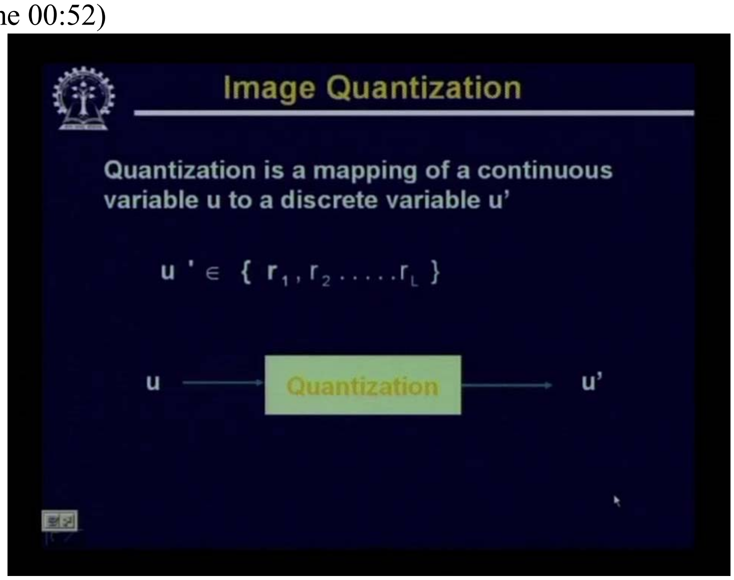
Digital Image Processing
Prof. P. K. Biswas
Department of Electronics and Electrical Communications Engineering
Indian Institute of Technology, Kharagpur
Module 01 Lecture Number 06
Quantizer Design

(Refer Slide Time 00:17)

Hello, welcome to the course on Digital Image Processing. The first phase of the image digitization process, that is quantization and we have also seen through the examples of these reconstructed image that if we vary the sampling frequency below and above the Nyquist rate, how the quality of the reconstructed image is going to vary. So now let us go to the second phase that is quantization of the sample values

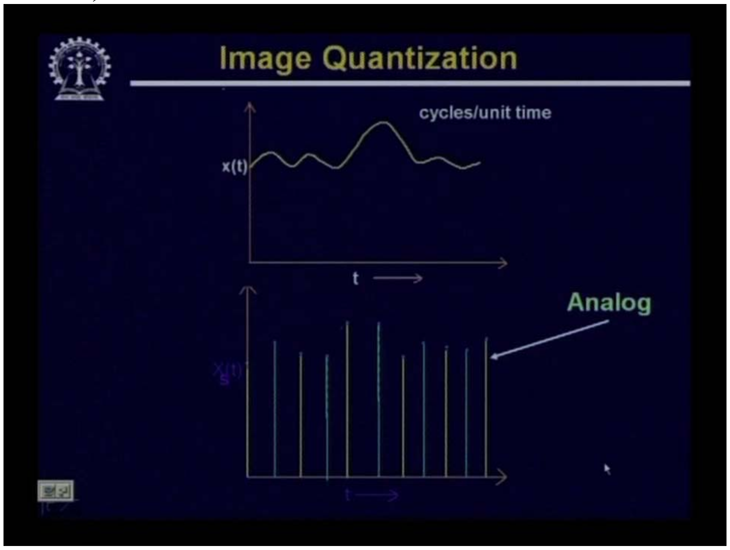
(Refer Slide Time 00:52)



Now this quantization is a mapping of the continuous variable u to a discrete variable u', where u' takes values from a set of discrete variables. So if your input signal is say u, after quantization the quantized signal becomes u', where u' is one of the discrete variables as shown in this case as r_1 to r_L. So we have L number of discrete variables from r_1 to r_L and u' takes a value of one of these variables.

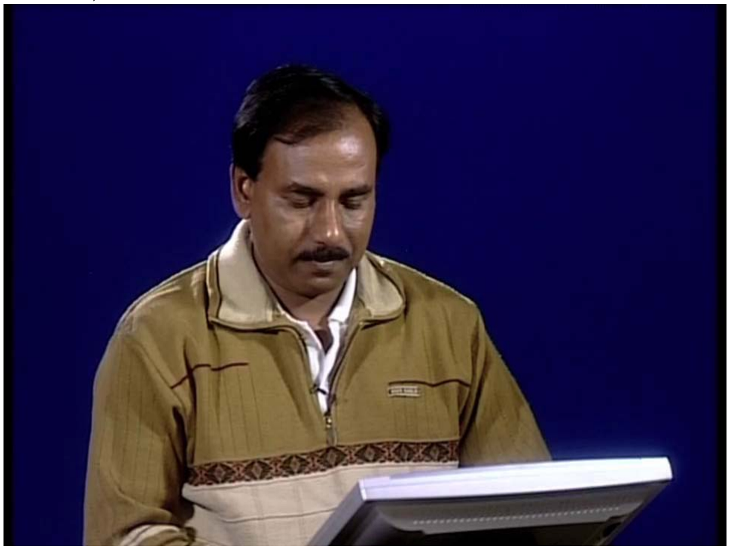
Now what is this quantization?

(Refer Slide Time 01:36)



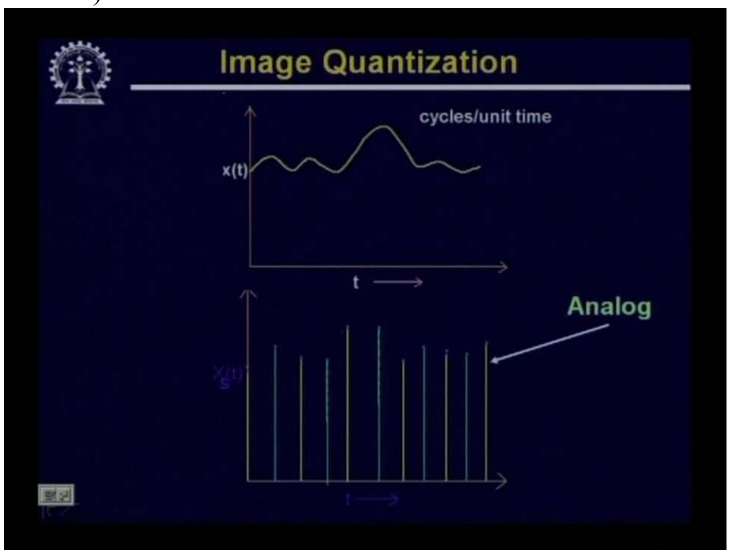
You find that after sampling of a continuous signal, what we have got is a set of samples. These samples are discrete in time domain, Ok.

(Refer Slide Time 01:50)



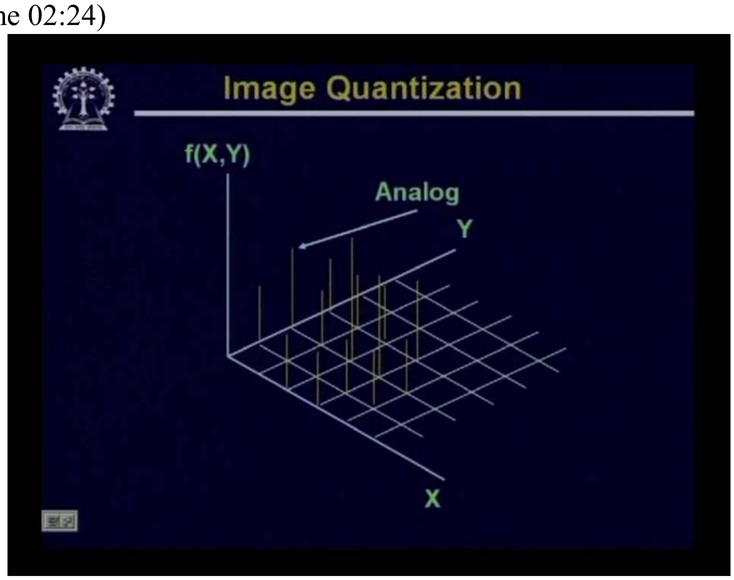
But still every sample value is an analog value. It is not a discrete value. So what we have done after sampling is, instead of considering all possible time instants, the signal values at all possible time instants, we have considered the signal values at some discrete time instants.

