Digital Signal Processing

Prof. S. C. Dutta Roy

Department Of Electrical Engineering

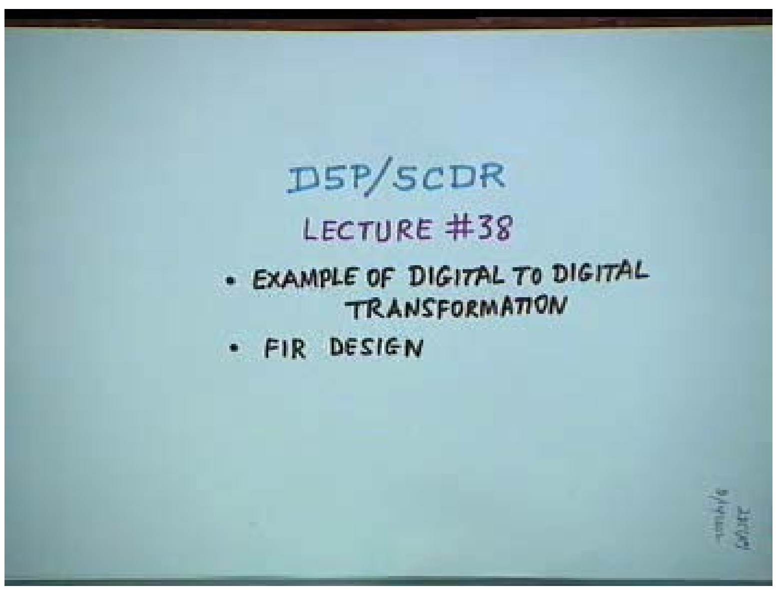
Indian Institute of Technology, Delhi

Lecture – 38

FIR Design

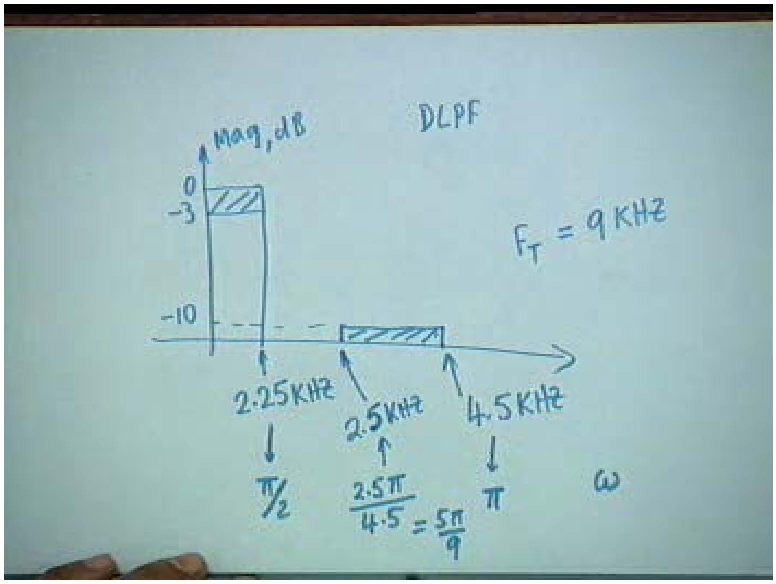
This is the 38th lecture. We take an example of Digital to Digital Transformation and then enter into the domain of FIR design.

(Refer Slide Time: 01:11 -01:24)



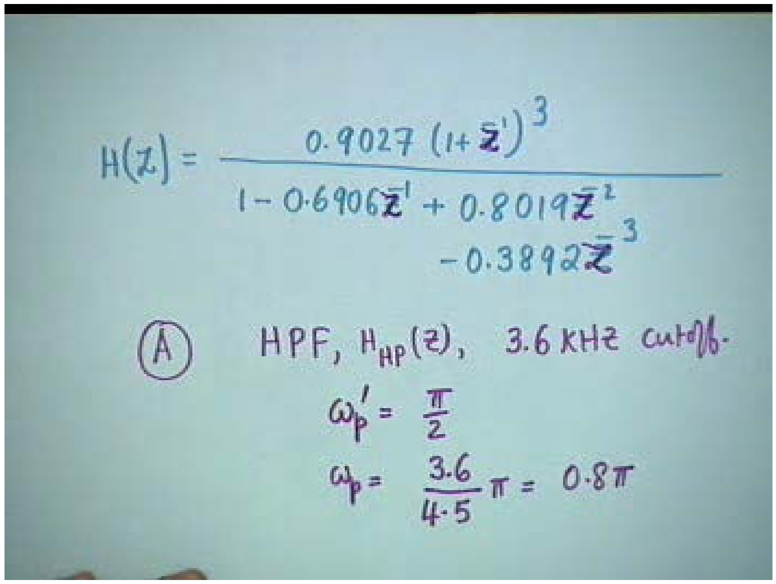
In the last lecture, we took the example of an IIR band stop filter design and discussed about digital to digital frequency transformation. We also saw that these transformations can be derived in a very simple manner if you know the analog transformations. As an example of application of Digital to Digital Transformation, we consider the low pass filter with tolerances as shown in the next slide.

(Refer Slide Time: 01:59 -04:48)



This is to be the prototype filter which we shall transform to various other kinds of filters. The passband tolerance is – 3dB and the passband edge is 2.25 kHz; this is how the actual specifications are given and you have to convert each frequency to its normalized digital frequency before you start designing. The stopband starts at 2.5 kHz and the sampling frequency F_T is given as 9 kHz and therefore the end of base band is 4.5 kHz. And you also notice that in terms of ω, 2.25 kHz corresponds to π/2 because 4.5 kHz corresponds to π. And 2.5 kHz corresponds to 2.5π/4.5 = 5π/9; also Δs is – 10dB. By the usual procedure, you find the normalized analog filter and then convert the analog normalized filter to digital low pass filter by the Bilinear Transformation and get a transfer function, which satisfies these specs. Using this as our prototype, we are going to use now the Digital to Digital Transformation instead of the BLT.

