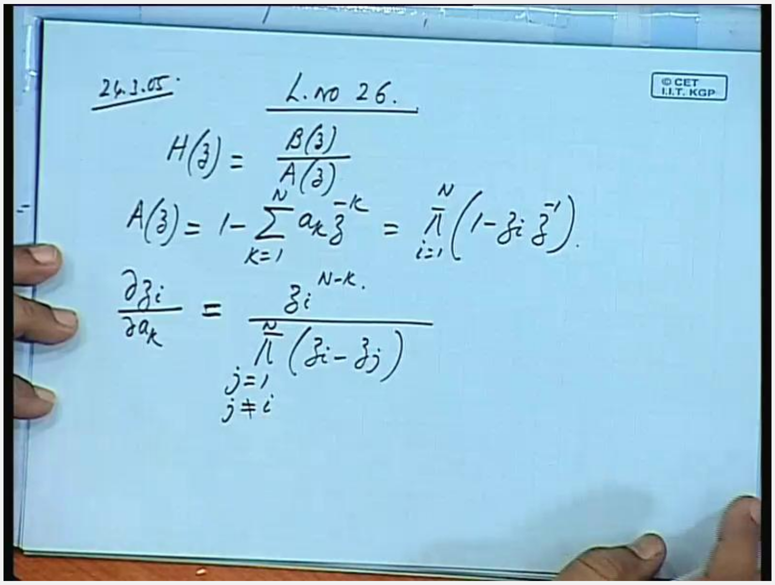
Digital Signal Processing
Prof. T. K. Basu
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 26
Effects of Quantization

(Refer Slide Time: 00:50)



24.3.05.

L. NO 26.

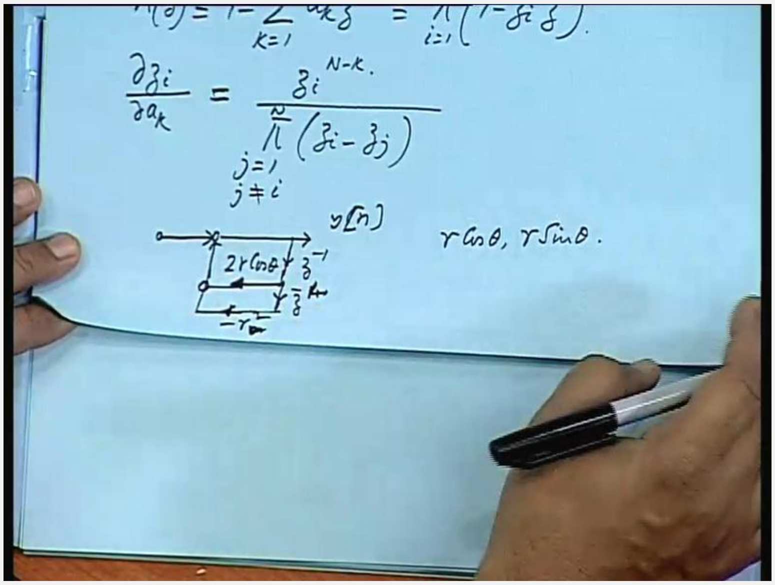
$$H(z) = \frac{B(z)}{A(z)}$$

$$A(z) = 1 - \sum_{k=1}^{N} a_k z^{-k} = \prod_{i=1}^{N} \left(1 - z_i z^{-1}\right).$$

$$\frac{\partial z_i}{\partial a_k} = \frac{z_i^{N-k}.}{\prod_{\substack{j=1 \\ j \neq i}}^{N} (z_i - z_j)}$$

Good morning friends. We shall be continuing with our discussion on Quantization. Last time we saw, if the transfer function H z is written as the ratio of two polynomials and then say for any one of them, you can write the polynomial in this form. We can also write in the product form, where z i will be the roots of this, the factors i varying from 1 to N. We saw the sensitivity delta z i by delta a k with the parameter a k is z i to the power N minus K divided by this product z i minus z j, where j equal to 1 to N, j not equal to i.

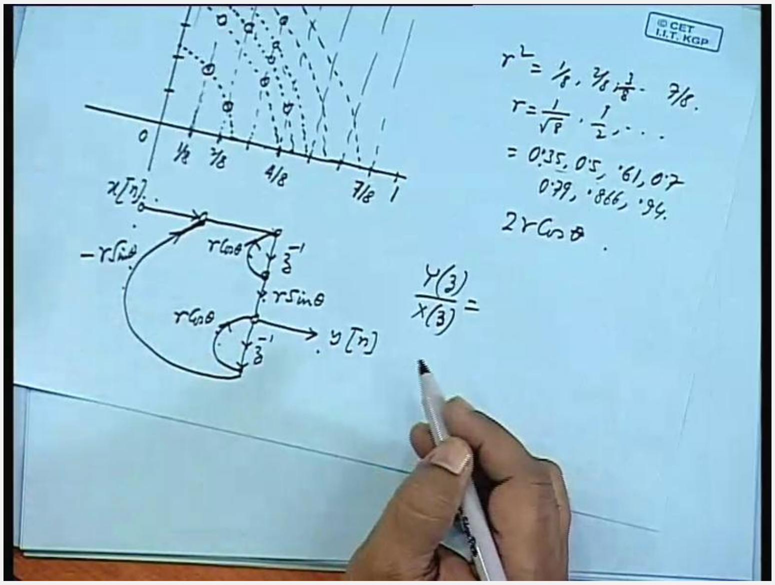
(Refer Slide Time: 02:24)



$\frac{\partial z_i}{\partial a_k} = \frac{z_i^{N-k}}{\prod_{j=1, j\neq i}^{N} (z_i - z_j)}$
$y[n]$
$2r\cos\theta$
z^{-1}
$-r^2$
$r\cos\theta, r\sin\theta.$

Let us see the representation of a Biquards, you can have say the output is y n z to the power minus 1, suppose the roots are complex minus r square z to the power minus 1. For this, what are the possible values of the roots, you know the roots are r cos theta and r sin theta for this.