(Refer Slide Time 02:12)



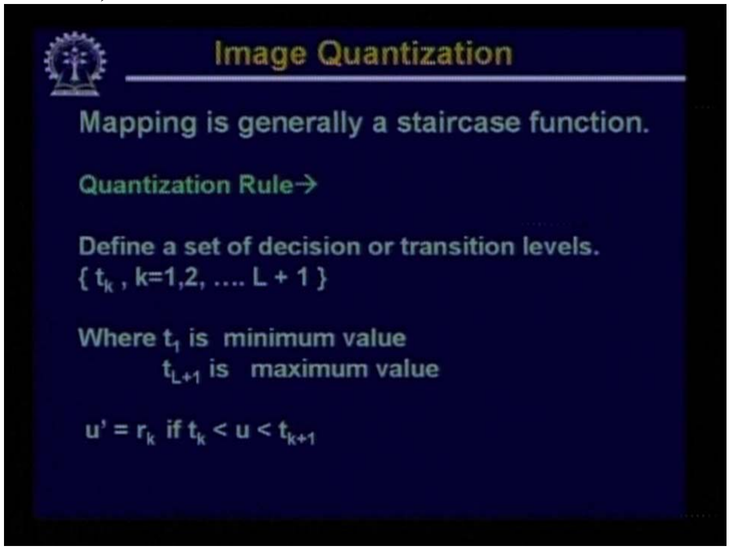
And at each of these discrete time instance, I get a sample value. Now the value of this sample is still an analog value. Similar is the case with an image.

(Refer Slide Time 02:24)



So here, in case of an image the sampling is done in two-dimensional grids where at each of the grid locations, we have a sample value which is still analog. Now if I want to represent a sample value on a digital computer, then this analog sample value cannot be represented. So I have to convert this sample value again in the discrete form. So that is where the quantization comes into picture.

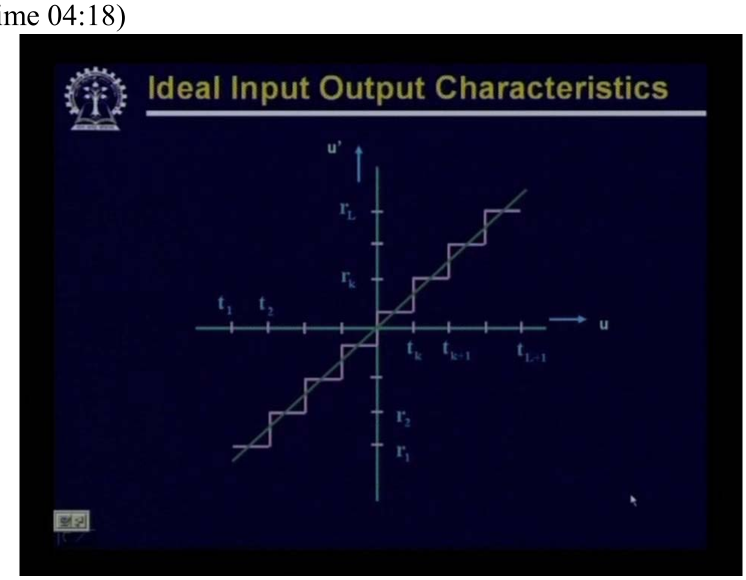
(Refer Slide Time 02:57)



Now this quantization is a mapping which is generally a staircase function.

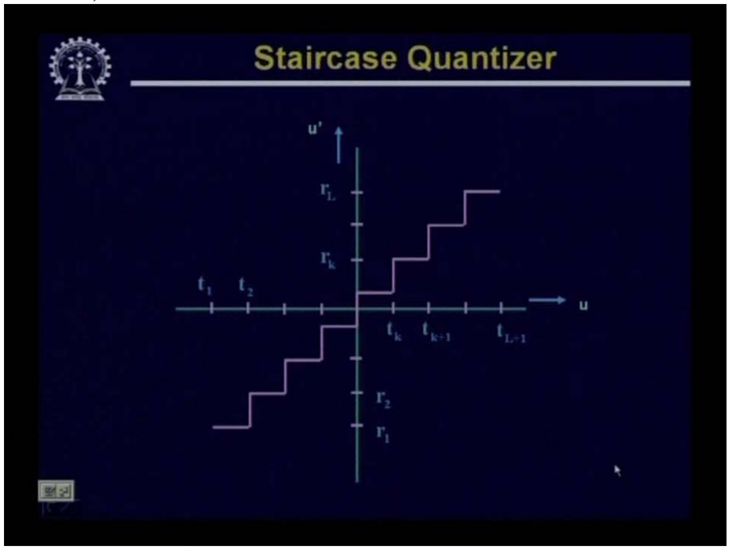
So for quantization what is done is, you define a set of decision or transition levels which in this case has been shown as transition level t_k, where k varies from 1 to L+1. So we have defined a number of transition levels or decision levels which are given as t_1, t_2, t_3, t_4 up to t_{L+1}, Ok and here t_1 is the minimum value and t_{L+1} is the maximum value. And you also define a set of the reconstruction levels that is r_k. So what we have shown in the previous slide that the reconstructed value u' takes one of the discrete values r_k, so the quantized value will take the value r_k if the input signal u lies between the decision levels r_k and r_{k+1}. So this is how you do the quantization.

(Refer Slide Time 04:18)



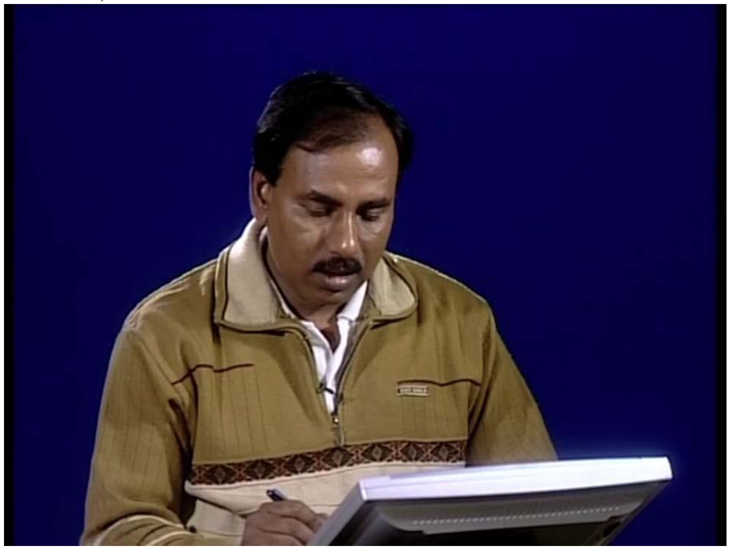
So let us come to this particular slide.

(Refer Slide Time 04:21)



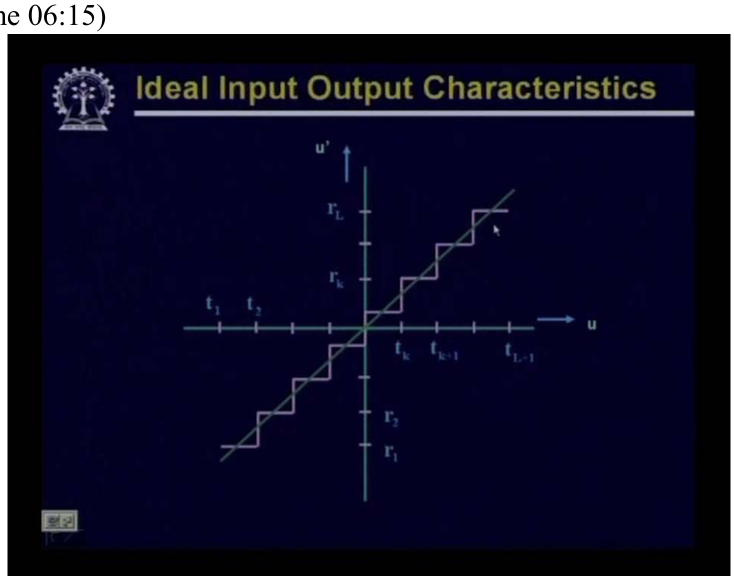
So it shows the input output relationship of a quantizer. So it says whenever your input signal u, so along the horizontal direction we have put the input signal u and along the vertical direction we have put the output signal u' which is the quantized signal. So this particular figure shows that if your input signal u lies between the transition levels t_1 and t_2, then the reconstructed signal or the quantized signal will take the value r_1. If the input signal lies between t_2 and t_3, the reconstructed signal or the quantized signal will take a value r_2. Similarly, if the input signal lies between t_k and t_{k+1}, then the reconstructed signal will take the value of r_k and so on. So given an input signal which is analog in nature, you are getting the output signals which have, which is discrete in nature. So the output signal can take only one of these discrete values. The output signal cannot take any arbitrary value.