(Refer Slide Time: 05:01 – 10:08)

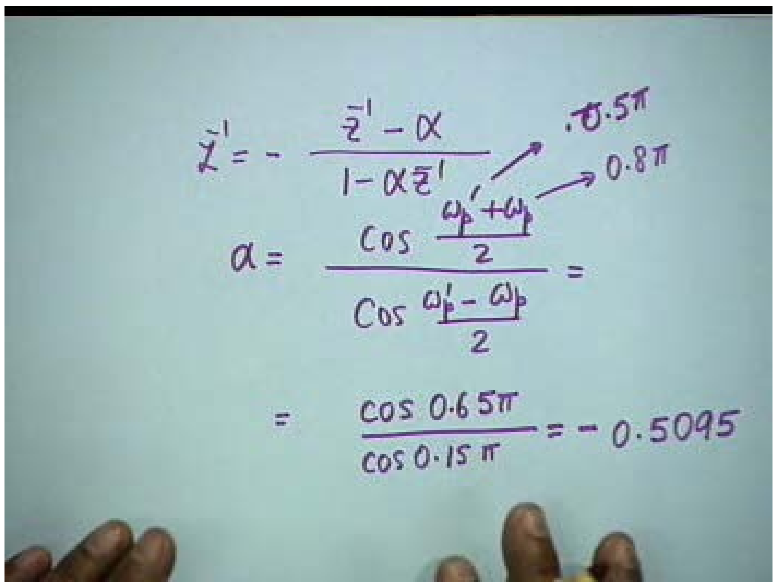


For the prototype digital LPF we use Z for complex frequency and z for the corresponding variable in the transformed filter. Recall that in analog case also, the complex frequency variable was transformed from S to s. So from Z we shall go to z for other kinds of filters. The prototype LPF is of 3rd order and we attempted Chebyshev to get the transfer function $H(z) = 0.9027 (1 + Z^{-1})^3/(1 - 0.6906 Z^{-1} + 0.8019 Z^{-2} - 0.3892 Z^{-3})$. Sometimes this transfer function is provided to you; otherwise you have to find it out.

From this we shall derive other kinds of filters. I wanted to illustrate the Digital to Digital Transformation so I did not write the intermediate steps. We saw two alternative routes: you can go from digital specs to analog specs and to analog normalized low pass filter, and then transform it to analog other kind. Finally, you use bilinear transformation. Here, you interchange these two steps.

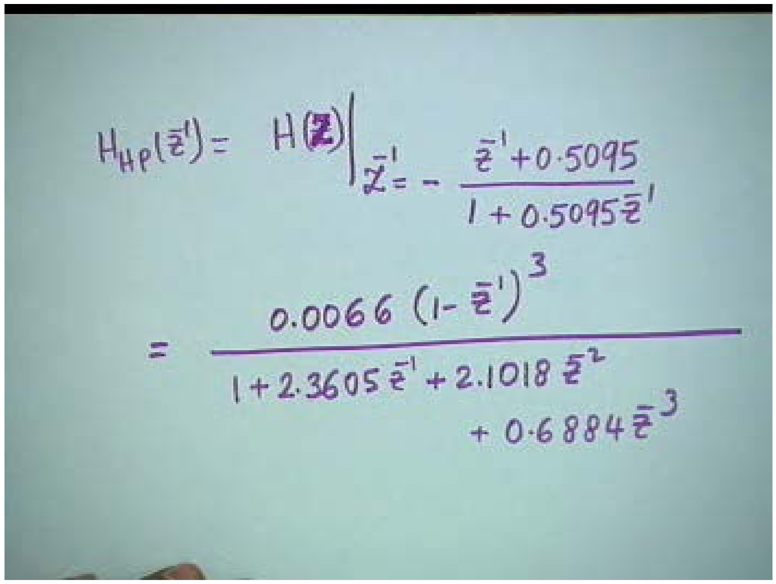
That is, to the normalized analog filter you apply Bilinear Transform; so you get a low pass digital filter whose ω_p' is not normalized. Then you apply Digital to Digital Transformation; this is the alternative route. But the advantage is that if you have standardized the prototype filter, then you can obtain any other kind of filter with the same tolerances without going through the process of analog filter again. That is, we simply replace Z^{-1} by the appropriate function of z^{-1} and that is the advantage. We design three filters. First we design a high pass filter $H_{HP}(z)$ with cutoff at 3.6 kHz. Now, ω_p' for low pass filter is equal to $\pi/2$. For the HPF to be designed, ω_p obviously is $(3.6/4.5) \times \pi = 0.8\pi$.

(Refer Slide Time: 10:13-11:41)



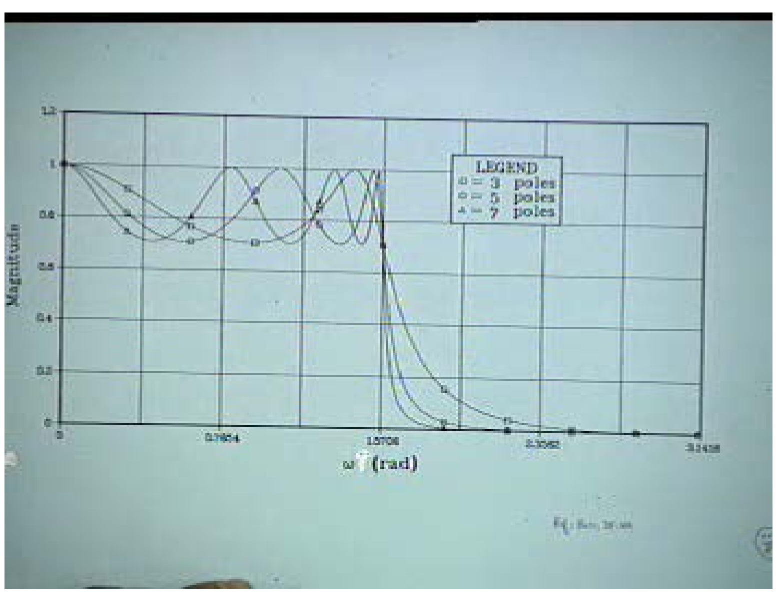
Our transformation from low pass to high pass, as we know is $Z^{-1} = -(z^{-1} - \alpha)/(1 - \alpha z^{-1})$ where the only parameter to be calculated is α. Here α is cos $[(\omega_p + \omega_p')/2]/\cos[(\omega_p' - \omega_p)/2]$. In this case ω_p' is 0.5π and ω_p is 0.8π, therefore α becomes $\cos 0.65\pi/\cos 0.15\pi = -0.5095$.

