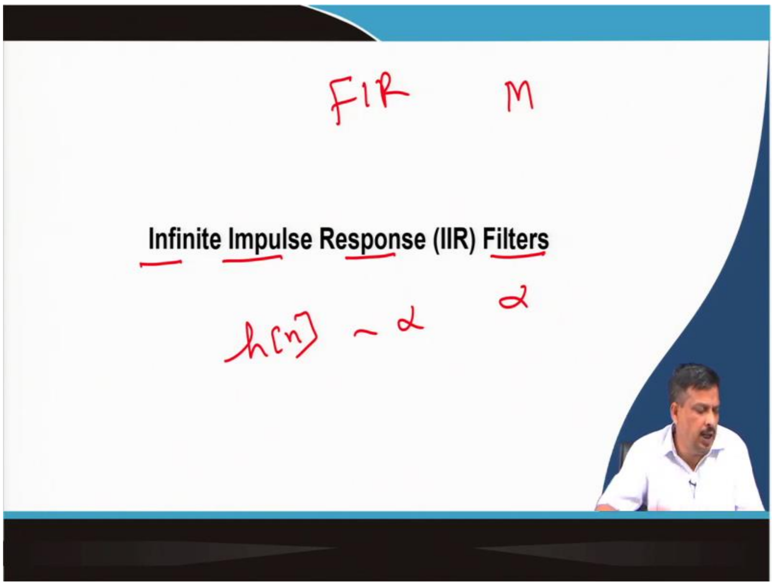
Signal Processing Techniques and its Applications
Dr. Shyamal Kumar Das Mandal
Advanced Technology Development Centre
Indian Institute of Technology, Kharagpur

Lecture - 41
Infinite Impulse Response (IIR) Filters

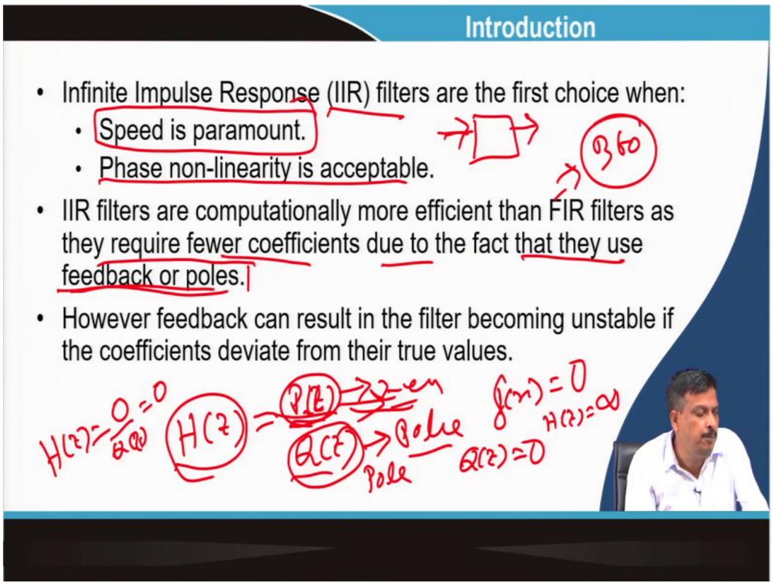
So, we talk about the Infinite Impulse Response Filter.

(Refer Slide Time: 00:28)



So, we have talked about the FIR filter and finite impulse response filter. So, FIR, where M is the finite order of the filter, is finite. Here, I am talking about the infinite impulse response filter. So, where I have an h[n] whose length is infinite, it is from minus infinity to infinity. So, an infinite length filter is an impulse response length filter. So, that is called an IIR filter, ok.

(Refer Slide Time: 00:58)



Now, why IIR filter? You may say that, sir, I know the FIR filter and the design is very simple, so why should I use the IIR filter? Now you know that the IIR filter is computationally more efficient than the FIR filter, which I have already discussed. Because the IIR filter required fewer coefficients due to the use of feedback or poles, I will describe it in detail.

Why is this feedback called pole? Because when I say IIR filter, when I say H(z), H(z) is nothing but less the P(z) by Q(z). H(z) is the z domain representation of the filter transfer function, and you know the solution of P(z) is called 0, and the solution of Q(z) is called a pole. Why?

Why is the solution of Q(z) called pole and the solution of P(z) called 0? So, when I say $x^2 + 2x + 1$. I said it is a second-order equation. It is a two solution. What is the physical meaning of the solution? The physical meaning is that if I put that value in that equation, the equation value becomes 0.

So, when I found out the solution of an equation, we said f of x solution. When I found out the solution, we called f of x equal to 0. Now, if I see that H(z), why is the P(z) solution of P(z) called 0? Because H(z), if I put that value in P(z), then P(z) becomes 0. So, 0 by Q(z) is become 0.