(Refer Slide Time 05:34)



Now let us see that what is the effect of this? So as we have shown in this second slide that ideally we want that whatever is the input signal, the output signal should be same as the input signal and that is necessary for the perfect reconstruction of the signal. But whenever we are going for quantization, your output signal, as it takes one of the discrete set of values, is not going to be same as the input signal always. So in this, in this particular slide, again we have shown the same staircase function

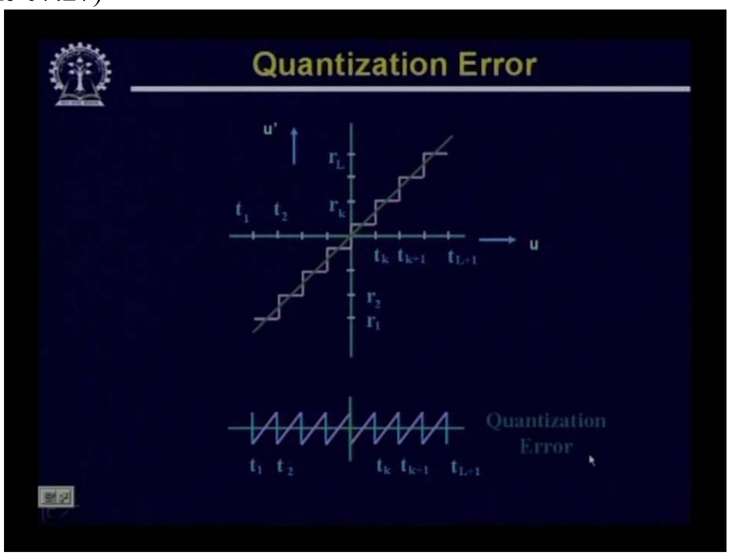
(Refer Slide Time 06:15)



where along the horizontal direction we have the input signal and in the vertical axis we have put the output signal.

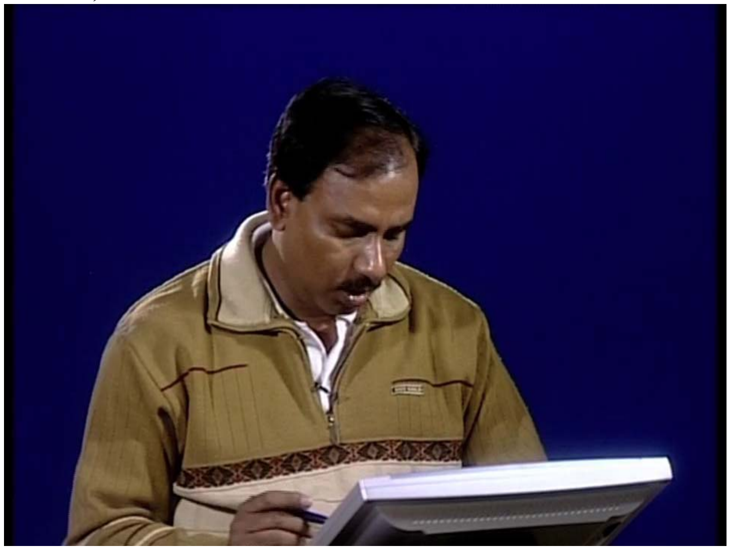
So this pink staircase function shows what is the quantization function that will be used and this green line which is inclined at an angle of 45° with the u axis, this shows that what should be the ideal input output characteristics. So if the input output function follows this green line, in that case, for every possible input signal I have the corresponding output signal. So the output signal should be able to take every possible value. But when you are using this staircase function, in that case, because of the staircase effect, whenever the input signal lies within certain region, the output signal takes a discrete value. Now because of this staircase function, you are always introducing some error in the output signal or in the quantized signal. Now let us see that what is the nature of this error.

(Refer Slide Time 07:27)



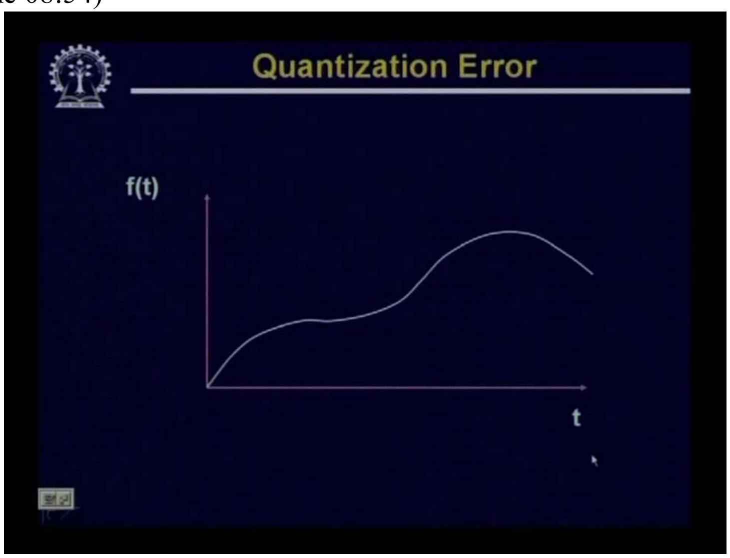
Here we have shown the same figure. Here you find that when this green line which is inclined at 45° with the u axis crosses the staircase function, at this point whatever is your signal value, it is same as the reconstructed value. So only at these crossover points, your error in the quantized signal will be 0. At all other points, the error in the quantized signal will be a non-zero value. So at this point the error will be maximum which will, maximum and negative, which will keep on reducing. At this point, this is going to be 0, and beyond this point again it is going to increase. So if I plot this quantization error, you find that the plot of the quantization error will be something like this, between every transaction levels. So between t_1 and t_2, the error value is like this. Between t_2 and t_3, the error continuously increases. Between t_3 and t_4, error continuously increases and so on. Now what is the effect of this error on the reconstructed signal?

(Refer Slide Time 08:44)



So for that let us take again a one-dimensional signal f(t), which is a function of t as is

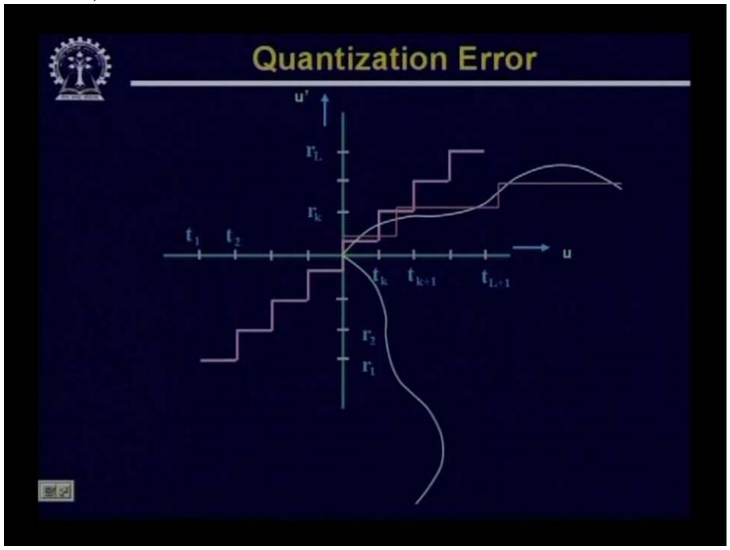
(Refer Slide Time 08:54)



shown in this slide

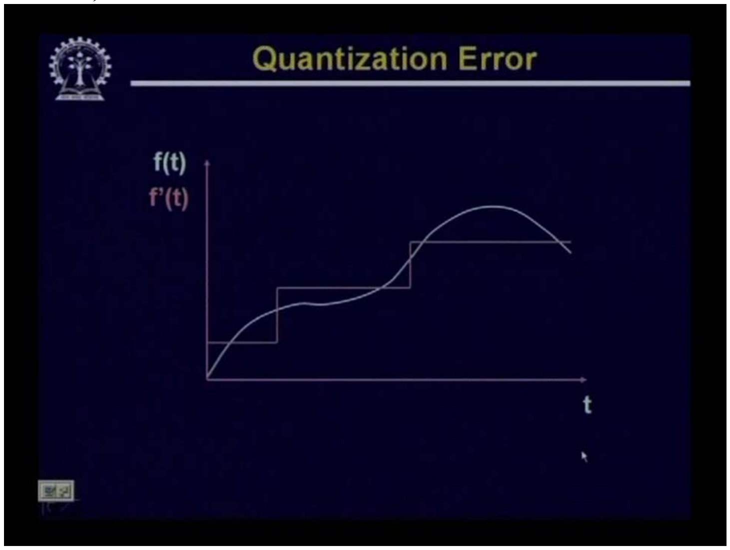
and let us see that what will be the effect of quantization on the reconstructed signal.