(Refer Slide Time: 03:38)



Now, this is a structure of just one factor, one quadratic factor, so what should be the location of the roots. If, we have say three bit representation of these parameter values 0, 1, 2, 3, 4, 5, 6, 7, 8 so this is 1, this is 7 by 8, this is 4 by 8 and this is 1 by 8, 2 by 8 and so on. Similarly, 1, 2, 3, 4, 5, 6, 7, 8; so what would be the values of r square, r square is this constant and this is 2 r cos theta and these are the things which are quantized. So, r square is quantized to any of these levels, so r square will be 1 by 8, 2 by 8 and so on up to 7 by 8, including 0.

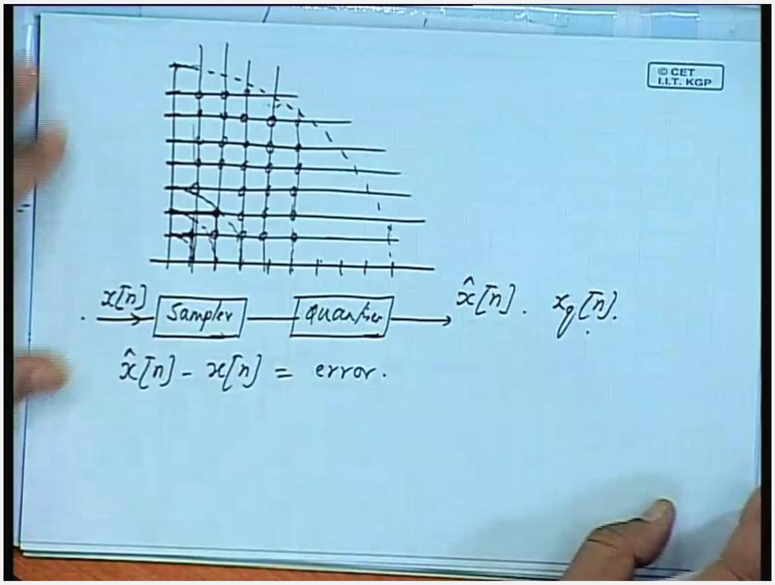
So, what would be the values of r will be 1 by root 8, 2 by root 8, that is half and so on, if you check it approximately 1 by root 8 is 1 by 2 root 2. So, approximately 0.35, 0.5, 0.61, 3 by 8 square root of that then 4 by 8, so 1 by root 2, 0.7 then 0.79 is their approximate values. Then 6 by 8 is root 3 by 2, so 0. 866 and 7 by 8, so that is approximately root 8 and 0.8775, so approximately 0. 94.

So, if I take radius of these values 0.35, so it is somewhere here this is 0.25, so somewhere here and then 0.5. Then 0.6 say somewhere here, 0.7, this is 50.6 should have been 0.6 is somewhat here, this is not there any way like that 0.7 will be somewhere closed to this 0.79. So, like this you can plot it will be somewhere here then 0.866 and 0.94 we can plot it 0.866 and 0.94, I am not plotting all of them. Then what will be the quantification of 2 r cosine theta, 2 is a multiplier of course.

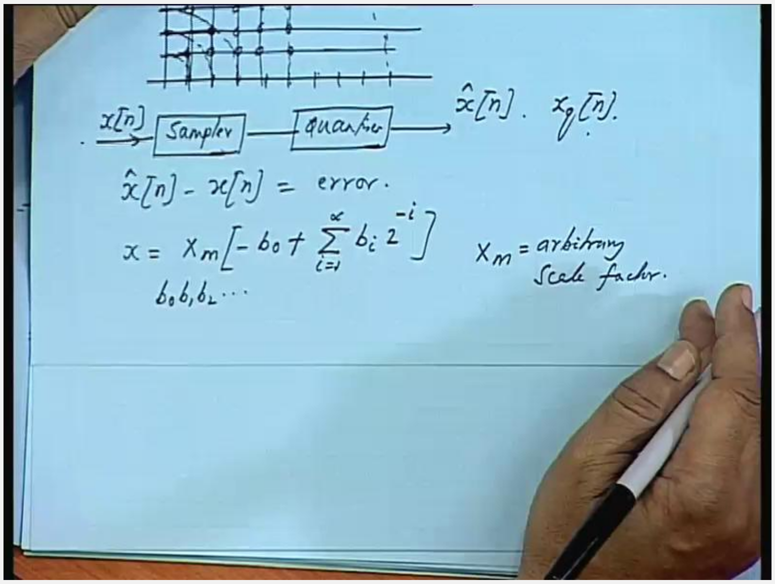
So, I can take even r cosine theta, so there will be just 1 by 8, 2 by 8, 3 by 8 and so on, 5 by 8, 6 by 8, 7 by 8, I have not drawn the other circles as yet, so these are the possible values of r cosine thetas and so on. So, you will find in this zone, they are verified they are all compressed around this these are the possible values the parameters can take with this structure. If you increase the bit representation that is instead of say 2 to the power 3, if I have 2 to the power 4 or 2 to the power 5, 2 to the power 6, if it is an 8 bit number, then this will be partitioned much more closely.

And hence, the distribution will appear more or less uniform; it will be still a little ratified on the lower side and little compressed on the higher side, what you get more discrete values. And then that the discritization will be very, very close. There is another representation of the same quadratic form, this is very interesting, this is r cosine theta, this is z to the power minus 1, this is r sin theta, this is output y n and this is r cosine theta. This is minus r sin theta, this is x n, you try to find out the output and input relation, take it as an exercise. Now, in this case, what would be the possible values of the parameters, they can be only r sin theta and r cosine theta, these are the constants, these are the constants to be quantized.

(Refer Slide Time: 10:51)



$\hat{x}[n] - x[n] =$ error.



$\hat{x}[n] - x[n] =$ error.