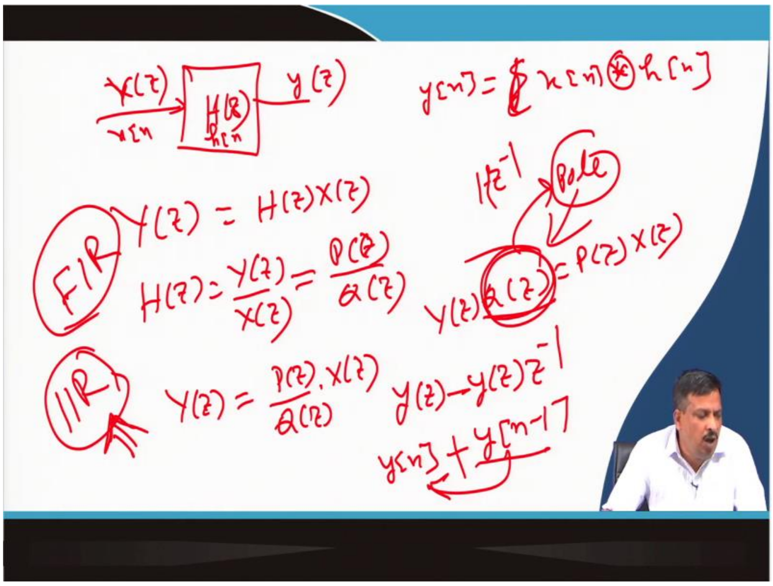
That is why the solution of P(z) is called 0. Now when the solution of Q(z) is equal to 0, that means that H(z) is infinite 1 by 0. So, those are called poles, and what is the physical significance of the pole and 0? Pole defines the resonant frequency of the system. The resonance will happen when the frequency response for which ω the response is resonated. That is why poles are related to resonance, and 0s is the at 0s value for that ω, for which Q(z) is 0.

The impulse of the response of the system is 0. So, the mod of H(z) is 0. So, those are called pole and 0. So, the IIR filter is not very efficient compared to the FIR filter. If the fir filter order is 360, you know I required convolution on 360 n equal to 360. So, n cross n 360 cross 360. But if I say second-order differential equation, you know how to realize, using structure 1 and structure 2, how many delays are required. So, on the implementation side, the computationally IIR filter is easier than the FIR filter. That is why the IIR filter is important, where speed is a criterion.

Suppose you want to design a system that is real-time. So, if speed is my requirement, then the IIR filter is the best. What is the problem? The problem is that the IIR filter may have a phase distortion. So, it cannot be an exact linear phase if it is not a linear phase. So, I want to design a system. If the requirement is a completely linear phase, then I cannot use the IIR filter.

So, in that case, I have to use the FIR filter. However, if some non-linearity is acceptable in the phase, then I can use the IIR filter. Now you said Q(z) is the position of the pole, which is nothing but a positive, which is nothing but feedback. Because why is it called feedback? I will show you. Why are Q(z) poles called feedback? So, you know, let us say, suppose I have a system.

(Refer Slide Time: 06:06)



I have a system called H(z). If I apply an input X(z), I get output Y(z). Now I know Y(z) is equal to H(z) multiplied by X(z) or if it is h[n] if it is x[n], then I know y[n] is equal to the convolution of x[n]; x[n] convolved with h[n] that I know. Now, what is the H(z)? I can say H(z) is nothing but a Y(z) divided by X(z).

So, H(z) in the form of P(z) by this can be in the form of P(z) by Q(z). So, if I say H(z) is in the form of P(z) by Q(z), then I can say Y(z) is equal to P(z) by Q(z) into X(z). Now, I can say Y(z) into Q(z) is equal to P(z) into X(z). Now, this Q(z) is what it is; let us say Q(z) is 1 minus z^{-1}. So, it is nothing but a Y(z) minus Y(z) into z^{-1}. So, that means it is in the time domain.

So, y[n] is equal to minus y[n-1]. So, there is a negative feedback of the previous output. So, if there is negative feedback on the previous output, it may also be a plus. There is a positive feedback of the previous output.

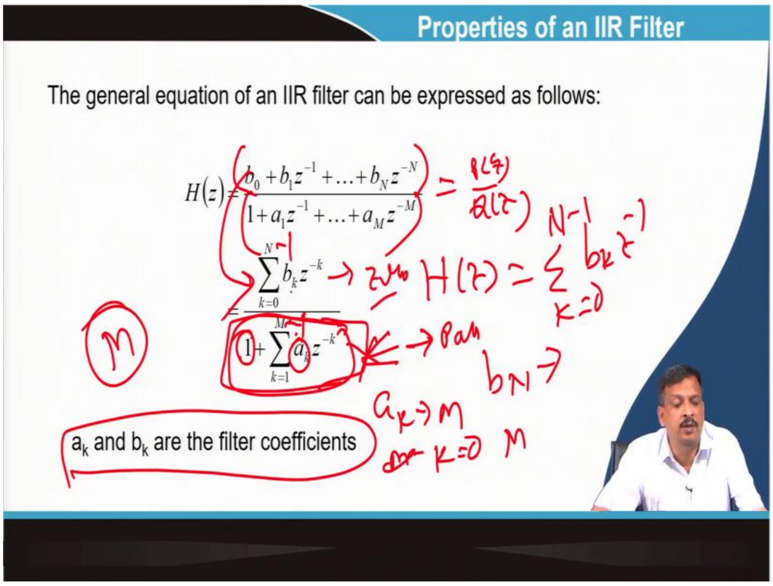
So, that is why who created this feedback Q(z)? What is the Q(z)? Q(z) is nothing but a pole. So, that is a pole related to the system's feedback. Is it clear? So, since there is feedback in the system, do you know how to design an oscillator? If there is positive feedback, then the system can oscillate and become unstable.