(Refer Slide Time 09:04)



So here we have plotted the same signal, Ok. So here we have shown the signal is plotted in the vertical direction so that we can find out what are the transition levels or the part of the signal which is within which particular transition level. So you find that this part of the signal is in the transition level say t_{k-1} and t_k. So when the signal, input signal lies between the transition t_{k-1} and t_k, the corresponding reconstructed signal will be r_{k-1}. So that is shown by this red horizontal line. Similarly, the signal from this portion to this portion lies in the range t_k and t_{k+1}. So corresponding to this, the output reconstructed signal will be r_k, so which is again shown by this horizontal red line. And this part of the signal, the remaining part of the signal lies within the range t_{k+1} and t_{k+2} and corresponding to this, the output reconstructed signal will have the value r_{k+1}. So to have a clear figure,

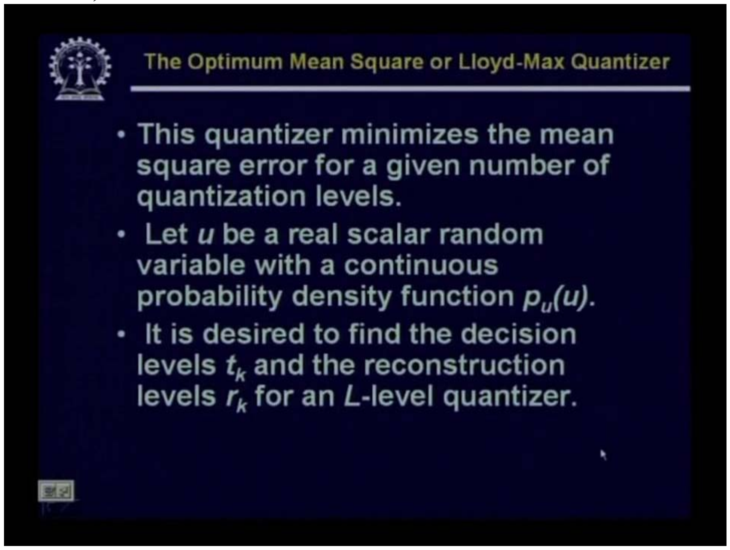
(Refer Slide Time 10:22)



you find that in this, the green curve, it shows the original input signal and this red staircase lines, staircase functions it shows that what is the quantization signal, quantized signal or f'(t).

Now from this, it is quite obvious that I can never get back the original signal from this quantized signal, because within this region the signal might have, might have had any arbitrary value. And the details of that is lost in this quantized form, quantized output. So because from the quantized signal I can never get back the original signal so we are always introducing some error in the reconstructed signal which can never be recovered. And this particular error is known as quantization error or quantization noise. Obviously the quantization error or quantization noise will be reduced, if the quantizer step size that is the transition interval say t_k to t_{k+1} reduces. Similarly, the reconstruction step size, r_k to r_{k+1}, that interval is also reduced.

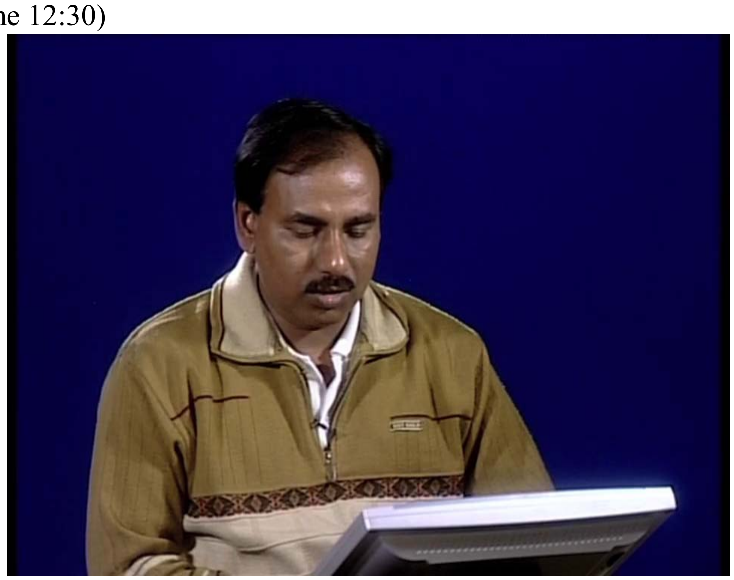
(Refer Slide Time 11:42)



So for quantizer design, the aim of the quantizer design will be to minimize this quantization error. So accordingly we have to have an optimum quantizer and this Optimum Mean Square Error quantizer known as Lloyd-Max Quantizer, this minimizes the mean square error for a given a given number of quantization levels.

And here we assume that let u be a real scalar random variable with a continuous probability density function $p_u(u)$. And it is desired to find the decision levels t_k and the reconstruction levels r_k for an L level quantizer which will reduce or minimize

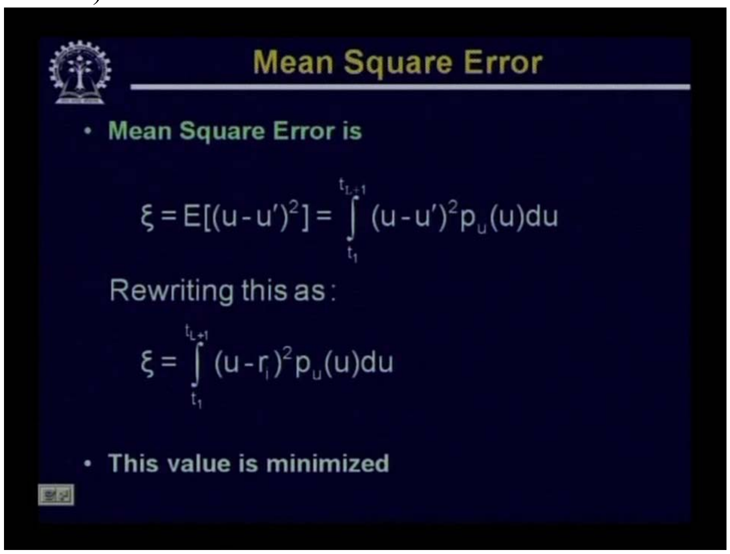
(Refer Slide Time 12:30)



the quantization noise or quantization error. Let us see how to do it.