(Refer Slide Time: 11:44 -16:05)



Now we have to replace Z; we get $H_{HP}(z) = H(Z)$ under the condition $Z^{-1} = -(z^{-1} + 0.5095)/(1 + 0.5095\ z^{-1})$ and the result comes as $H_{HP}(z) = 0.0066(1 - z^{-1})^3/(1 + 2.3605\ z^{-1} + 2.1018\ z^{-2} + 0.6884\ z^{-3})$. Let us see first how the prototype filter response is like; it is a Chebyshev filter whose response is shown in the next slide.

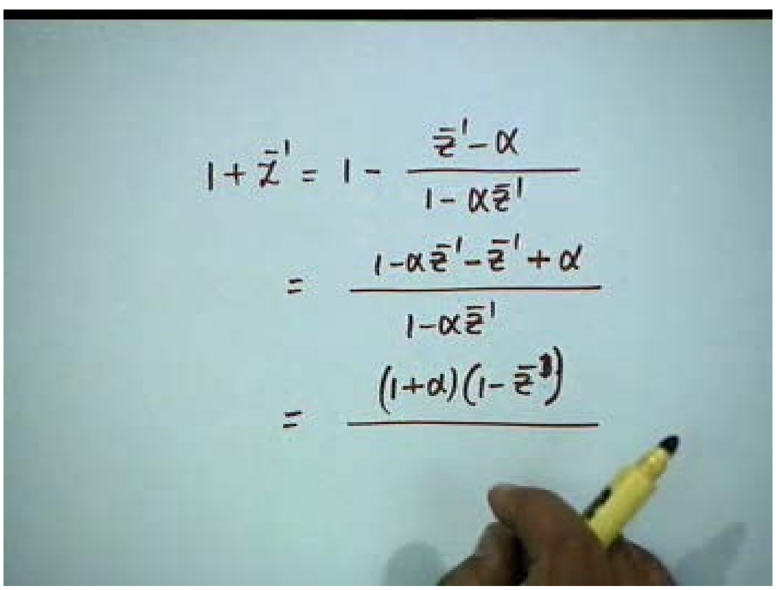
(Refer Slide Time: 13:23 -15:21)



There are three different responses: Minimum order is 3 and that is shown by the rectangles. Please see the equal ripple in the pass band, between 0.707 and 1. There are three peaks and dips. On the other hand, for an order of 5, the number of oscillations increases; for order = 7, it increases further. What is the argument in favor of drawing this picture? We want to illustrate the following. If we increase the order we over satisfy the stopband whereas order = 3 may be just enough. Usually, one designs higher order filters also, because the 3rd order may not be enough due to quantization errors.

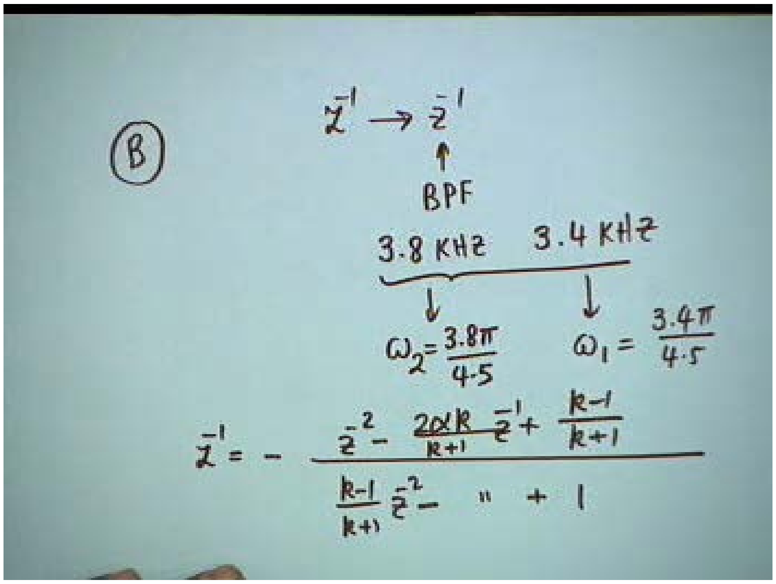
Now all the three filters in the figure are of odd orders, so that all of them start from magnitude = 1 at d.c. With the 5^{th} & 7^{th} orders, we increase the cost but we provide better stopband rejection. There is no change in the pass band tolerance; it is only in the stop band we have over satisfied. Also, the number of ripples increases as the order increases. Now there are several things to notice about this filter. One thing is that this is a high pass filter because at z = 1, the magnitude is 0. How does $(1 - z^{-1})^3$ come? Our numerator for the prototype LPF was $(1 + z^{-1})^3$; how did this convert into this?

(Refer Slide Time: 16:11 – 17:43)



It happens because the transformation is an all pass function and $1 + Z^{-1} = 1 - [(z^{-1} - \alpha)/(1 - \alpha z^{-1})]$ whose numerator can be written as $(1 + \alpha)(1 - z^{-1})$ and when you raise it to the power 3, the numerator becomes $(1 - z^{-1})^3$. It is a property of the Digital to Digital Transformation and this property is the reflection of the all pass nature of the transformation because Z^{-1} to z^{-1} is all pass and this will happen. For example, in band pass filter the numerator will be $(1 - z^{-2})$ and because it is third order it will be raised to the power 3. But in a band stop filter, you cannot say this. Band stop is rejection at some intermediate frequency so it would be slightly more complicated. It is good to examine every result critically as to whether you have made a mistake or not, whether what you were expecting has been met or not. You can also verify in the high pass function whether at $z^{-1} = -1$ i.e. $\omega = \pi$, the result is equal to 1; it may not come exactly 1 because of truncation. In all probability it will deviate. Let us look at a band pass design.