So, stability is one of the important factors that depends on this Q(z) that pole. So, if the poles are not properly placed, the design filter, which is nothing but a system, maybe an unstable system. But any system that I want should be stable.

So, that is the problem. So, the first drawback of the IIR filter is the linearity of the phase. The second drawback is the stability of the filter. So, these are the two drawbacks of designing an IIR filter compared to an FIR filter. But if I say computational complexity, the IIR filter is much more superior to the FIR filter.

So, when I require a real-time application where speed is paramount or speed is a requirement, then I go for the IIR filter. But I have to carefully design that IIR filter so that it is stable and the phase distortion is not that much. So, that is the difference between the FIR filter and the IIR filter.

(Refer Slide Time: 09:56)

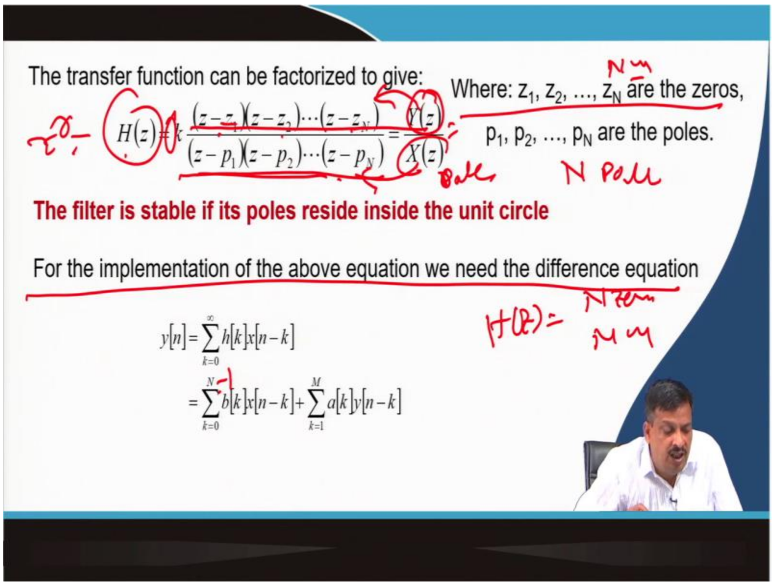


So, how do I design how what should be the expression of an IIR filter? What is a general solution for an IIR filter? So, I know IIR filter H(z) is nothing but a P(z) by Q(z). So, it is nothing but a 2 polynomial, one up and one down. So, this polynomial can be written this way: also, k is equal to 0 to n bk z^{-k}. Ok or not? If it is an FIR filter, this feedback is absent.

So, I know H(z) is equal to k equal to 0 to N bk z^{-k} in case of an N minus 1 I can say or N minus 1 I can say or M minus 1. If it is k equal to 0 to N minus 1, then it is N number of bk order is bk is N. If it is k equal to 1 to M, then the order of ak is equal to M ok. So, I can say k is equal to that if it is said that if I consider this one is at M equal to 0 k equal to 0, this whole thing is 1, then I can say the order of M is equal to M, then I say it is M minus 1. So, M minus 1 plus 1 means the total order is M. And this bk and ak are called the filter coefficient.

So, FIR filter this portion is not there, but IIR filter this portion is there those are related to the pole, and those are related to the 0.

(Refer Slide Time: 11:38)

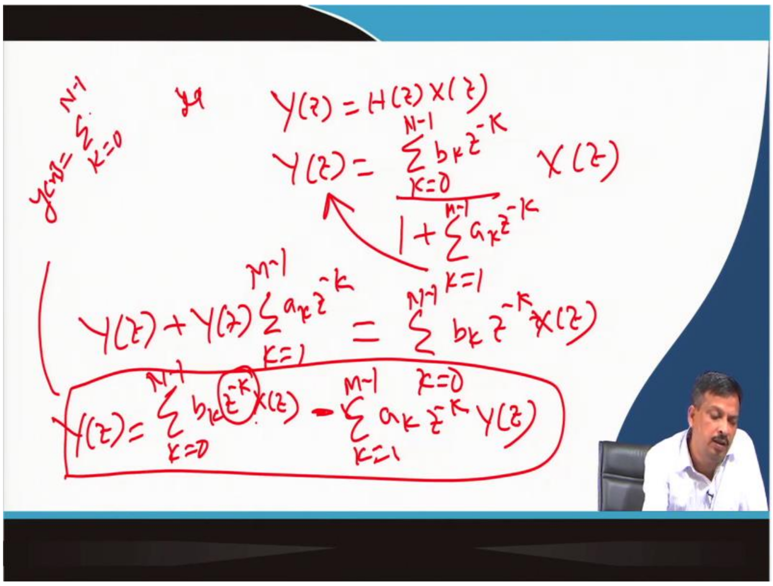


So, let us say there is a 2 polynomial. So, every polynomial can be written as a factor. This k is a constant, and I can write down the factors in the upper side Y(z) in this factor and X(z) in this factor. So, these z 1 z 2 z n are called 0 position. Because I know H(z) is X(z) by Y(z) or P(z) by Q(z), whatever the solution of Y(z) is nothing but a H(z) equal to 0. So, those are called 0 position solutions of X(z), which is nothing but a 0, but at that time, H(z) is infinite, where the system is resonating.