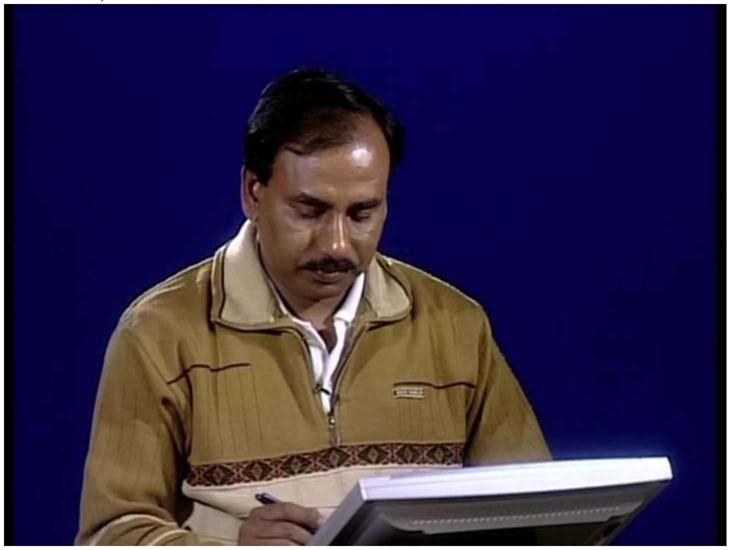
Now you remember that u is the input signal and u' is the quantized signal. So the error of reconstruction is the input signal

(Refer Slide Time 12:52)



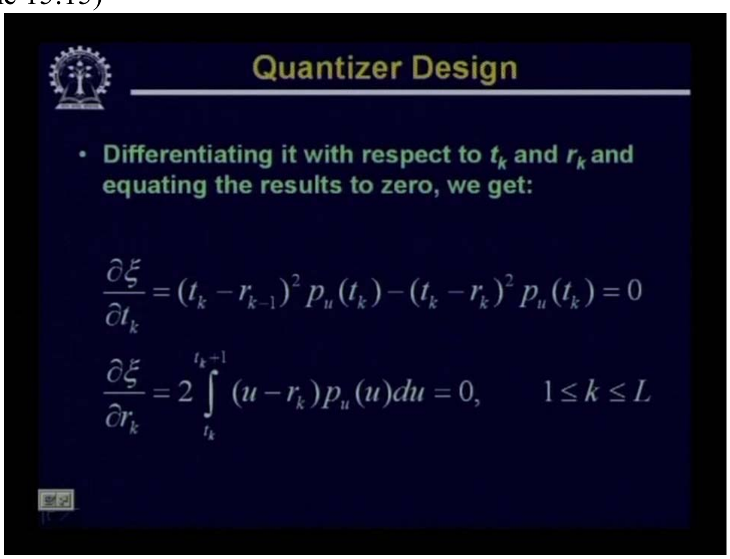
minus the reconstructed signal. So the mean square error is $\xi = E\left[(u-u')^2\right] = \int_{t_1}^{t_{L+1}} (u-u')^2 p_U(u)d(u)$, you find that, you remember that t_1 was the minimum transition level and t_{L+1} was the maximum transition level. So if I just integrate this function, $\xi = \int_{t_1}^{t_{L+1}} (u-u')^2 p_U(u)d(u)$, I get the mean square error. This same integration can be rewritten in this form as $\xi = \int_{t_i}^{t_{i+1}} (u-r_i)^2 p_U(u)d(u)$ because r_i is the reconstruction level of the reconstructed signal in the interval t_i to t_{i+1}. So I integrate this, $\int_{t_i}^{t_{i+1}} (u-r_i)^2 p_U(u)d(u)$, then I have to take a summation of this for i equal to 1 to L, Ok. So this modified expression will be same as this and this tells you that what is the square error of the reconstructed signal. And the purpose of designing the quantizer will be to minimize this error value.

(Refer Slide Time 14:44)



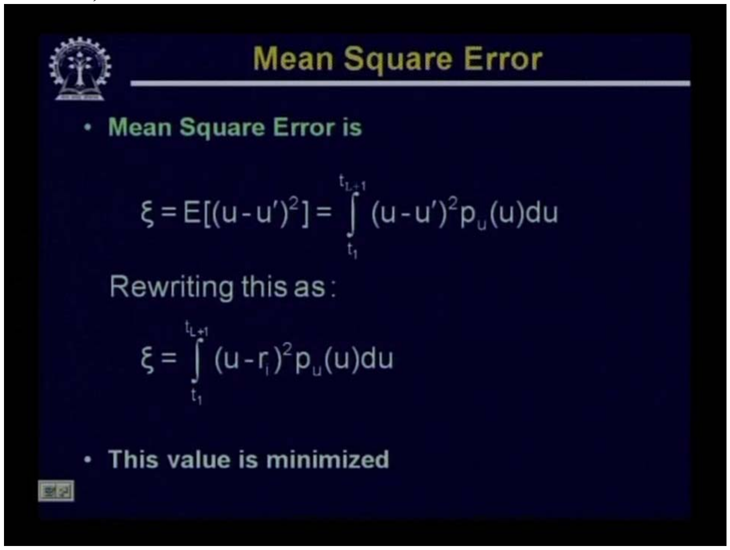
So obviously from school level mathematics we know that for minimization of the error value, because now we have to design the transition levels and the reconstruction levels which will minimize the error, so the way to do is, to do that is to differentiate the error function, the error value with t_k and with r_k and equating those equations to 0. So if I differentiate this particular

(Refer Slide Time 15:15)



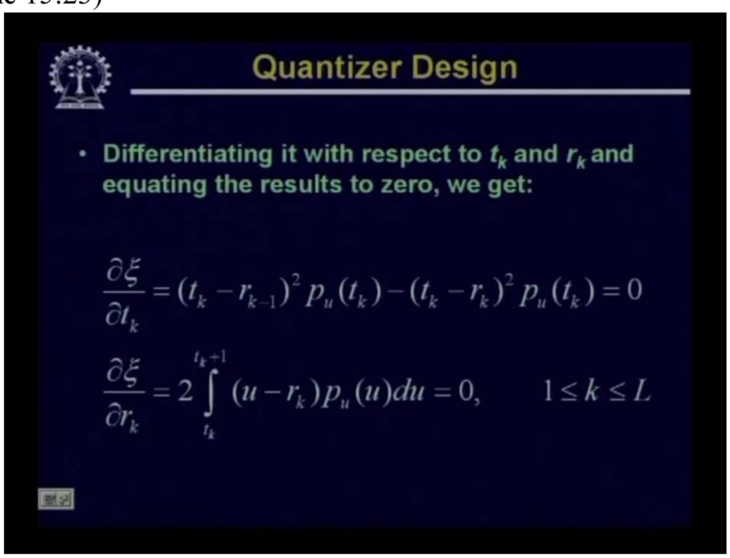
error value

(Refer Slide Time 15:16)



$\xi = \int_{t_i}^{t_{i+1}} (u-r_i)^2 p_U(u)d(u)$, in that case what

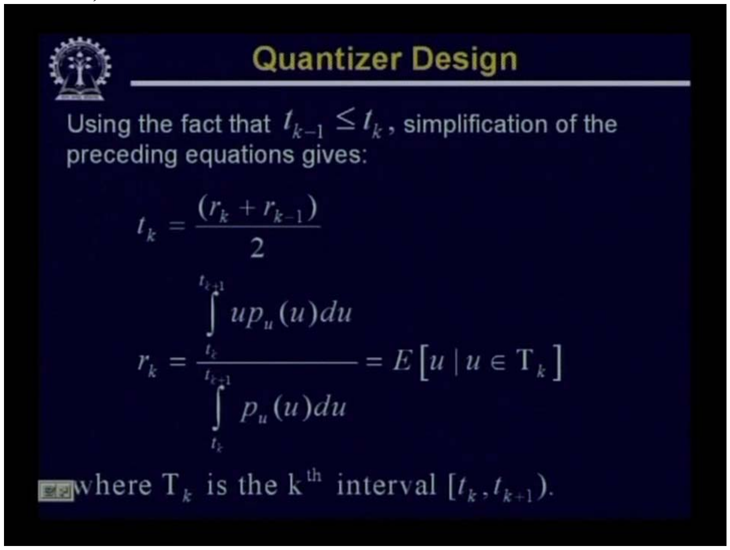
(Refer Slide Time 15:23)



I get is, ξ is the error value, $\frac{\partial \xi}{\partial t_k} = (t_k - r_{k-1})^2 p_U(t_k) - (t_k - r_k)^2 p_U(t_k) = 0$.