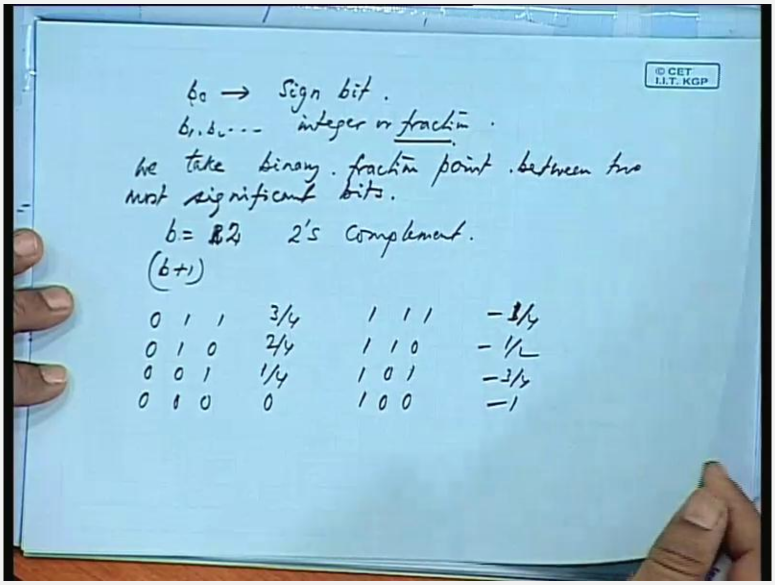
$x = X_m\left[-b_0 + \sum_{i=1}^{\infty} b_i 2^{-i}\right]$ $X_m =$ arbitrary Scale factor.

$b_0 b_1 b_2 \cdots$

So, they look like 1 by 8, 2 by 8, 3 by 8, 4 by 8 and so on. If you draw circles actually need not draw all the circles, it will find these are the grid positions is it not r cosine theta, can take only these possible values. Similarly, r sin theta can take only these possible values, so it will be just a straight grid is bounded by this that is all, so these are the grid function ate, it is a uniform grid, one thing, it does not exceed this value and so on.

Now, this is all about parameter quantization, now let us see the signal, when it quantizes the signal, what is the kind of error that we get, we have the signal x n and then sampler and then you have quantizer. So, this is a quantized quantity x hat n or sometimes you write x q n quantized output, so we are having the difference between these two, will be the error x at n minus x n is the error. Now, any real number x here expressing as some X m in to minus b 0 plus b i 2 to the power minus i, in the 2's complement form, if I have the number b 0, b 1 and b 2 etcetera. Then x is represented accurately by this, if you get go up to an infinite number bits X m is any arbitrary scale factor.

(Refer Slide Time: 14:44)



$b_0 \rightarrow$ Sign bit.

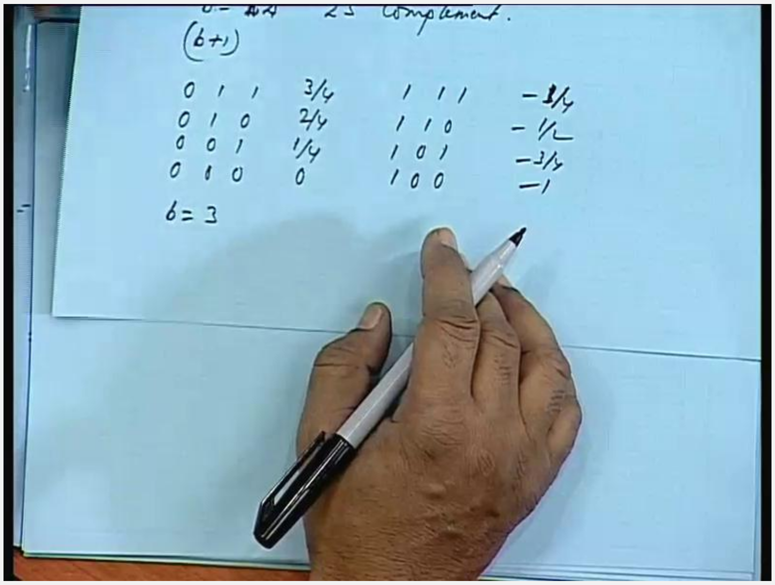
$b_1, b_2 \cdots$ integer or fraction.

We take binary. fraction point. between two most significant bits.

$b = 2$ 2's complement.

$(b+1)$

0	1	1	3/4	1	1	1	−1/4
0	1	0	2/4	1	1	0	−1/2
0	0	1	1/4	1	0	1	−3/4
0	0	0	0	1	0	0	−1



$b = 2$ 2's complement.

$(b+1)$

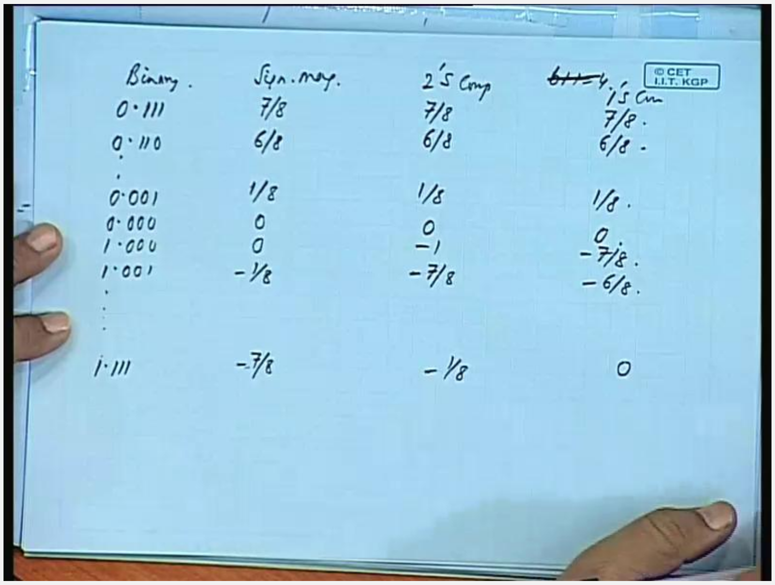
0	1	1	3/4	1	1	1	−1/4
0	1	0	2/4	1	1	0	−1/2
0	0	1	1/4	1	0	1	−3/4
0	0	0	0	1	0	0	−1

