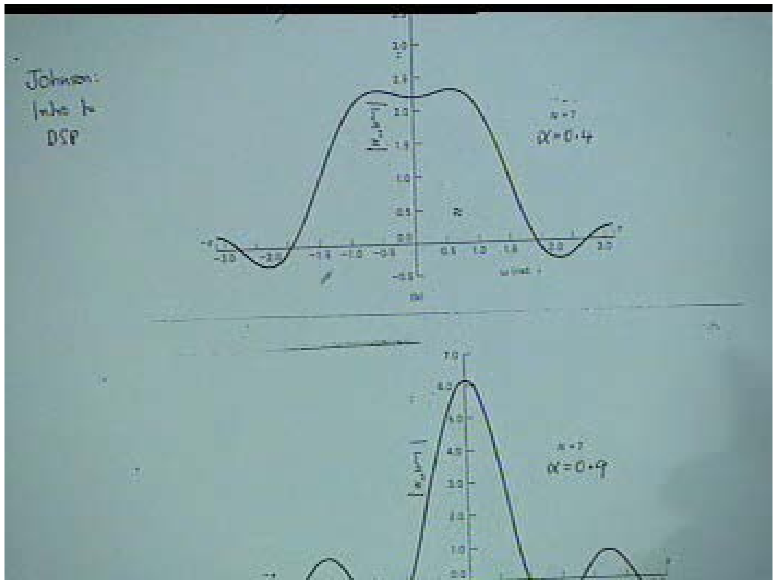
The effect of increasing the length of the window is very similar to Hann window except that Hamming allows for a little more reduction in side lobe height.

(Refer Slide Time: 41:19 - 42:38 min)

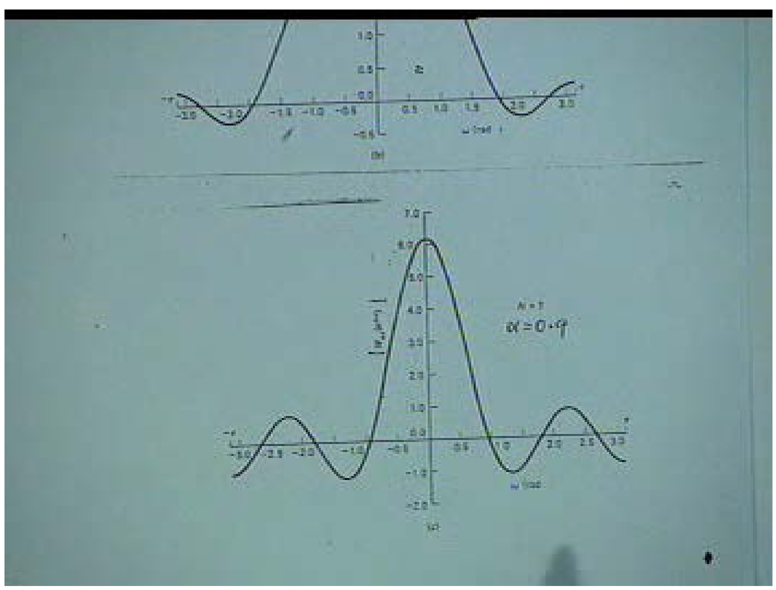


For N = 7, you notice that the main lobe width has increased beyond 4 radians, (refer slide). It can be decreased to approximately 1 if you raise the length to 25 but then the side lobe height also increases. It is always a compromise between main lobe width and side lobe height. You try the windows and whatever is acceptable, you use it. These windows are simple because they are very easy to calculate. On the other hand, Dolph Chebyshev uses Chebyshev functions, so you require a table or you have to calculate it every time. Similarly the Bessel functions also need to be calculated. They are tabulated but not for all values. Obviously they cannot be tabulated for continuous values. But cosine function is very easy to calculate. The next window we consider is the so called Generalized Hamming. Generalized Hamming window is $\alpha - (1 - \alpha) \cos 2\pi n/(N - 1)$. Now you vary α to suit your requirements. If we vary α then the spectrum changes shape like that shown in the next slide. However, $\alpha = 0.5$ appears to be a good compromise, and nothing substantial is gained by varying α.