That is why this is called pole position. So, poles are resonance 0s, and I can say the system output response is 0. Nodes and anti-nodes bode plot all of you have heard about that part. So, pole 0 in control theory, you have heard about it. So, I can say there is if there is an N number of P N is there. So, I can say N number of poles are there if N number of zs are there. So, I can say N number of 0s are there. So, I can say H(z) is a transfer function that consists of N number of 0 and N number of poles. This may be deeper than the N number of poles and the N number of 0.

So, if I write it down in the derivative form, it is in differential equation form. So, let us take a slide here.

(Refer Slide Time: 13:14)



So, differential equation form. So, what is there? I know Y(z) is equal to H(z) into X(z) ok. So, how do I write down X z? Y(z) is

$$Y(z) = \sum_{k=0}^{N-1} b_k z^{-k}$$

I write or let us say N k then it will be N plus 1 order.

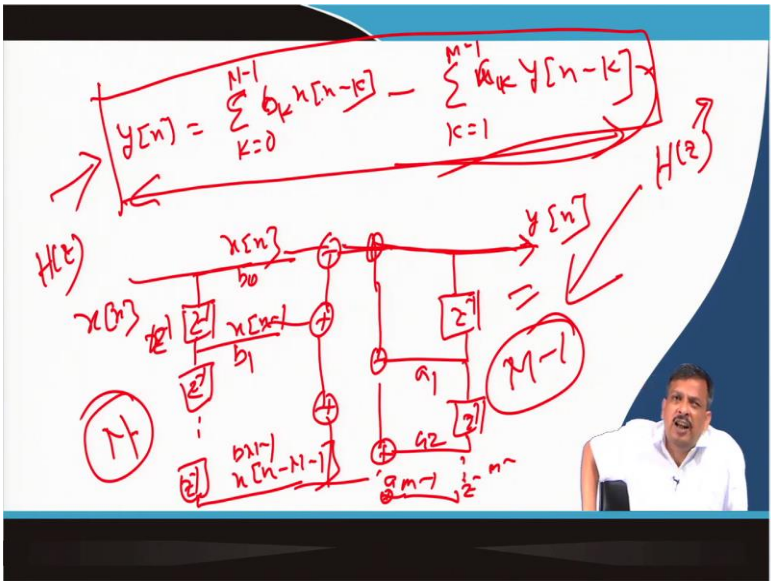
So, it's better to write N minus 1, ok? Divided by 1 plus k equal to 1 to M minus 1 ak z^{-k}, let us say ak z^{-k} into X(z). So, if I say this, it will multiply here. So, I can say Y(z) plus Y(z) into k equal to 1 to M minus 1 ak z^{-k} is equal to k equal to 0 to N minus 1 bk z^{-k} X(z) ok.

Now, what is Y(z)? I can say Y(z) is equal to this and will be on this side. So, I can say

$$Y(z) = \sum_{k=0}^{N-1} b_k z^{-k} X(z) - \sum_{k=1}^{M-1} a_k z^{-k} Y(z)$$

if you suit this equation, when I say the time domain equation. So, Y(z) gives me y[n]. Here, I can say that the input signal is delayed by the k number of samples. So, I can say it is nothing but ak equal to 0 to N minus 1. So, let us say instead of writing here, I will write on another slide. So, from this equation I can say.

(Refer Slide Time: 15:46)



So, Y(z) is related to the y[n] time domain. Then this will be k equal to 0 to N minus 1 ak z^{-k}. So, I can say this is nothing but ak equal to 0 to N minus 1 bk into x[n] minus k x of X(z) minus k. So, z^{-k} x[n] minus k minus k equals 1 to M minus 1 ak y[n] minus k. So, this one is the feedback y[n] minus k. So, this is the form of differential equation form of the differential equation of H(z).