$b = 3$

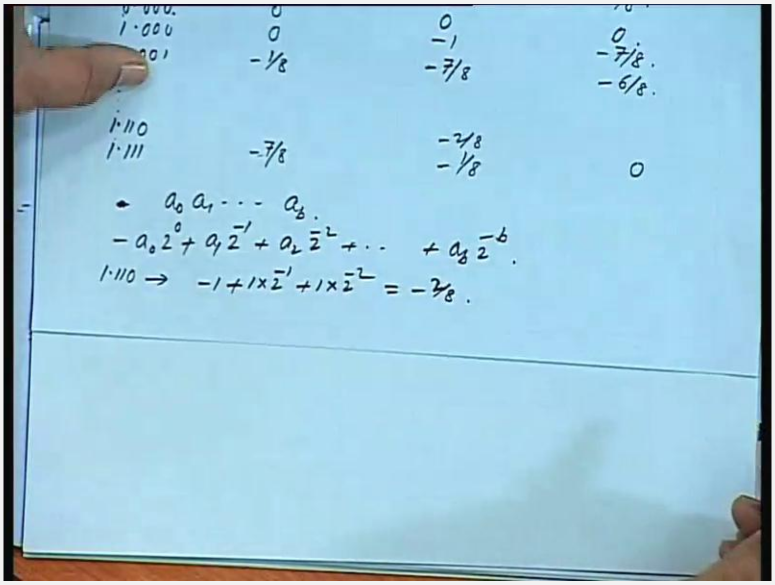
And the 2’s complement the left most bit, this is representing the sign and rest of them are integer or fraction, we take it as fraction. So, between two numbers, we take binary fraction point between two most significant digits, between two most significant bits, so for b equal to 2, so the 2's complement, rather b equal to 3, b equal to 2, the total number of bits is b plus 1, one for the sign bit.

So, b equal to 2 the 2's complement is like, this we write 3 by 4, 2 by 4, 1 by 4 and 0, then 1, 1, 1 as minus 3 by 4, 1, 1, 0 minus half 1, 0, 1, you should minus 1 by 4, minus 3 by 4 and 1, 0, 0 as minus 1. If, you have a three bit modeling that is, if you have b is equal to 3.

(Refer Slide Time: 17:37)



Binary.	Sign.mag.	2's Comp	~~b=4~~. 1's Com
0·111	7/8	7/8	7/8.
0·110	6/8	6/8	6/8.
⋮			
0·001	1/8	1/8	1/8.
0·000	0	0	0.
1·000	0	−1	−7/8.
1·001	−1/8	−7/8	−6/8.
⋮			
1·111	−7/8	−1/8	0



0·000.	0	0	0
1·000	0	−1	−7/8.
1·001	−1/8	−7/8	−6/8.
⋮			
1·110		−2/8	
1·111	−7/8	−1/8	0

$a_0 a_1 \cdots a_b.$

$-a_0 2^0 + a_1 2^{-1} + a_2 2^{-2} + \cdots + a_b 2^{-b}.$

$1 \cdot 110 \rightarrow -1 + 1 \times 2^{-1} + 1 \times 2^{-2} = -2/8.$

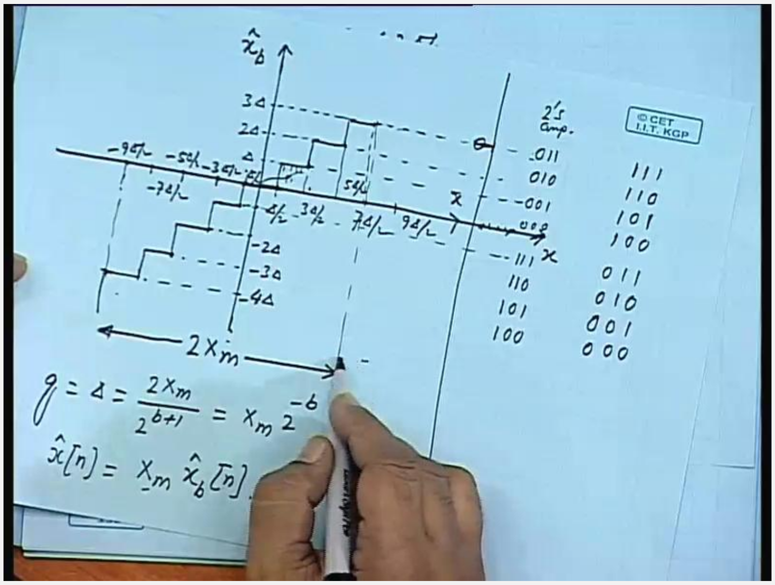
Now, let us write in 2's complement 1's complement and simple sign magnitude, so what was the represent 0.111, the decimal point is here, I did not show.

This is the decimal point, so similarly for three bits to represent the fraction and 1 bit, here so it will be b plus 1 is 4, so this a binomial binary representation, this is the sign magnitude representation. This is 2's complement and this is 1's complement, so is 0.111 will represent 7 by 8, 7 by 8, 7 by 8, 0.110, similarly 6 by 8, 6 by 8, 6 by 8, so like that when it comes to 0.01, it will be 1 by 8, 1 by 8, 1 by 8, the change starts now 0. 000 will it be 0, 0, 0.

Then, 1.000 will be 0 minus 1 minus 7 by 8, 1.000 will be minus 1 by 8 minus 7 by 8 minus 6 by 8 and lastly if you continue like this will 1.111, it should be minus 7 by 8 minus 1 by 8 and 0. If you remember, we discussed in 1's complement notation 0. 000 is same as 0.111 is like plus minus 0.

Similarly, 0 here is 0.000, so in general suppose we have a 0, a 1 up to a b, it is b number of bits total number of bits is b plus 1. Then the number is in 2's complement is a 0 minus a 0 plus 2 to the power 0 plus a 1, 2 to the power minus 1 plus a 2, 2 to the power minus 2, and so on. So, for example, 1.110 will be minus 1, 2 to the power 0, means always 1 plus a 1 is 1 in to 2 to the power minus 1 plus 1 in to 2 to the power minus 2, so that gives me minus 1 by 8. After this it will appear somewhere in between 1.111 is this coming 1.110 is one-forth and half, so how much is it minus 2 by 8 is it, so it will be 2 by 8, it will tally with the previous 1.