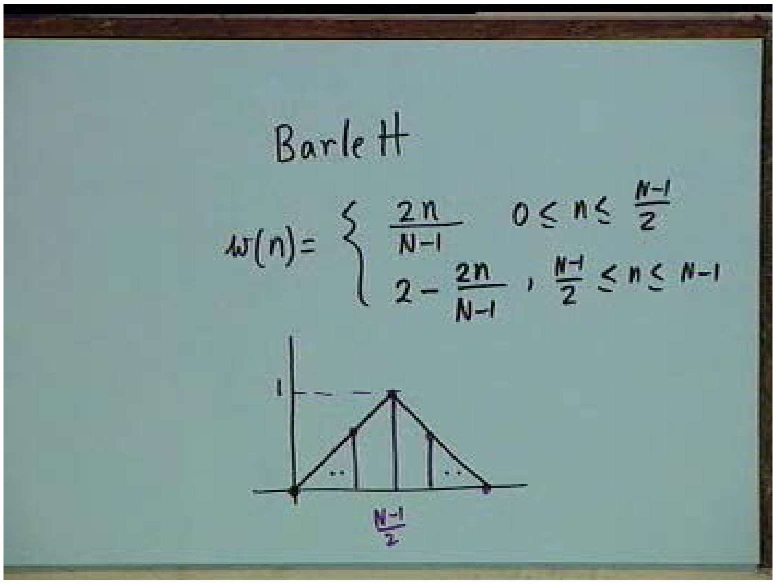
(Refer Slide Time: 44:08)



(Refer Slide Time: 44:33 min)

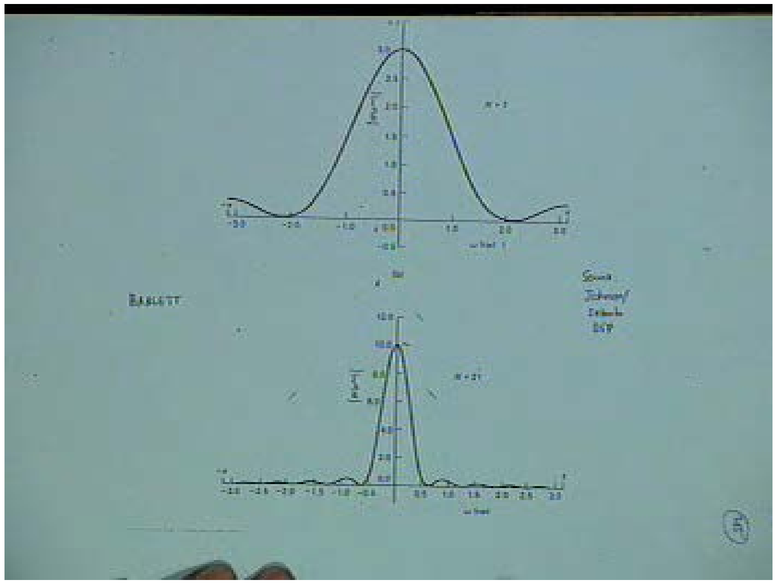


(Refer Slide Time: 44:57 - 46:51 min)



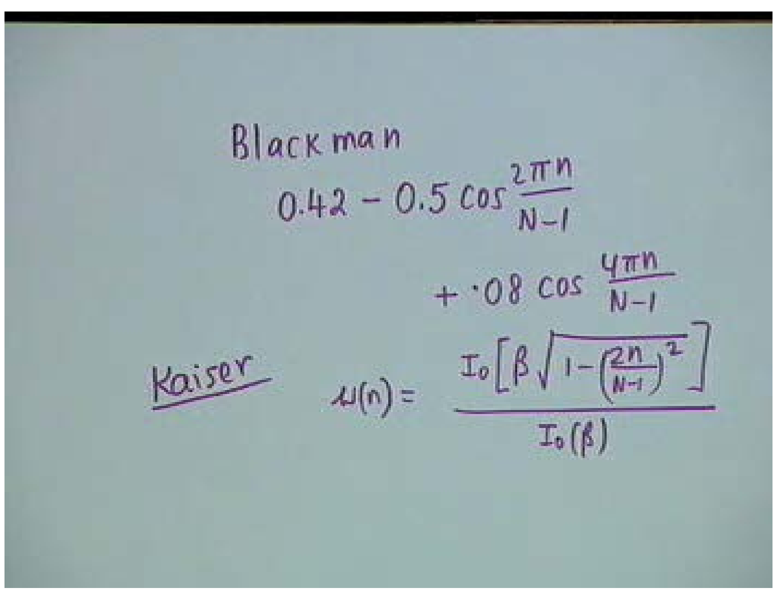
Bartlett suggested a simple window, called the triangular window. In this, w(n) = 2n/(N – 1), 0 ≤ n ≤ (N – 1)/2 and [2 – 2n/(N – 1)] for (N —1)/2 ≤ n ≤ N —1. The value rises along a straight line and falls along a straight line, after reaching a maximum of unity at n = (N – 1)/2; the calculation is very simple. But in common with the Hann window, it has the disadvantage that the end samples are zero. It is called Linear Window or Triangular Window or a Bartlett Window. There is one distinct feature of Bartlett window, namely that the spectrum is always positive. That is, the pseudo-magnitude is positive and it does not undershoot. In all the figures we saw so far, pseudo-magnitudes go positive as well as negative.