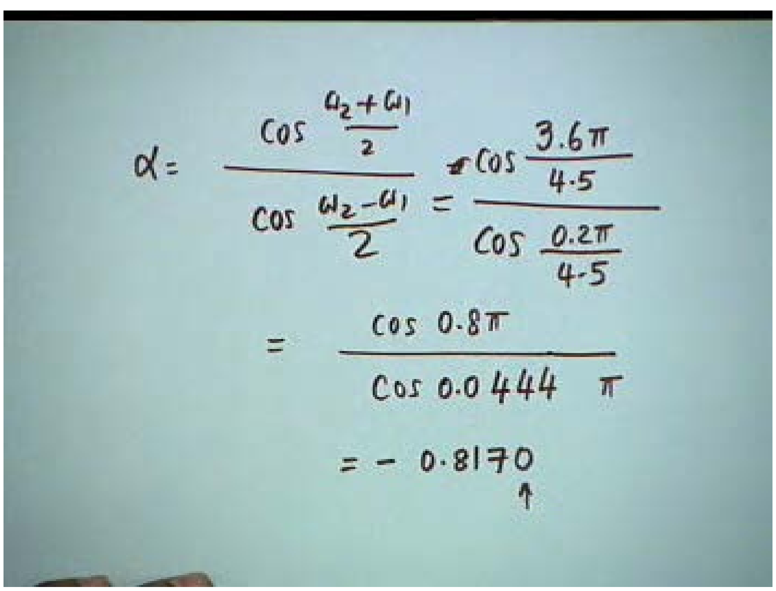
(Refer Slide Time: 18:40 -22:12)



Suppose if it is required to go from Z^{-1} to z^{-1} where the transformed filter is to be a band pass with cutoffs at 3.8 kHz and 3.4 kHz; the tolerance is the same as in the prototype. Here, we are not touching the stop band at all. Whatever stop band edges come have to be accepted. Digital to Digital Transformation may not always achieve the desired stop band. We are hoping that this will be met and then only Digital to Digital Transformation is useful; otherwise it is not because we are not touching the stop band. You have to plot, there is no other way. At the end of any design exercise in digital filter you have to make a plot and see whether the specifications are met with because of too many steps where error may creep in; the maximum error creeps in because of truncation.

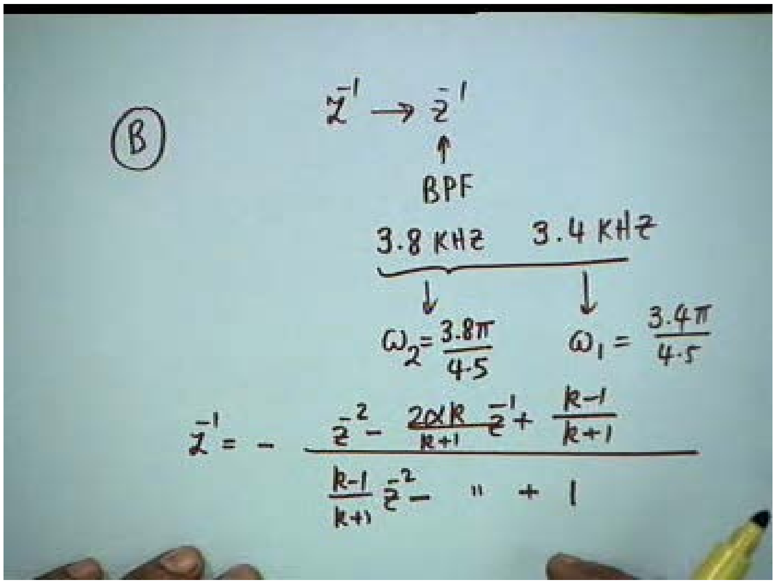
And, in Digital to Digital Transformation there is no guarantee that the new filter satisfies the stop band specifications because the only specification we are considering are the pass band edge(s) and there is no way we can take care of stop band edge(s). So this is a limitation on digital to digital transformation. On the other hand, the first route we followed takes care of the pass band as well as the stop band. In this example, the first thing we do is to find out ω_2 and ω_1. Here ω_2 would be $3.8\pi/4.5$ and ω_1 would be $3.4\pi/4.5$. The negative sign in the transformation occurs in high pass and also in band pass. In low pass to low pass, the sign is positive; low pass to band stop also, the sign is positive. Here our transformation is $Z^{-1} = -[z^{-2} - (2\alpha k/(k + 1))\, z^{-1} + ((k - 1)/(k + 1))]/([(k - 1)/(k + 1)]\, z^{-2}$ − the same term as in the numerator + 1). We have to find out α and k to calculate these coefficients.

(Refer Slide Time: 22:14 – 24:31)



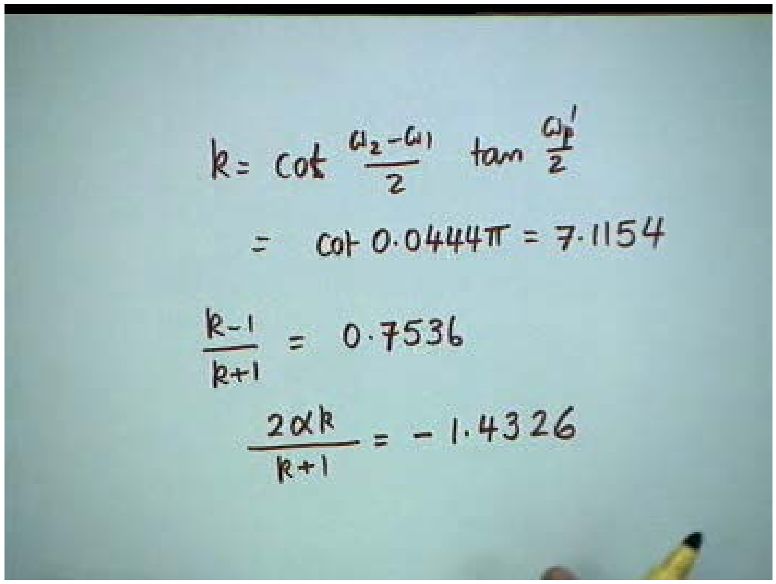
Now $\alpha = [\cos (\omega_2 + \omega_1)/2]/[\cos (\omega_2 - \omega_1)/2]$.