(Refer Slide Time: 22:28)



$$q = \Delta = \frac{2X_m}{2^{b+1}} = X_m 2^{-b}$$

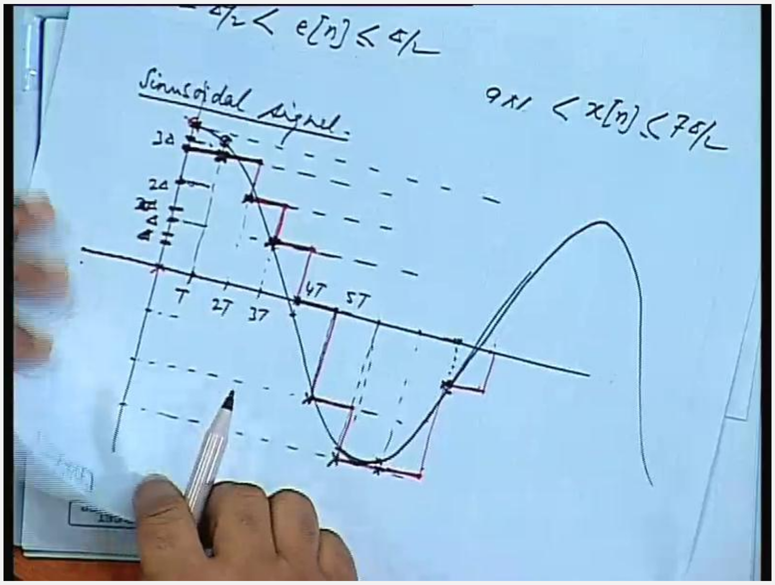
$$\hat{x}[n] = X_m \hat{x}_b[n].$$

Now, suppose we have a signal varying from some plus magnitude to minus magnitude, so you scale it by that factor X m and then we try to represent by this quantized variables, so let me sketch it. So, this level is delta this level is 2 delta, this level is 3 delta, this is delta by 2, 3 delta by 2, 5 delta by 2, 7 delta by 2 and it continues. Again, 9 delta by 2 and so on is the variable x and this is a quantized variable x hat, similarly here this is minus delta, minus 2, delta minus 3 delta, minus 4 delta.

This is minus 3 delta by 2, this is minus delta by 2 minus 5 bit delta by 2 minus 7 delta by 2 minus 9 delta by 2, so this is 0,11, I will show it here this level is 011, 010. This level is 001, this is 000, then this one is 111, 110, 101, 100, this is 2's complement, if take off set off set code, this will be just 1 less on this side will be 111, 110, 101. This one will be 100, similarly this one; you can fill up 011, then 010, 001 and 000.

Suppose, this is the range you call twice X m, so the quantized, what the quantization level, that is delta is twice X m divided by 2 to the power b plus 1, b plus 1 is the code length total. So, that will be X m, 2 to the power minus b, this is the value of delta, x hat n is X m in to x b hat n, if I call it x b hat, so this is basically scaled between plus minus 1. You can say and that that has to be finally multiplied by this factor X m, so that will give you the actual quantized value.

(Refer Slide Time: 28:21)



$\Delta/2 < e[n] \leq \Delta/2$
Sinusoidal signal.
$9\Delta/2 < x[n] \leq 7\Delta/2$
3Δ
2Δ
Δ
T
2T
3T
4T
5T

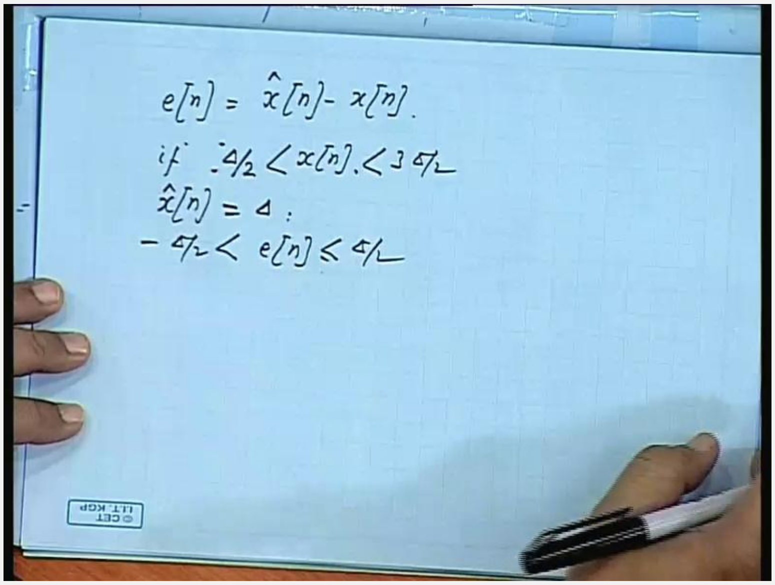
Suppose, you have a sinusoidal signal varying between say plus minus 1, 2 to 4 to 6, 2 to 4 to 6, suppose we have something like this. Now, you may have quantization levels like this, say delta somewhere here delta, 2 delta, 3 delta, I have chosen the lengths little shorter delta, 2 delta and say 3 delta. These are the levels say and suppose the sampling time is T 2, T 3, T 4, T 5, T, and so on, so at this moment, what we are sampling is this value and what will that be represented as 3 delta.

So, this is the sample value sample value, I put as a cross at 2 T our value is here that is approximated as this quantity at 3 T, it is coming over here and that is sampled as delta itself. At 4 T it is this it will be close to 0 and 5 T, it is here it will be like this and at 6 T is here, so it will be sampled as this. So, the values are like this, if I join this sinusoid will appear as say previous to this, I have not shown you the sample value.

If, I start a sampling here, so it will be also like this, so it will be a function like this, it will continue till the next instant, then it will come here, then it will be like this and will be like this, then it will be like this. Similarly, if I sampled here that will also be approximated here, if I take the next sample, somewhere here it will be like this, so it will be approximated here and so on. So, you can see may be somewhere closed to 0, this will be the representation that this red blocks, that will be the representation of the sinusoid, it will be changing.

If, I change the sampling time, it will be approximating closed to the sinusoid, if I have more number of bits, so that is why for a very correct representation, you should have a large number of bits. And also, you should have a very high sampling rate, if our sampling rate is high, then atleast these blocks will not this error, will not continue for such a long time.