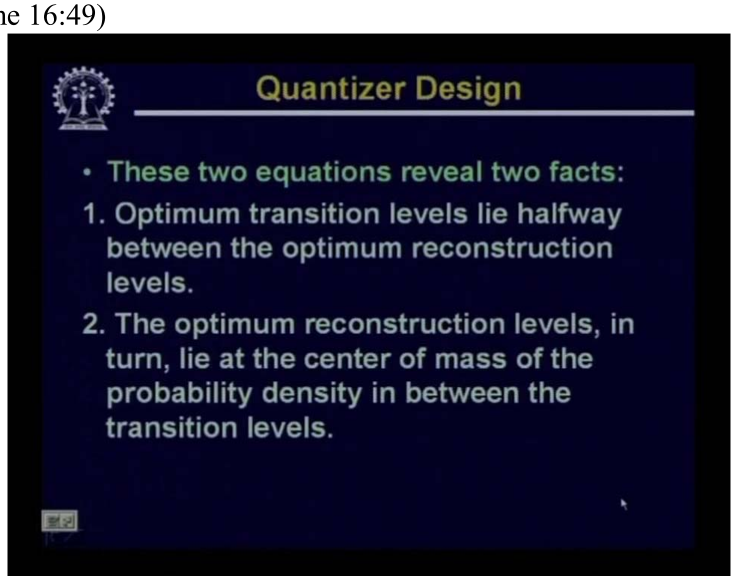
Similarly, the second equation, $\frac{\partial \xi}{\partial r_k} = 2\int_{t_k}^{t_{k+1}} (u-r_k)p_U(u)du = 0$, where the integration has to be taken from t_k to t_{k+1}. Now by solving these two equations

(Refer Slide Time 16:01)



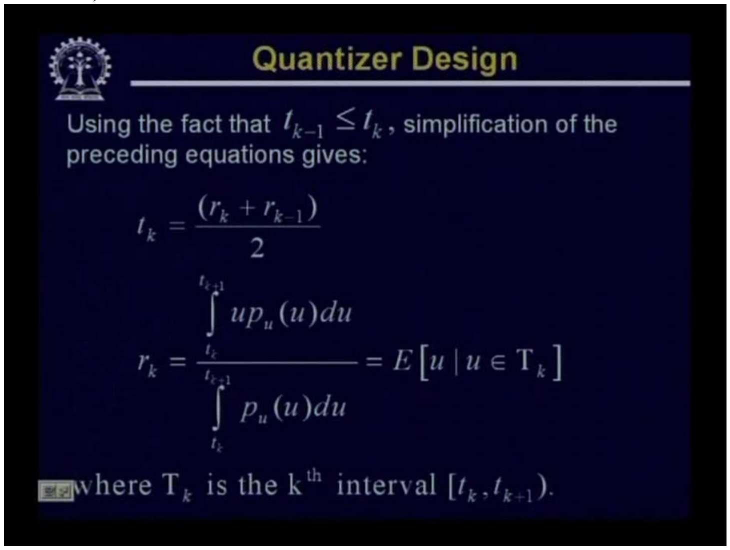
and using the fact that t_{k-1} is less than t_k we get two values, one is for transition level and other one is for the reconstruction level. So the transition level t_k is given by $t_k = \frac{(r_k + r_{k+1})}{2}$ and the reconstruction level r_k is given by $r_k = \frac{\int_{t_k}^{t_{k+1}} u p_U(u) du}{\int_{t_k}^{t_{k+1}} p_U(u) du}$. So what we get from these two equations? You find that these two equations tell

(Refer Slide Time 16:49)



that the optimum transition level t_k lie halfway between the optimum reconstruction levels So that is quite obvious

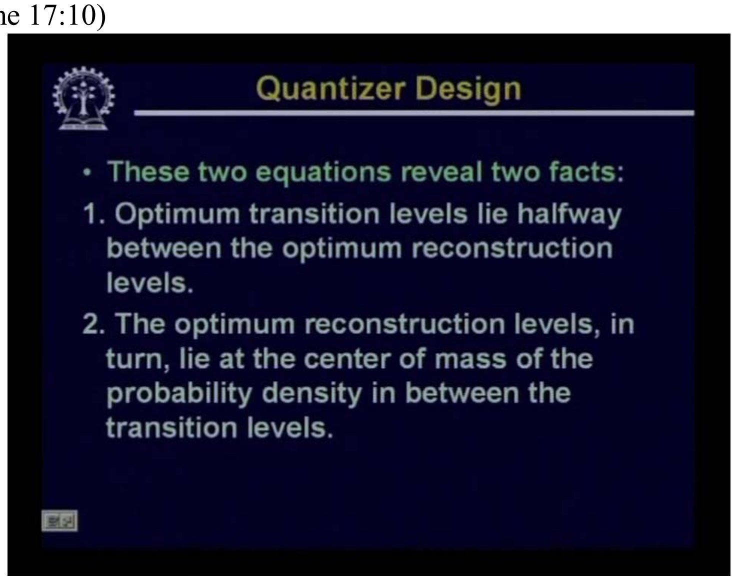
(Refer Slide Time 16:57)



because $t_k = \frac{r_k + r_{k-1}}{2}$. So this transition level lies halfway between r_k and r_{k-1}.

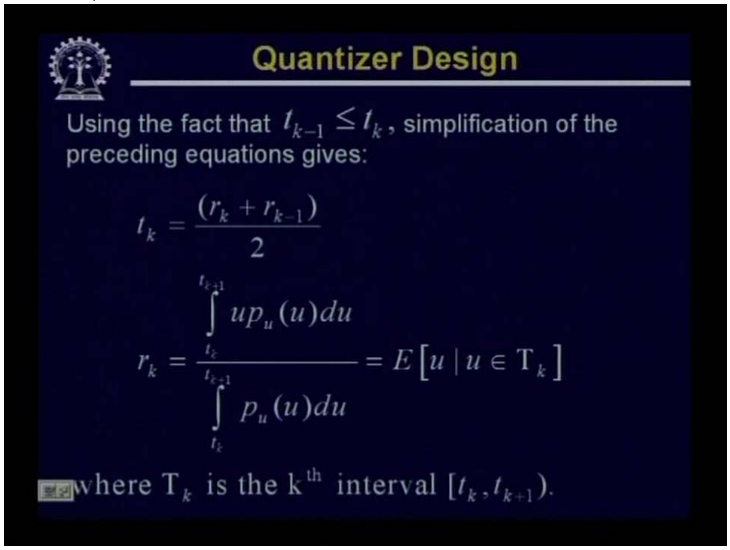
And the second observation is that,

(Refer Slide Time 17:10)



the optimum reconstruction levels in turn lie at the center of mass of the probability density in between the transition levels. So which is given by the second equation

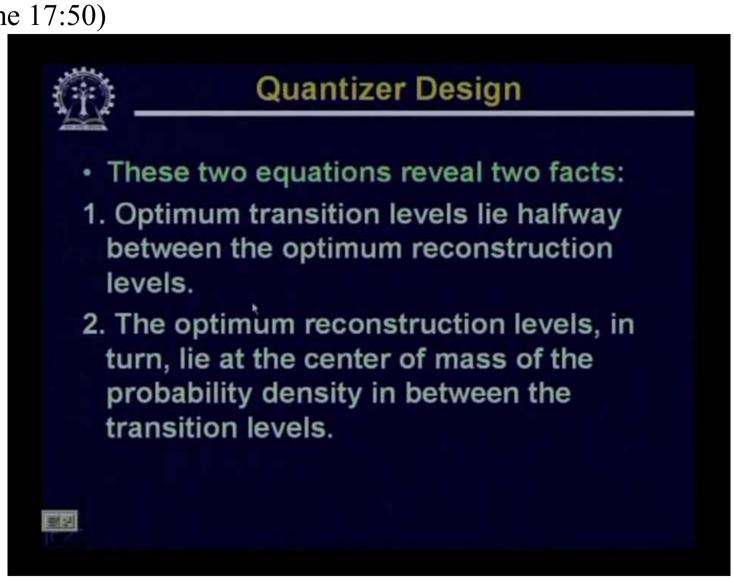
(Refer Slide Time 17:22)



that $r_k = \dfrac{\int_{t_k}^{t_{k+1}} u p_U(u) du}{\int_{t_k}^{t_{k+1}} p_U(u) du}$. So this is nothing but the center of mass of the probability density between the interval t_k and t_{k+1}.