(Refer Slide Time: 22:28 -22:51)



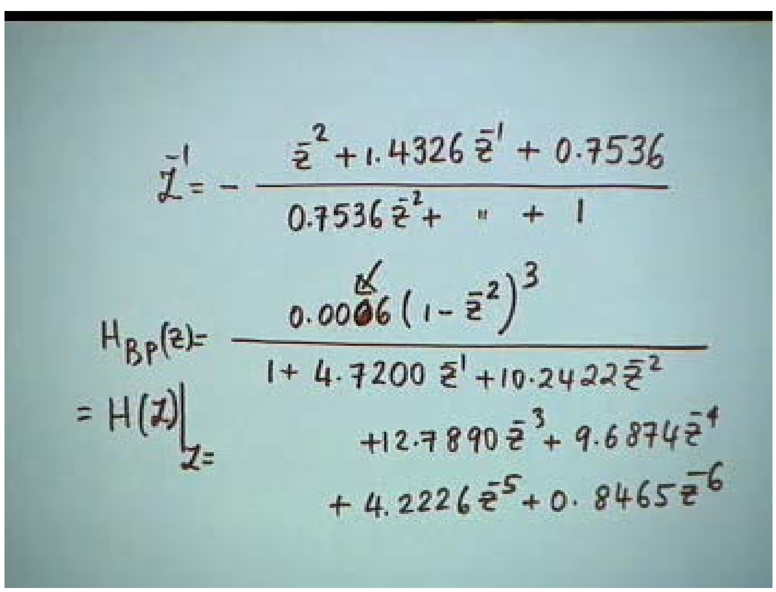
Note that we did not convert ω_1 and ω_2 into decimal numbers. We shall do so only when we cannot keep them in fractions any more due to the apprehension of quantization error creeping in and polluting the result. What we get here is $\alpha = \cos 0.8\pi/\cos 0.0444\,\pi = 0.8170$. The last 0 is significant because this 0 tells you that the next digit is less than 5. You must follow this discipline and if you truncate at 4 places you must truncate all the numbers and even if it is 0, you must show it.

(Refer Slide Time: 24:33-25:39)



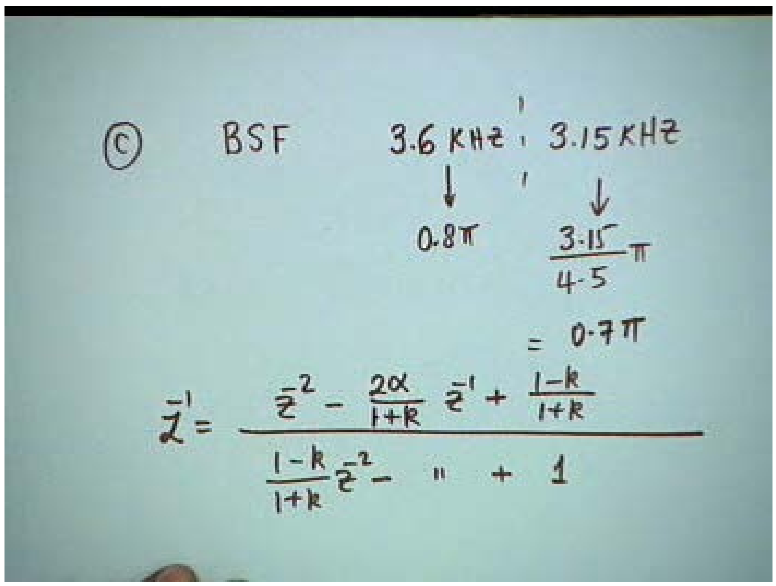
Now k = cot $[(\omega_2 - \omega_1)/2]$ tan $\omega'_p/2$ = cot $0.0444\pi = 7.1154$. The next step would be to calculate the constants; one of them is $(k - 1)/(k + 1) = 0.7536$ and the other is $2(\alpha k)/(k + 1) = -1.4326$. It is advisable to write the transformation in specific form.

(Refer Slide Time: 25:41 – 29:45)



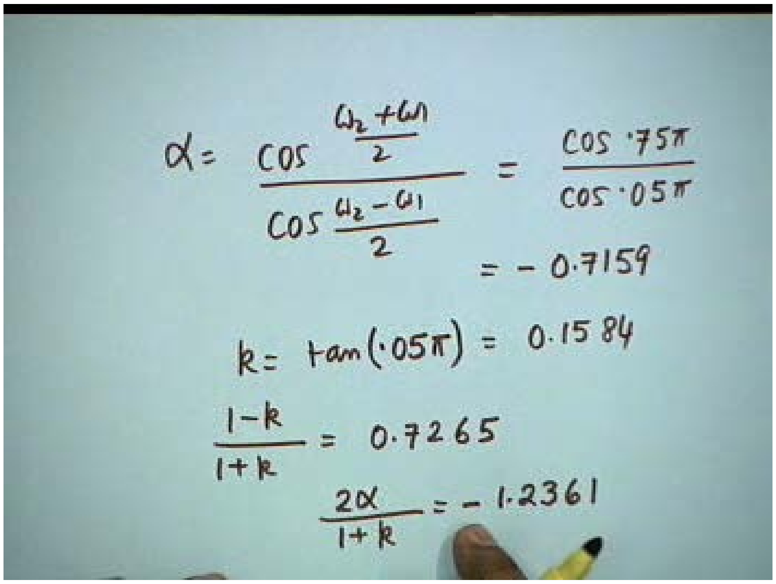
The transformation is $Z^{-1} = (z^{-2} + 1.4326z^{-1} + 0.7536)/(0.7536z^{-2} + 1.4326\ z^{-1}/1)$. If you substitute this in the expression for H(z), you get $H_{BP}(z)$ as shown in the last slide. Finally (Refer Slide Time:29:47–32:28) let us design a Band Stop Filter (BSF) with pass band edges at 3.6 kHz and 3.15 kHz; we have no control over stop band edges.