(Refer Slide Time: 33:28)

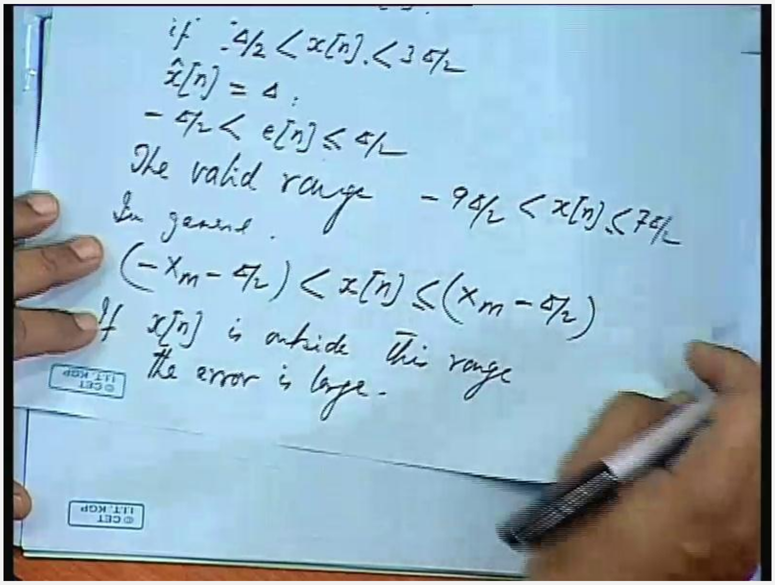


$e[n] = \hat{x}[n] - x[n].$

if $\Delta/2 < x[n] < 3\Delta/2$

$\hat{x}[n] = \Delta$;

$-\Delta/2 < e[n] \leq \Delta/2$



if $\Delta/2 < x[n] < 3\Delta/2$

$\hat{x}[n] = \Delta$;

$-\Delta/2 < e[n] \leq \Delta/2$

The valid range $-9\Delta/2 < x[n] \leq 7\Delta/2$

In general.

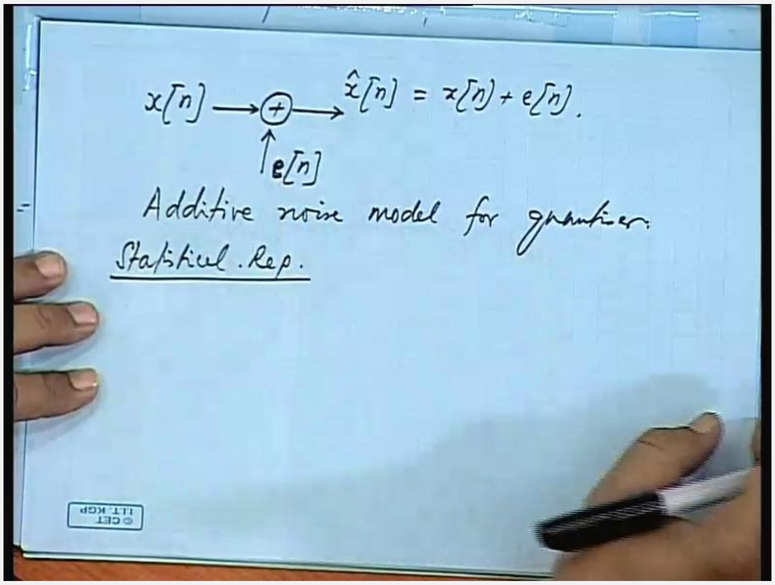
$(-X_m - \Delta/2) < x[n] \leq (X_m - \Delta/2)$

If $x[n]$ is outside this range the error is large.

So, error e n is x hat n minus x n, so if you have delta by 2 x n 3 delta by 2, suppose x n is lying between delta by 2 and 3 delta by 2 between this point and this point, the actual signal is passing through this value, so this is the value of x. So, the error so x hat n will be taken as delta and minus delta by 2, so error is between plus minus delta by 2 is it not it can be it can vary only between these 2 points. So the valid range is minus 9 delta by 2 to 7 delta by 2, you have seen it here, from minus 9 delta by 2 to 7 delta by 2, this is the range of variation of x.

In general, it will be minus X m minus delta by 2, plus X m minus delta by 2, so this is X m minus delta by 2, that is minus X m minus delta by 2 and plus X m minus delta by 2 whole thing has been shifted by delta by 2 on 2 side both sides. If x n is outside this range, for example in the sinusoid, you have chosen less number of steps, if it is outside and if it is more, then this will not be the error will be very large, error will not be restricted within this. Error is large not within this limit, so if you quantize and if you restrict your value of x n within this limit.

(Refer Slide Time: 36:59)

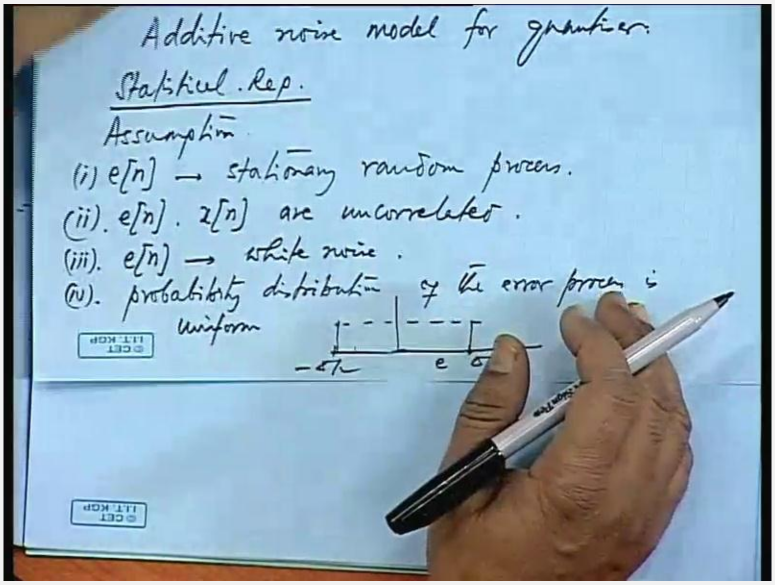


$x[n] \longrightarrow \oplus \longrightarrow \hat{x}[n] = x[n] + e[n].$

$\uparrow e[n]$

Additive noise model for quantizer.

Statistical. Rep.