So this optimum quantizer

(Refer Slide Time 17:50)



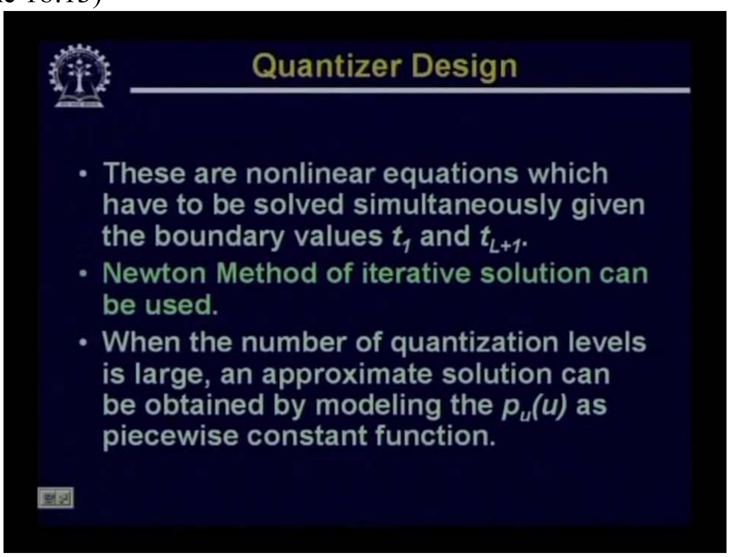
or the Lloyd-Max Quantizer gives you the reconstruction value, the optimum reconstruction value

(Refer Slide Time 17:56)

and the optimum transition levels in terms of probability density of the input signal.

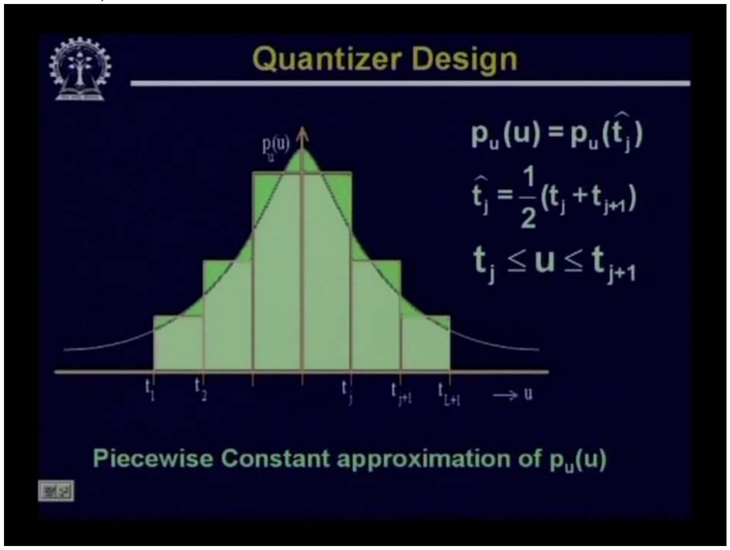
Now you find these two equations are non-linear equations

(Refer Slide Time 18:13)

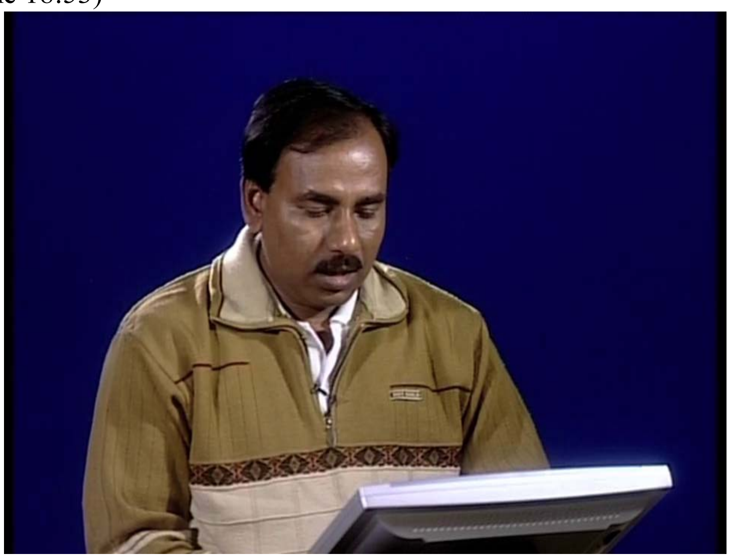


and we have to solve these non-linear equations simultaneously given the boundary values t1 and t_{L+1} and for solving this one can make use of the Newton method, Newton iterative method to find out the solutions. An approximate solution or an easier solution will be when the number of quantization levels is very large. So if the number of quantization levels is very large you can approximate $p_u(u)$, the probability density function as piecewise constant function. So how do you do this piecewise constant approximation?

(Refer Slide Time 18:52)

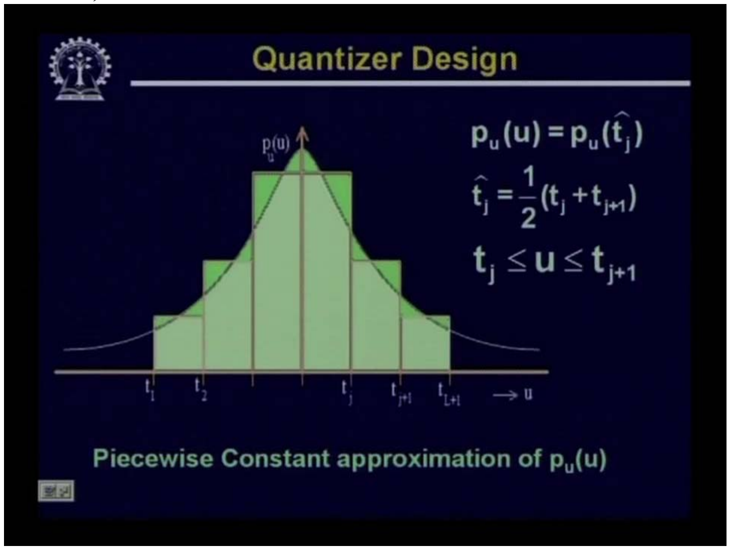


(Refer Slide Time 18:53)



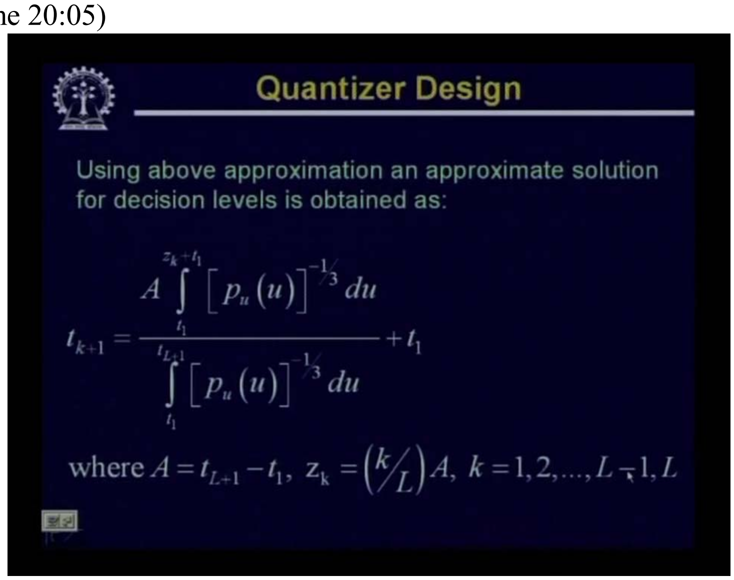
So in this figure you see that is a probability density function has been shown

(Refer Slide Time 18:58)



which is like a Gaussian function. So we can approximate it this way, that in between the labels t_j and t_{j+1}, we have the min value of this as $\hat{t}_j$, which is halfway between t_j and t_{j+1}, and within this interval we can approximate $p_U(u)$, where $p_U(u)$ is actually a non-linear one, we can approximate this as $p_U(\hat{t}_j)$. So in between t_j and t_{j+1}, that is in between every two transition levels, we approximate the probability density function to be a constant one which is same as the probability density function at the midway, halfway between these two transition levels. So if I do that, this continuous probability density function will be approximated by staircase functions like this. So if I use this approximation and recomputed those values, you will find