(Refer Slide Time: 47:14 - 47:49 min)



When N is increased from 7 to 21, once again the same phenomena occurs, that is, the side lobe height increases but the main lobe width decreases. There is nothing much to choose between Hann and Bartlett. In Hamming, you are utilizing the full length. Many other windows have been proposed and they are still being proposed. One is the Blackman window.

(Refer Slide Time: 48:04 - 51:17 min)

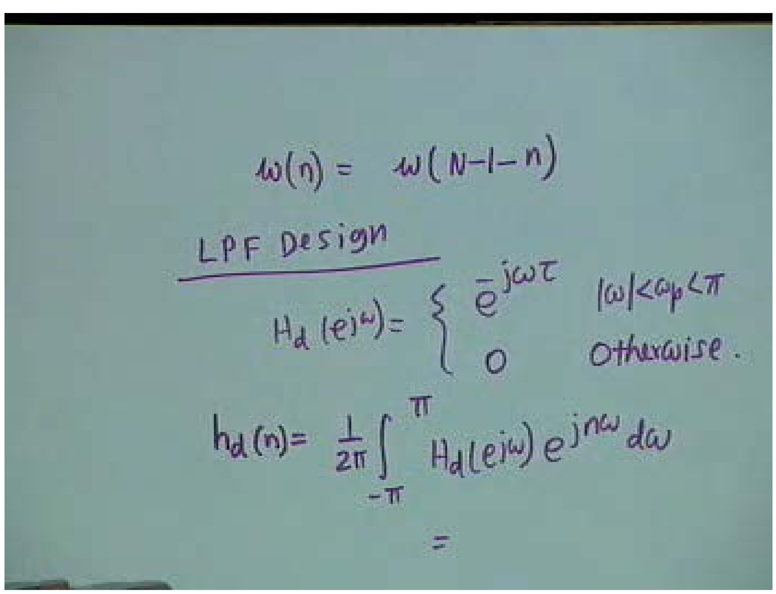


Blackman suggested that we use one cosine and the second harmonic also. He suggested using $0.42 - 0.5 \cos [2\pi n/(N - 1)] + 0.08 \cos[4\pi n/(N - 1)]$. So the maximum still remains 1. When $n = (N - 1)/2$, $w(n)$ becomes maximum, equal to 1. There is no reason why you cannot extend the series further. But then this is not worth doing because there is a kind of an uncertainty relationship between the main lobe width and the side lobe height. If one improves, the other deteriorates. And this is a reflection of Heisenberg's famous uncertainty principle. It shows up in many situations in electrical engineering.

For example, if a function is time limited, it cannot be band limited. The more it is time limited the more is the spread in the frequency. It shows its teeth in amplifier rise time and bandwidth. The smaller the rise time, the larger is the bandwidth that you require. The Dolph Chebyshev window requires Chebyshev functions.

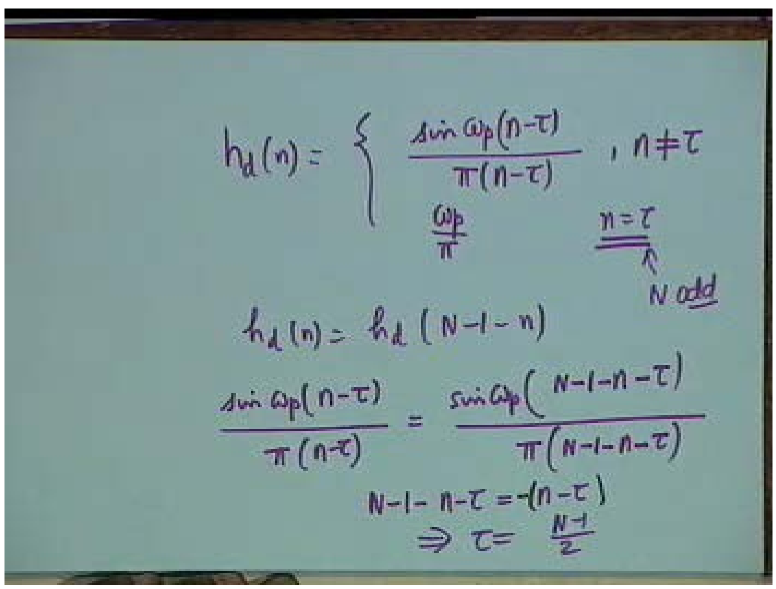
The Kaiser window uses the Bessel functions and the relationship is $w(n) = I_o . [\beta \sqrt{(1 - [2n/(N - 1)]^2)} / I_o (\beta)$ It is not simple to compute I_o. You notice that in all these windows there is something that we took care of, i.e. w(n) was taken as a symmetrical window.