Additive noise model for quantizer.

Statistical. Rep.

Assumption.

(i) $e[n] \rightarrow$ stationary random process.

(ii). $e[n]$. $x[n]$ are uncorrelated.

(iii). $e[n] \rightarrow$ white noise.

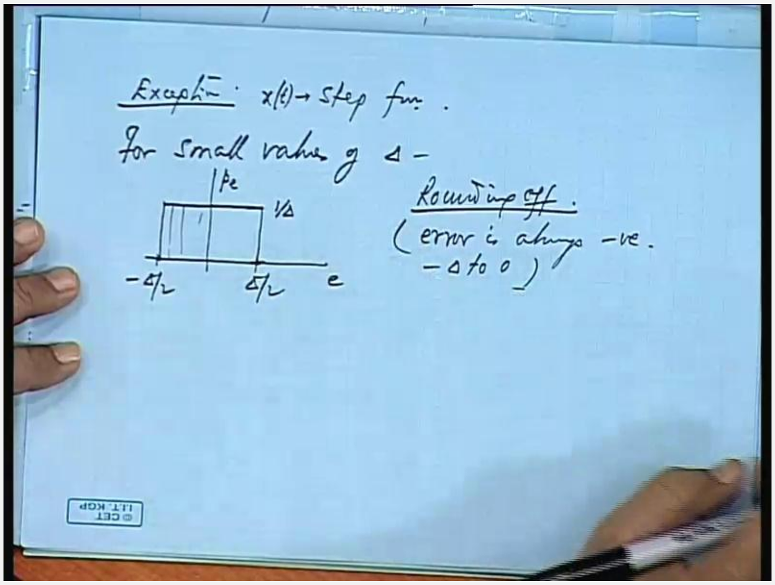
(iv). probability distribution of the error process is uniform

Then, we can have a linear model for error, additive noise can be representing the error, we can write x n plus e n is equal to x hat n, now this is known as additive noise. So, this is the model for the quantizer, now what will be the statistical representation of this error, their noise, we are making certain assumptions at this stage about the property of this noise e n. The first assumption is e n is stationary random process, it all depends on the signal, if the signal is varying, then you can have a random process representation.

Suppose, it is a step function, then it does not change is it not, so that time we cannot have that is an exception, but for the general signals it will be always changing. So, you can assume that the error is a stationary and random, error is stationary means it is restricted within minus delta by 2 to plus delta by 2 and it can vary within that with equal probability, so it is a random process. Then next assumption is en has got no relation with x n, e n and x n are un correlated, then the variables which decide e n are also un correlated, that means any noise. When, it is not correlated by any factor, we call it white noise, say for example a very good example will be you measure the voltage of our node in a circuit or may be this supply voltage. You keep measuring, you will find it is suppose to be 220, sometimes at this moment, it may be 219, next moment it may be 220.5.

So, it is keep on changing fluctuating around that value 220 and how much is that fluctuation that is unpredictable, there can be an electromagnetic noise outside, there may have loose connections somewhere and somebody may be switching on or off, some load. So, that may cause some line disturbance, so that is an aggregate of various factors various on certain factors and that variation can be taken as that noise can be taken as a white noise. So, e n is a white noise, a white noise means it has all possible frequency components, then probability distribution of the error process is uniform. That means the error can occur, say this is from minus delta by 2 to plus delta by 2 with uniform probability, it may be anywhere with equal uncertainty.

(Refer Slide Time: 42:18)



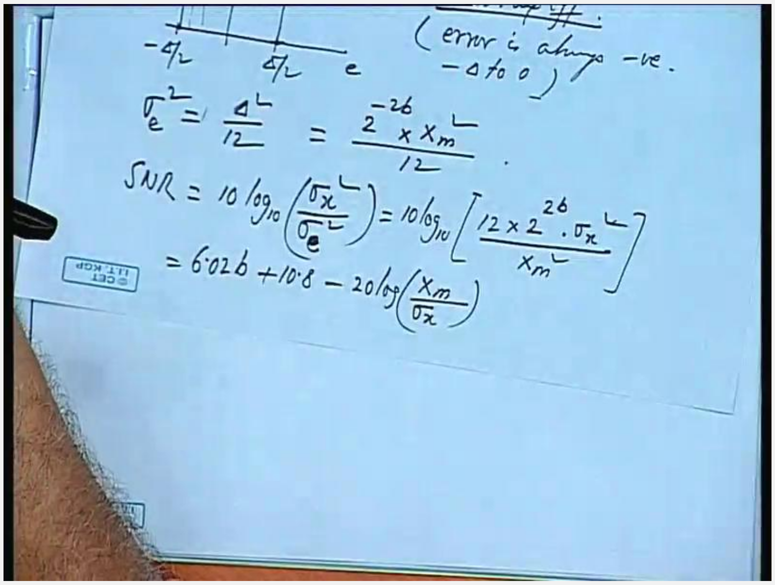
Exception: $x(t) \rightarrow$ step fn.

for small value of Δ -

p_e
$1/\Delta$
$-\Delta/2$
$\Delta/2$
e

Rounding off.

(error is always -ve.
$-\Delta$ to 0)



$-\Delta/2$ $\Delta/2$ e

(error is always -ve.
$-\Delta$ to 0)

$$\sigma_e^2 = \frac{\Delta^2}{12} = \frac{2^{-2b} \times X_m^2}{12}.$$

$$SNR = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 10\log_{10}\left[\frac{12 \times 2^{2b} \cdot \sigma_x^2}{X_m^2}\right]$$