(Refer Slide Time: 29:47-32:28)



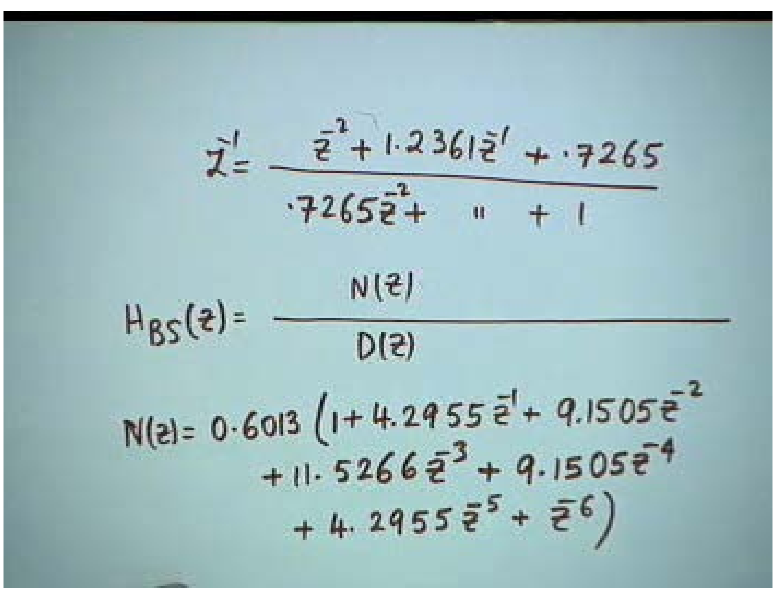
Now, 3.6 kHz corresponds to 0.8π and 3.15 kHz corresponds to 0.7π and therefore the transformation now is $Z^{-1} = [z^{-2} - (2\ \alpha/(1 + k))\ z^{-1} + ((1 - k)/(1 + k))]/[((1 - k)/(1 + k))\ z^{-2} - (2\ \alpha k/(1 + k))\ z^{-1} + 1]$. We shall have to calculate α, k and then the constants (1 – k)/(1 + k) and 2α/(1 + k). α has the same formula as in the band pass case, that is, cos α is $\cos[(\omega_2 + \omega_1)/2]/\cos[(\omega_2 - \omega_1)/2]$ = cos 0.75π/cos 0.05π and this calculates to − 0.7159.

(Refer Slide Time: 32:53 – 34:04)



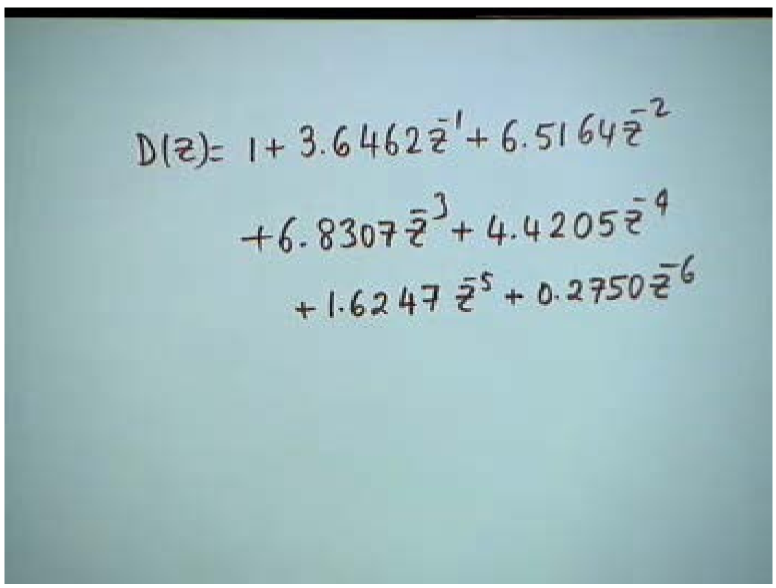
Here k is tan 0.05π tan π/4 = 0.1584. And under this condition (1 – k)/(1 + k) calculates to 0.7265 and 2α/(1 + k) becomes − 1.2361. The transformation is $z^{-1} = (z^{-2} + 1.2361\, z^{-1} + 0.7265)/(0.7265\, z^{-2} + 1.2361\, z^{-1} + 1)$.

(Refer Slide Time: 34:08 – 36:47)



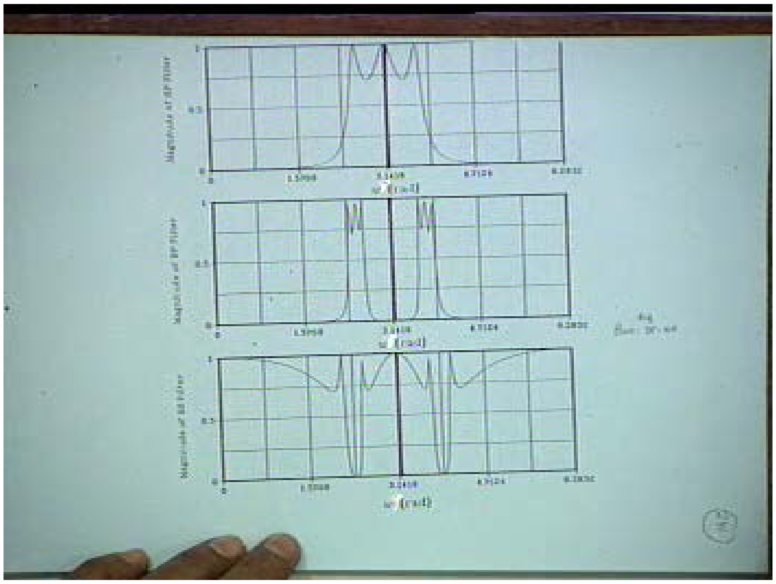
You can get the transfer function of the band stop filter by putting this in the low pass prototype transfer function. If we write $H_{BS}(z) = N(z)/D(z)$, then $N(z) = 0.6013(1 + 4.2955\ z^{-1} + 9.1505\ z^{-2} + 11.5266\ z^{-3} + 9.1505\ z^{-4} + 4.2955\ z^{-5} + z^{-6})$. Symmetry of the coefficients indicates that it is a linear phase polynomial. The numerator is linear phase but not the total filter because the phase is affected by the denominator as well. Why did it come like this? Once again it is also the reflection of the fact that the transformation is all pass. All pass property gave rise to $1 - z^{-1}$ in the high pass and $(1 - z^{-1})^2$ in the numerator of band pass and it gives a linear phase here. Linear phase is a reflection of the interchange or the reversal of the coefficients in the all pass.