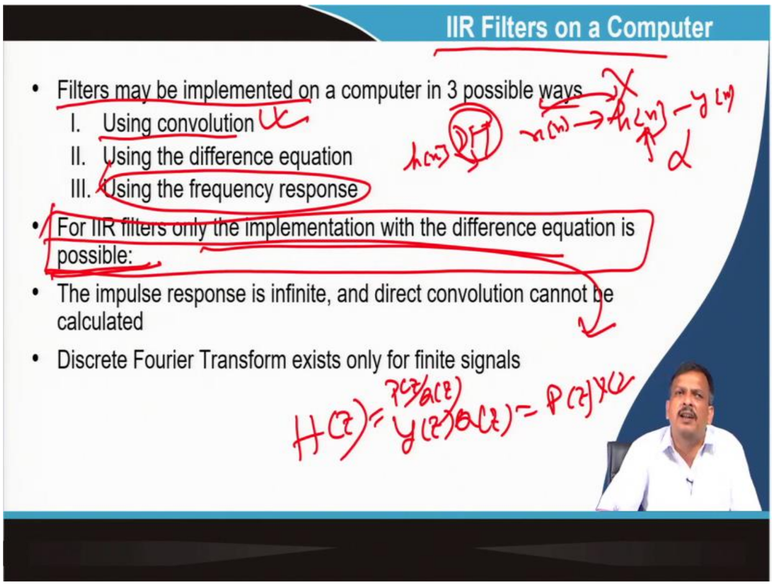
Now, if I told you, can I realize this equation? How do I implement this equation? I want to implement this equation told me. How do I implement it? Very simple, this is nothing but a x[n]. This is nothing but a y[n]. So, discrete structure 1 or structure 2. So, I know that on this side, there will be a z^{-1}, z^{-1}, and dot dot dot z^{-1}. So, the first one is k equal to 0; that means z to the power that the x that k equal to 0 means x[n]. So, this is my x[n] this is my x[n-1]. So, I required x[n] minus N minus 1 sample.

So, N number of delays is required, all will be added up. So, this is multiplying by this is b 0, this is b 1, this is b N minus 1, and all will be added together then fit to here. Then y[n] has to be delayed by z^{-1} I. This will be multiplied by a 1, then again z^{-2}, multiplied by a 2 dot $z^{-(M-1)}$, multiplied by an m minus 1.

All will be added up to get the y[n]. So, this is a structured 2 realization. I will interchange this one with this one so I can get the structure. So, this is structure 1 realization. I can interchange this one, and I can get the Structure 2 realization. So, instead of N number of delay here and N my M minus 1 number of delay here, I may require only N number of delay. Which one is the highest? Structure 2.

So, once I am able to form the differential equation of the IIR filter transfer function, which is H(z), which is the filter transfer function. I can able to implement the filter using structure 1 or structure 2 realization. So, ultimately, what do I have to determine? I have to determine the H(z). Once I get H(z), I can able to implement it using either structure 1 or structure 2 realizations. How do I determine H(z)? So, where do I determine H(z)?

(Refer Slide Time: 19:32)

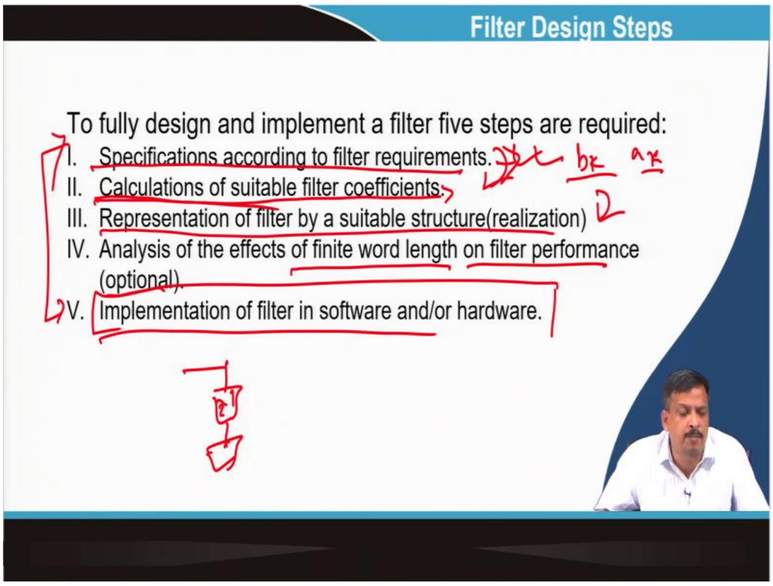


So, what will be given to me? The only thing given to me is the filter specification or desired frequency response of the filter. So, from there, I have to derive H(z). And once I get H(z), I can implement it using the computer. So, computer IIR filters on a computer. So, using convolution. So, filter what are the common methods for filter implementation? If I know h[n], I know if I have an x[n], then I can take the convolution, and I get the filter output convolution. But if h[n] is infinite, the convolution is not possible.

So, the IIR filter cannot be implemented using the convolution method. Now, if I am able to express the filter in differential equation form, then I can easily implement it using structure 1 and structure 2 realization. So, I can only implement the IIR filter using these methods—frequency methods, which I cannot do. Because once I said that transfer I since h[n] is infinite duration, I cannot take discrete Fourier transform because discrete Fourier transform requires a finite length. So, I cannot do that. So, what are the methods remaining? Only differential equation form.

What is the requirement for this form? I have to find out H(z). Once I get H(z), I know Y(z) into Q(z) is equal to P(z) into X(z), where H(z) is nothing but a P(z) by Q(z). Once I do that, I can write down the realization of Structure 1 and Structure 2 and implement the IIR filter.