(Refer Slide Time: 51:25 - 53:33 min)



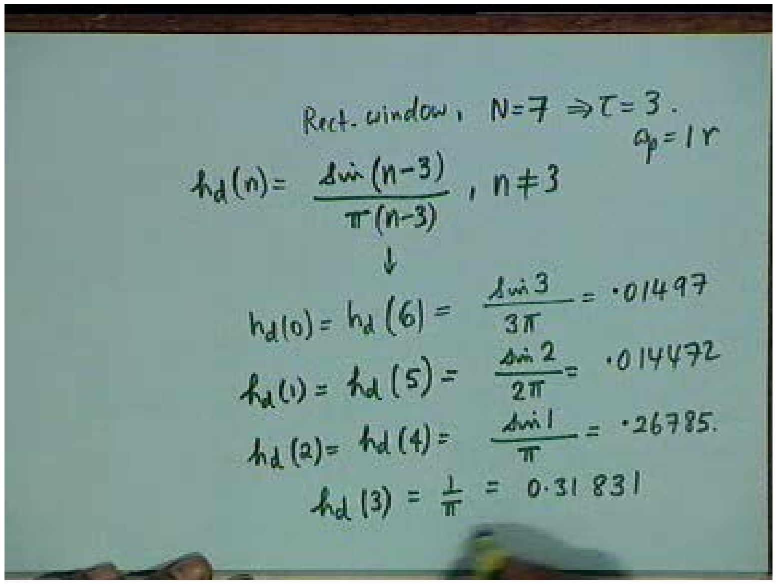
We have taken w(n) = w (N – 1 – n) because of linear phase constraint. If you want linear phase, $h_d(n)$ is symmetrical; otherwise linear phase shall be not be maintained. We take the specific case of FIR low pass design. What we want is H_d $(e^{j\omega}) = e^{-j\omega\tau}$ for $|\omega| \leq \omega_p \leq \pi$ and 0 otherwise; this is the ideal low pass filter. Obviously we have fixed our N. What should be τ? τ is (N – 1)/2. Once τ is given, you have no choice and the only thing you can play with are the window functions. If you find $h_d(n)$ corresponding to this, you have to use the inverse Fourier formula that is [1/(2π)] $\int_{-\pi}^{\pi} H_d(e^{j\omega})\ e^{jn\omega}\ d\omega$. And if you substitute for H_d $(e^{j\omega})$, the lower limit shall be substituted by – ω_p and the upper limit will be + ω_p. It is a very simple integration.

(Refer Slide Time: 53:36 - 56:57 min)



The result is: $h_d(n) = \sin [\omega_p (n - \tau)]/[\pi(n - \tau)]$ provided n is ≠ τ. If n = τ then obviously this will be ω_p/π. When is this possible? this is possible only when τ is an integer and therefore N is odd. We started with linear phase and said that τ = (N – 1)/2. You can prove that if we start with $e^{-j\omega\tau}$, then τ must be equal to (N – 1)/2; it is very simple. Our requirement is $h_d(n) = h_d$ (N – 1 – n). Therefore $\sin [\omega_p(n - \tau)]/[\pi(n - \tau)]$ should be = $\sin [\omega_p(N - 1 - n - \tau)]/[\pi (N - 1 - n - \tau)]$. They should be equal independent of the value of n. $N - 1 - n - \tau = -(n - \tau)$ and this gives $\tau = -(N - 1)/2$.

(Refer Slide Time: 57:14 - 1:00:00 min)



Therefore if we use a rectangular window then $h_d(n)$ of length 7 means $\tau = 3$ and let $\omega_p = 1$ radian. We shall work out this example with various windows and see what the effect is. Suppose ω_p is 1 radian, then $h_d(n) = \sin(n-3)/[\pi(n-3)]$, $n \neq 3$. And with rectangular window $h_d(n)$ is same as $h(n)$. So $h_d(0)$ shall be same as $h_d(6) = \sin(3)/(3\pi) = 0.01497$. $h_d(1)$, the same as $h_d(5)$, should be equal to $\sin(2)/(2\pi)$ and that comes out as.014472 and $h_d(2) = h_d(4) = \sin(1)/\pi$. This comes as 0.26785. And finally $h_d(3)$ is $1/\pi$ that is 0.31831. Next time we will show how the frequency response looks like with this kind of a window.