(Refer Slide Time: 36:50 -39:10)



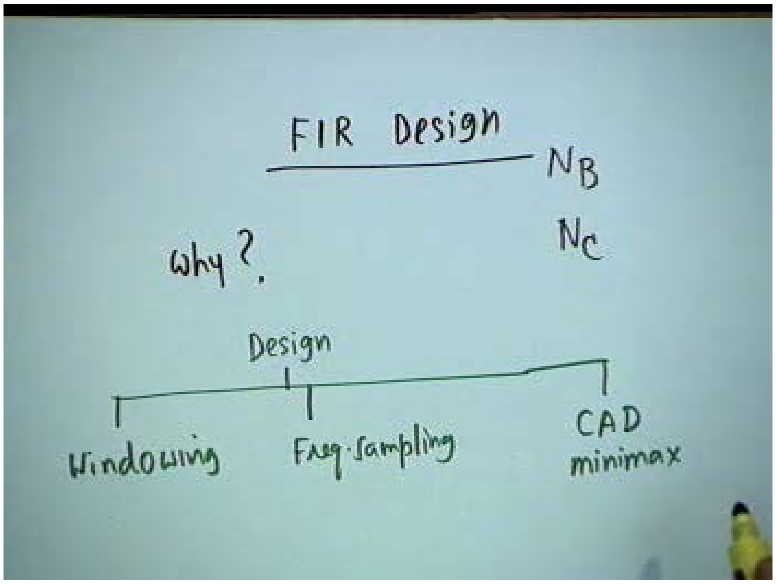
The denominator polynomial of $H_{BS}(z)$ is not linear phase; it cannot be because it is a dangerous situation. Where do the zeros of a linear phase polynomial lie? They lie in reciprocal pairs and therefore you have poles of magnitude $z > 1$ i.e. outside the unit circle and the filter would have been unstable. Here, $D(z) = 1 + 3.6462\ z^{-1} + 6.5164\ z^{-1} + 6.8307\ z^{-3} + 4.4205\ z^{-4} + 1.6247\ z^{-5} + 0.2750\ z^{-6}$. It is not guaranteed that it will satisfy the specs so you will have to plot it and see.. And this effort is worth making so that no catastrophe occurs at a later date.

(Refer Slide Time: 39:43 — 41:13)



Now comes the question of frequency responses. The three frequency responses are given in the figure, plotted not up to π but up to 2π. Our prototype was a Chebyshev filter so there are ripples in the passband. It is suggested that you find out the center frequency in both cases using $\alpha = \cos(\omega_0)$.

(Refer Slide Time: 41:17 – 47:29)



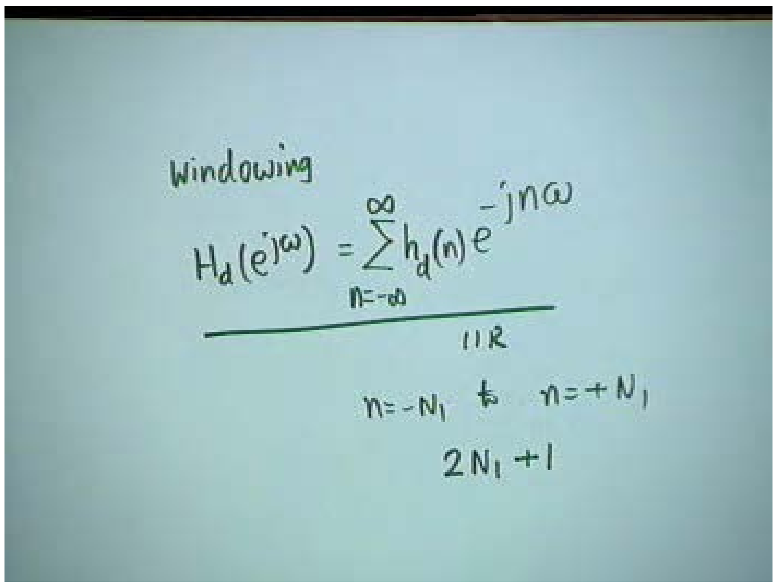
At this point we can look into FIR design. In IIR design, you do the design very confidently because you have an analytical base and everything is obtained in a closed form. FIR design, unfortunately, is not so good. In FIR there is a lot of uncertainty. And at every step you are not sure whether you are proceeding in the correct manner and whether the specs are satisfied or not; so it has a blind start.

In IIR, there were analytical formulas for calculating the Butterworth order and Chebyshev order. There is nothing like this in FIR. In FIR, you only have empirical formulas which may or may not work. If empirical formulas give the order of 17 you might have to use an order of 20 or 21. Empirical formulas always have this kind of tolerance and uncertainty. But despite these drawbacks, what is the real need to use FIR? It is because FIR is unconditionally stable and has a linear phase. Linear phase is a strict requirement in many situations.

For example, in data processing, if a rectangular pulse becomes smeared because of delay distortion, then it does not convey what you wish to convey. In speech processing, linear phase is a strict requirement so you have to use it. There are two advantages: one is, it is linear phase and the other is that it is unconditionally stable. There is a third advantage. If you have a non-causal FIR then you can make it causal by simply shifting the impulse response to the right by multiplying the transfer function by the required number of delays. So realizability of FIR is not a problem; it is not a great advantage but it is one of them. And the disadvantage is that you have to use, for the same specs, a much larger order than what is required in an IIR design. So the cost goes high. What can be done by the 2^{nd} order IIR may require a 20^{th} or 30^{th} order FIR. And if you implement it by convolution in the time domain, it becomes a very slow process. Nevertheless, this is not a disadvantage because convolution can always be calculated by DFT. If h(n) and x(n) have to be convolved, you find H(k) and X(k), multiply the two, and take the inverse DFT. DFT is also a slow process but can be speeded up by using FFT.

For FIR, the disadvantage of slowness or speed of processing can be overcome by using FFT. There are only two semi-analytical design methods, semi-analytical because none are exact like IIR. The two semi-analytical methods are: windowing and frequency sampling. Of course, you have the computer aided methods for design of optimal filters, which are also called mini max designs, by using Remez exchange algorithm in which you minimize the maximum error. Standard programs are available but are costly. Often before going to computer aided design, one uses windowing or frequency sampling to get a rough design and then you polish this design with the help of a computer aided design procedure.