(Refer Slide Time: 21:42)



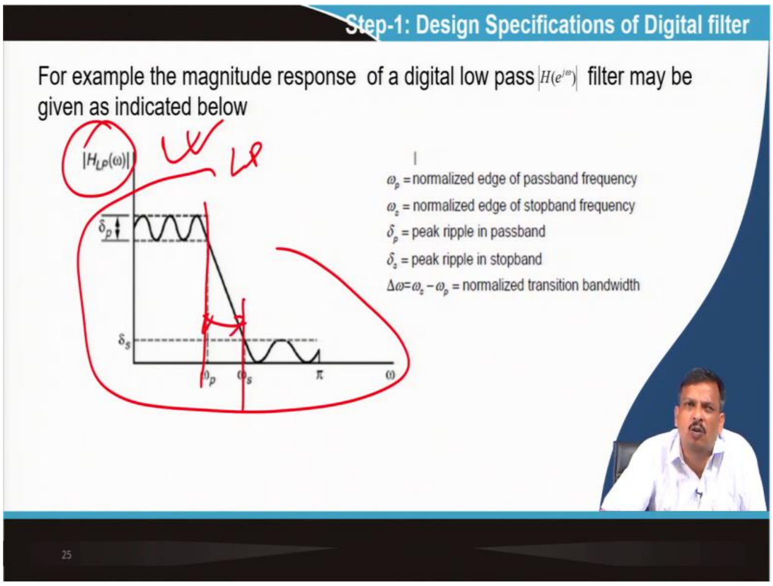
So, if you summarize that filter design step for the IIR filter, that is a five-step—specification according to the filter requirement. So, I have to require a filter specification, let us say design and IIR filter low pass IIR filter cut off frequency 2-kilo hertz transition bandwidth this refill this.

So, that specification is according to the filter requirement—the calculation of the suitable filter coefficient. So, bk and ak, I have to define, and I have to find out. I have to find out bk and a k; once I get that bk and ak value, I can form the differential equation form, and I can implement it using structure 1 or structure 2 realization.

So, first, find out the specification from the given description, write down the filtered frequency response and design the specification, then calculate ak bk from that frequency response, and then implement using realization discrete structure 1 or 2 realizations. Then, analyze the effect of finite word length or filter parameter. It may be a truncation error. Implementation of filters in software or hardware is nothing but a delay.

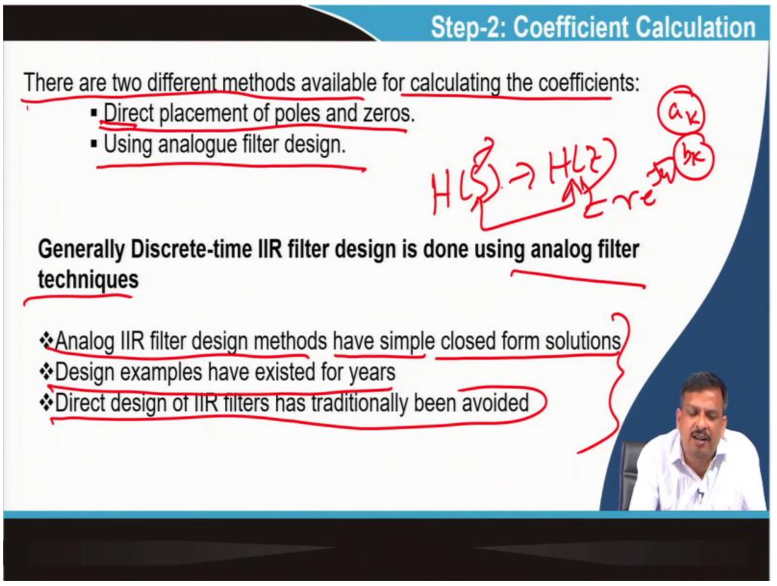
So, if I say that I want to delay the implementation of this gap. What is nothing but a z^{-1}, which means one sample delay? So I can easily implement it using the computer. So, those are the 5 steps to follow for the design and IIR filter.

(Refer Slide Time: 23:11)



So, the first step is the design specification, the same as the FIR filter. So, this is nothing but the magnitude response of the low pass filter, which is a pass frequency, stop bandage frequency, transition bandwidth, ripple in the pass band, and ripple in the stop band.

(Refer Slide Time: 23:34)



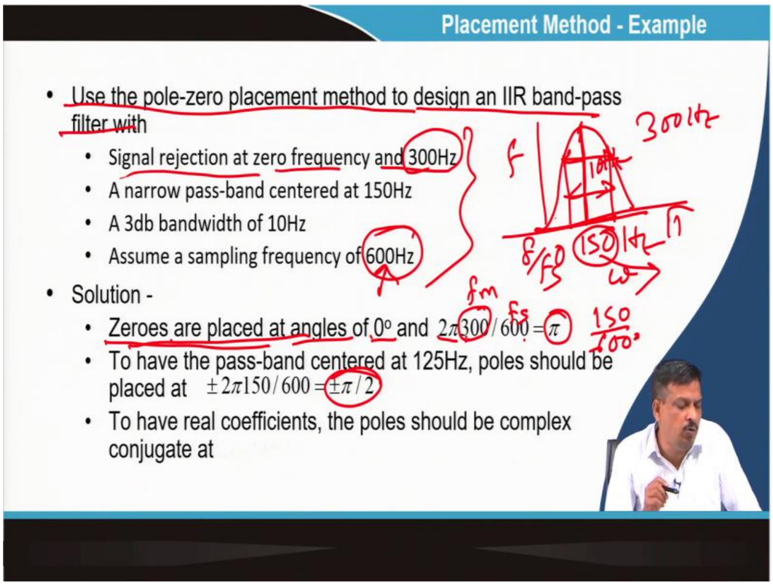
So, all are the same requirements here. The next step is a calculation of the coefficient. How do I derive ak and b k? That is my target. There are two different methods available for calculating the coefficient. One is called the direct placement of pole and zero; one is called using the analogue filter method. So, there are two methods available to compute ak and bk. One is called direct placement; another is called using analogue filtered design methods.