$$= 6.02b + 10.8 - 20\log\left(\frac{X_m}{\sigma_x}\right)$$

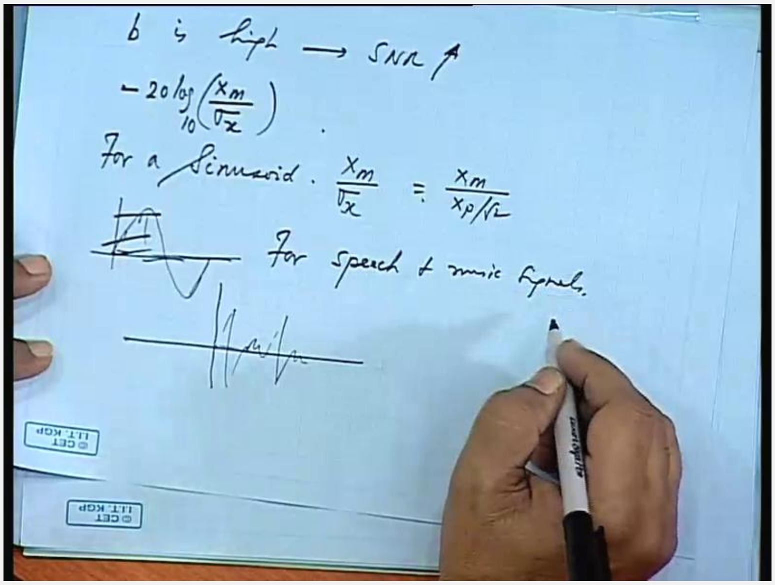
So, this will be in the only exception is for a signal like a step function, if x n is step or x t is a step function, then you have a you do not have this kind of an error. If the signal is very complicated like the speech signal or music signal, where you have large variation in the… And very frequent variations, this thing this kind of error modeling will be very accurate for small values of delta. The uniform distribution, it will be more close to a uniform distribution, this is the probability, this will be of height 1 by delta total probability is 1.

So, this is the probability density function, if you go for is a truncation, if you go for rounding off, then error is always negative 5..356. I round it off to is the other way around a truncation. It will be always negative, is the rounding off error is always negative between minus delta and 0, this think over, this is for rounding off, it will be delta to 0.

So, from here sigma e squared will be delta squared by 12, you take this square of the error integrate and then find out from their sigma e square is 2 to the power minus 2 b into X m square divided by 12. So, signal to noise ratio SNR, which is defined as sigma x square, variance of the signal divided by variance of the error is 10 log of substitute for sigma e square this quantity.

So, it will be twelve into 2 to the power 2 b into sigma x squared divided by X m squared, so that gives me 6.02 b plus 10. 8 minus 20 log of X m by sigma x, x square I have taken that 2 on this side, so that has become 20 and I have just inverted it, so it has become minus. Now, if you see from here signal to noise ratio, if it is high that means the noise component is very much reduced and it will be high, if I increase 1 bit more, if I have the representation 1 bit more, I gain by 6 d B almost.

(Refer Slide Time: 47:08)

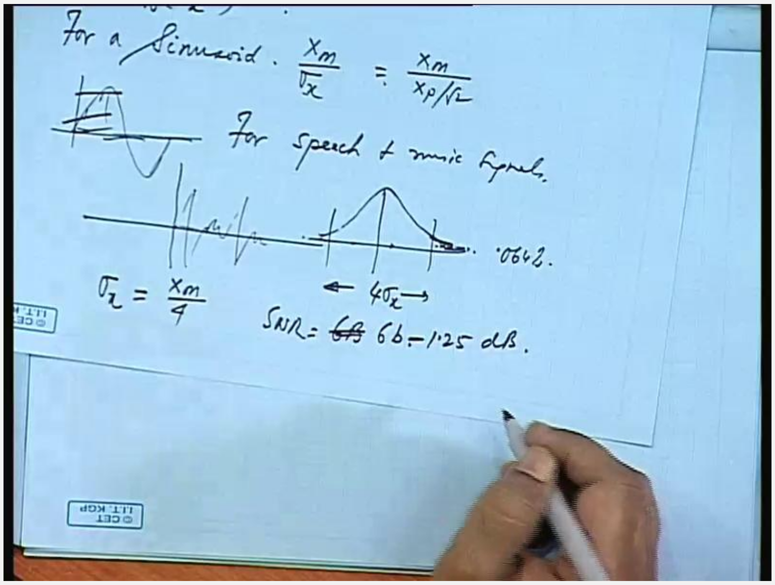


b is high → SNR ↑

$-20\log_{10}\left(\frac{X_m}{\sigma_x}\right)$.

For a Sinusoid. $\frac{X_m}{\sigma_x} = \frac{X_m}{X_p/\sqrt{2}}$

For speech + music signals.



For a Sinusoid. $\frac{X_m}{\sigma_x} = \frac{X_m}{X_p/\sqrt{2}}$

For speech + music signals.

.0642.

$\sigma_x = \frac{X_m}{4}$

$\leftarrow 4\sigma_x \rightarrow$

SNR = ~~6B~~ 6b - 1.25 dB.

So, if b is high SNR goes up, that means the noise component reduces by 6 d B per bit and this 10ds to this term minus 20 log of X m by sigma x, for a sinusoid X m by sigma x at the most. This will be x peak by root 2 is it not, the variance for a sinusoid, you can say variance for a sinusoid is it is a 0, mean process, so its RMS value is basically giving you variance, so it is x P by root 2.

So if you restrict if you restrict x P, if you take a very small sinusoid the this quantity will be large, so SNR ratio will be small that means, if you take a very small signal compare to the available range, then you are likely to have very poor SNR. So and if you exceed it, if you take a very large quantity, very large value of x P, then there will be chopping, then also you suffer a lot. So, when you are go you are going for a to d conversion, you must be careful, you should choose the scale, such that it matches with the range of variable.

For again, signals like speech signal or music signal the variation is very large and the amplitude is always centered around the 0 value, it changes on both sides so for speech and music signals. Now, it can be shown it tends to be a normal distribution, in a normal distribution, if I take 4 times sigma x, if sigma x is the variance, 4 times this, then how many what percentage of the points will be lying above this or below.

This it is about 0. 064 percent very, very small fraction 3 sigma covers almost 99 more than 99 percent, so 0.064 percent above 4 sigma, so practically no signal is present here. Therefore, if you had adjust sigma x to X m by 4, then SNR will be 6 b, b means that bit length minus 1, minus 1.25 d B. So, that is coming from here X m by sigma x is 4, so 20 log of 4 is 0.6, so 20 to 0.6, 12, so minus what 10.8 minus approximately 12, that gives me this. So, this is all about very brief description about the error, that we come across in quantization and we have already discussed about the quantization of parameters. In the next class will take up some examples.

Thank you very much.