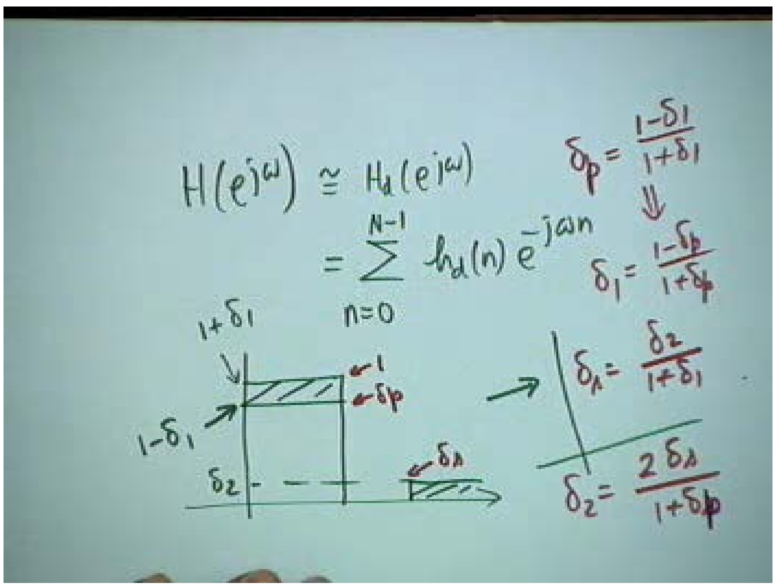
(Refer Slide Time: 47:31- 50:44)



The philosophies used in the semi-analytical methods are extremely simple. In windowing, for example, you have been given a desired transfer function and you know that the transfer function is periodic, with a period of 2π. And any periodic function can be expanded in Fourier series. So you expand this into Fourier series of the form summation $h_d(n)$, $e^{-jn\omega}$ n going from $-\infty$ to $+\infty$; n = 1 gives you the fundamental. Notice that the Fourier series expansion of the desired transfer function to be denoted by $H_d(e^{j\omega})$ is same as the Fourier Transform of the sequence $h_d(n)$.

Therefore, $h_d(n)$ is the impulse response sequence and in general $h_d(n)$ will be of infinite length. Now, arbitrarily, you truncate it and this is where the uncertainty comes. If we take $n = -N_1$ to $n = +N_1$, then the resulting sequence will define a finite impulse response of length $2N_1 + 1$. Now this filter is obviously not realizable because impulse response is not 0 for $n < 0$. But the solution is very simple; you simply multiply the resulting transfer function by z^{-N_1}. There is no guarantee that you have satisfied the specs. This N_1 is purely arbitrary to start with. There are empirical formulas to get you an idea of the order you require. Instead of truncating from $-N_1$ to N_1, we will truncate from 0 to some N − 1. We shall call this $H(e^{j\omega})$.

(Refer Slide Time: 50:48- 55:40)

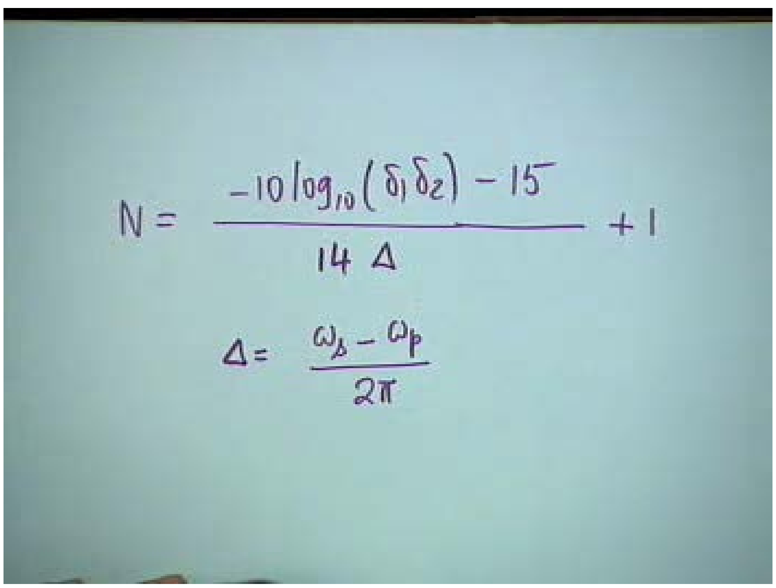


So we shall take $H(e^{j\omega})$, which will be an approximation to $H_d(e^{j\omega})$, as $\sum_{n=0}^{N-1} h_d(n)e^{-j\omega n}$ so that the length is N and we have thrown out all other terms in $H_d(e^{j\omega})$. I get a realizable filter H(z) of finite impulse response. We do not know before hand as to what this N should be, so it is a guess work. Fortunately for any type of filter, researchers investigated a very large number of filters on the computer. They made the designs and they came up with some simple empirical formulas for estimating the order that is needed, given the tolerances δp and δs.

Given the tolerances in the passband and stopband, there are some simple formulas for estimating the order. There are also complicated formulas which need not be applied unless the tolerances are very stiff. But these formulas were worked out not with the kind of tolerance scheme we are following. What we follow is: highest magnitude is 1 in passband and the lowest is δp, and in stopband highest magnitude is δs. These empirical formulas for the typical low pass filter use maximum as $1 + \delta_1$ and the minimum as $1-\delta_1$ and the maximum in the stopband as δ_2. The empirical formulas are in terms of δ_1 and δ_2. But converting from here to our specs is not a problem. All you have to do is to normalize by dividing by $1 + \delta_1$ which gives you δ_1 terms of δp as $(1 - \delta p)/(1 + \delta p)$. And $\delta s = \delta_2/(1 + \delta_1)$. And since δ_1 has been found out you can put this here and the result is $\delta_2 = 2\delta s/(1 + \delta p)$.

The formula for estimation, the only available one, required many hours of work on the computer by Bell Labs workers. When δs and δp are given, you have to find out the corresponding δ_1 and δ_2 and then substitute in the formula, which is:

(Refer Slide Time: 55:43 -57:13)



$N = \{[-10 \log_{10} (\delta_1 \delta_{2.}) - 15]/(14 \Delta)\} + 1$; + 1 is also not very important if the first term is much greater than unity. The quantity Δ is (stopband edge − passband edge)/(2π). This is a gross estimate. Once you get this estimate, you can go ahead designing the filter, but you must find out the actual response and see whether it is satisfied or not. We will continue this next time.