The most used methods are analogue filter design techniques because of these 4 3 reasons. IIR filter design method at simple closed-form solution analogue IIR filter. So, this was before the digital filter,. The analogue filters were there. So, this is an established theory, I can say. So, design examples have existed for a year. People have studied different methods, and they have established that they are good for designing analogue filters. Direct design of IIR filters has traditionally been avoided.

So, direct design would never do it. So, direct implementation of the pole and zero the use of these methods is very rare. In most cases, we go for the analogue filter design technique; that means, initially, we design that analogue filter and then convert it to the H(z). So, what is the analogue filter? An analog filter is nothing but a Laplace domain. You designed that analogue filter for Laplace domains Hs. Then, once you get the Hs, you can convert it to H(z).

What are the methods you think about? Z transform, Laplace transform, and all of you have done. So, how is the H of S converted into H(z)? You know the physical significance of z is nothing but a r $e^{j\omega n}$ $rne^{j\omega n}$, and S is also a complex number. So, there is a relationship between the S domain and the z domain, which is called bilinear transformation. So, that will be used ok.

(Refer Slide Time: 26:08)



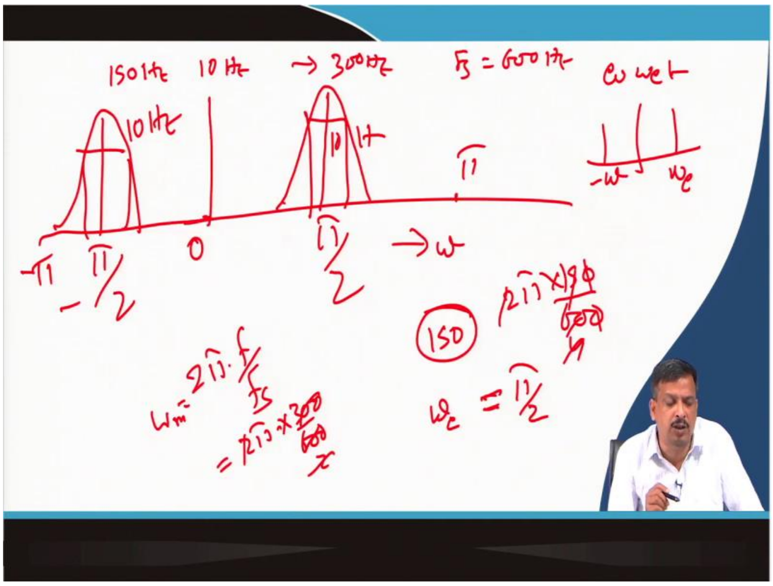
Let us say direct methods. This will be an example of the direct methods. So, what is the direct method? Use the pole-zero placement method to design an IIR band pass filter. Let us say this: I have to design. So, what I said is that I have to design and bandpass filter narrow pass band centre at. So, 150 hertz is the centered frequency, and the 3dB bandwidth is 3dB bandwidth is 10 hertz.

So, this distance is 10 hertz. So, in the narrow pass band pass filter, I want to design a centred frequency of 150 hertz and a 3dB passband of 10 hertz. And what is there? The single rejects signal rejection at 0 frequency and 300 hertz. That means the maximum frequency component of the signal is 300 hertz.

So, what are the sampling frequency requirements? 600 hertz more than 600 hertz. So, let us say I have a sampling frequency of 600 hertz. So, 0 is placed at the angle of 0 and 2π by. So, this ω is in radian, but it is given in frequency. You know f by F s is the normalized discrete frequency. So, this is f of m. This is F of s. So, 300 divided by 600 is the π. So, the maximum oscillation is 0 to π. Within that oscillation, I had to design a filter with a centred frequency of 150 hertz.

So, what is the value of ω? So, 150 hertz; that means 150 divided by 600 into 2π. So, that is nothing but a plus-minus π/2. Do you understand or not? So, what I said. So, what I want to do is that I have wanted to design and bandpass filter.

(Refer Slide Time: 28:33)



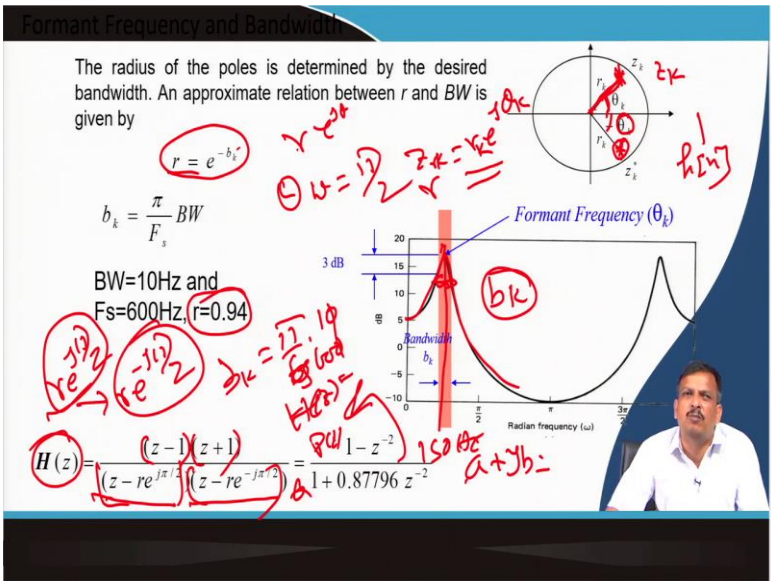
Whose centred frequency is 150 hertz, the bandwidth is 10 hertz, and the signal maximum frequency is 300 hertz. The sampling frequency F s is equal to 600 hertz. So, if I want to draw the magnitude spectra. So, let us see this is my 0, and this axis is in ω. So, I know what the highest rate of oscillation is in the case of a digital signal. It is nothing but a 2π into f by F s ω s ω m.

So, what is ω m is nothing but a 2π into 300 divided by 600, which is nothing but a 2 so π. Now, once I draw the low pass filter, it will be a folding in nature. So, if you know that if I say cos ω c t, it has a frequency response ω c plus and ω c minus. So, if I say I want to design a filter at a centred frequency of 150 hertz.

So, what is 150 hertz? 150 hertz is nothing but a 2π into 150 divided by 600. So, it is nothing but a π/2. So, ω is equal to ω c centred frequency is nothing but a π/2. So, it has a 2-centered frequency. One is this side also will be minus π. So, here will be π/2 minus, here will be π/2 plus. So, the actual frequency response will be 3dB down, which is 10 hertz, and this will also be here, where there will be 3 dB down 10 hertz.

So, that is my frequency response to that given filter specification. Now, I have to design the filter. So, where do I put that? How do I design the filter? So, I can say 0 is placed at angle 0 and π. 0s are plus in angle 0s and π, but if it is stable, then it should be the pole within a unit circle.

(Refer Slide Time: 31:02)



So, I have to find out the pole position. So, how do I find out the pole position? So, if you see the radius. So, if you see this is my pole position, let us say this is my pole position, which is z k. So, z k can be expressed in the 2-part. One is called r. So, this circle is a unit circle. So, my poles are within the unit circle. So, I can say that pole z k can be represented at r $e^{j\theta}$. So, r $e^{j\theta}$ let us say k and r k is this one. So, I know if my h[n] is real if my filter coefficient is real.

Then, I learned that these poles are ordered in complex conjugate. So, that means, if there is a positive θ k, there will be a negative θ k, and there will be another pole in here, which is the complex conjugate pole of this one. A pole can be written as r $e^{j\theta}$, r is the distance from the center, and θ is the angle.

So, instead of z is a+jb, a pole is a complex pole in the form of a+jb. So, a+jb can be written as magnitude, which is r polar form and θ, which is the angle ok. Now, since I know θ. What is the θ? Θ I know the centred frequency at 150 hertz ok. So, the centre frequency is 150 hertz, which is related to ω; ω is equal to π/2, which I have already done. Now, what is the value of r? Now, if you remember, I think it was in the first week.

I have said that bandwidth is a resonance. So, poles are related to the resonance. So, the resonance-centered frequency is 150 hertz, let us say, and the bandwidth is 10 hertz. So, bk represents the bandwidth. What is the relationship? r equal to e^{-bk} b k is in radian. But here, bandwidth is given in hertz. So, how do you convert? I know bk is nothing but a π by F s into a given bandwidth. So, I can say 10 divided by 600. So, I can say π by 60. So, bk is equal to π by 60, and r is equal to e^{-bk}.

So, e to the power minus π by 60. That will give me the value of r equal to 0.49 0.94. So, I know that the transfer function is real. So, the pole will occur in complex conjugate form. So, this is one pole, and this will be another pole. So, this one r $e^{j\pi/2}$. What is the complex conjugate? r $e^{-j\pi/2}$. So, I can say 2 poles are there one is e to the power r $e^{j\pi/2}$ or $e^{-j\pi/2}$.

So, I write down 2 poles, and two 0s are there. I write down z^{-1} and z to the power plus 1. So, this is my transfer function H(z). So, placement of the form of a given formant frequency and centred frequency, bandwidth, for that resonance bandwidth means the band bandwidth of the filter.

For the centred frequency, I can place the poles and 0s, and I can design the filter. So this is called the placement of pole 0 method. Is it clear? So, once I get H(z), I can easily implement H(z) using direct structure 1 or direct structure 2 method. So, it is nothing but a P(z) by Q(z). So, I can multiply Y(z), I can get that differential equation form, and I can implement using structure 1 or structure 2 methods.

Thank you.