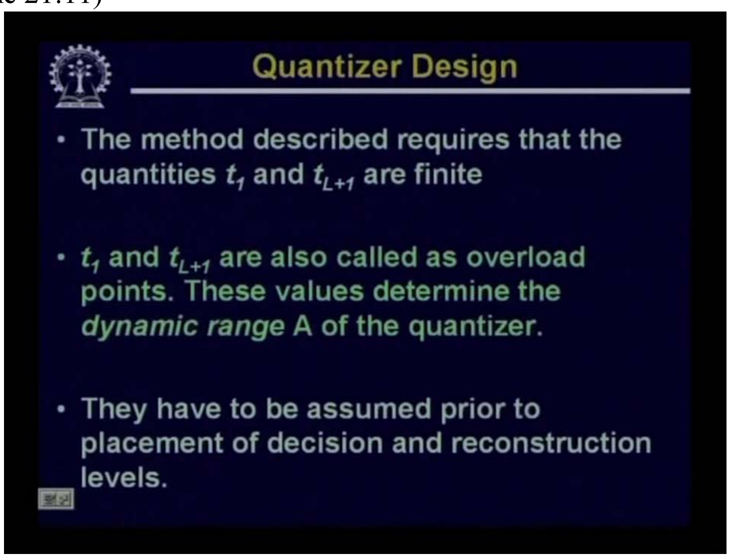
(Refer Slide Time 20:05)



that this t $t_{k+1} = \dfrac{A \int_{t_1}^{z_k + t_1} [p_U(u)]^{-1/3}\, du}{\int_{t_1}^{t_{L+1}} [p_U(u)]^{-1/3}\, du} + t_1$ where this A, the constant $A = t_{L+1} - t_1$ and we have said that t_{L+1} is the maximum transition level and t_1 is the minimum transition level and $z_k = \left(\frac{k}{L}\right) A$, where k varies from 1 to L.

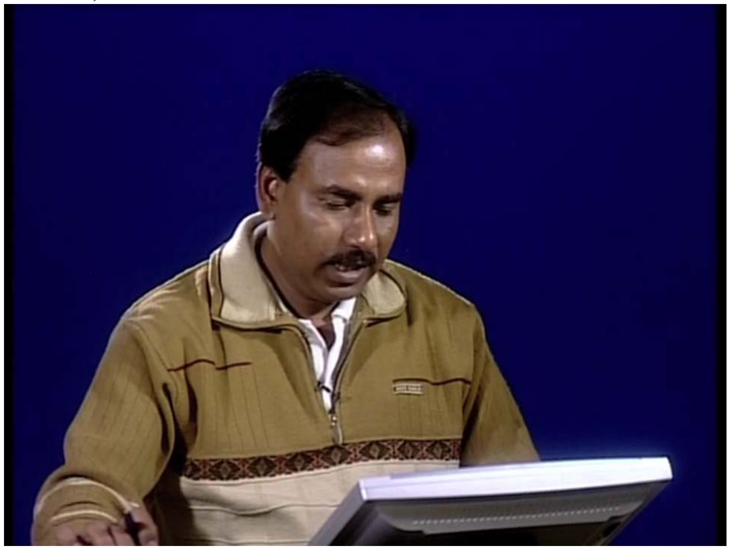
So we can find out t_{k+1} by using this particular formulation, when the continuous probability density function was approximated by piecewise constant probability density function.

(Refer Slide Time 21:11)



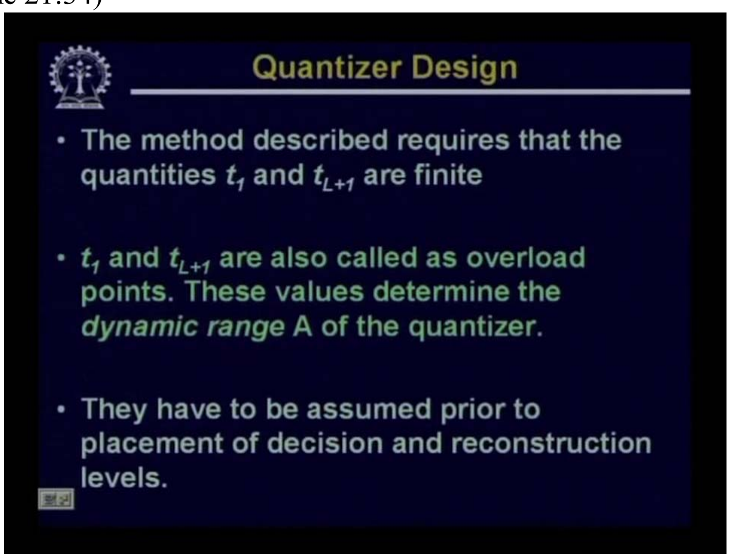
And once we do that, after that we can find out the values of the corresponding reconstruction levels. Now for solving

(Refer Slide Time 21:25)



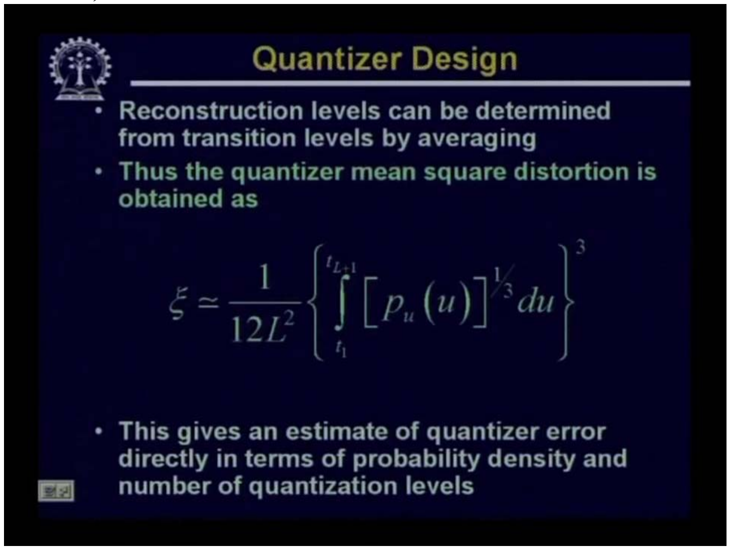
this particular equation, the requirement is that we have to have

(Refer Slide Time 21:34)



t_1 and t_{L+1} to be finite. That is the minimum transition level and the maximum transition level, they must be finite. At the same time, we have to assume t_1 and t_{L+1}, apriori before placement of decision and reconstruction levels. This t_1 and t_{L+1} are also called as overload points and these two values determine the dynamic range a of the quantizer.

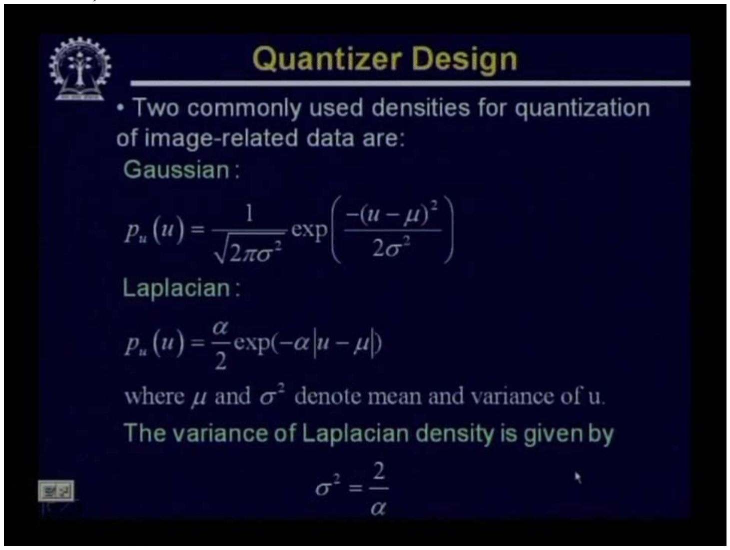
(Refer Slide Time 22:24)



So if you find that when we have a fixed t_1 and t_{L+1}, then any value less than t_1 or any value greater than t_{L+1}, they cannot be properly quantized by this quantizer; so this represents that what is the dynamic range of the quantizer

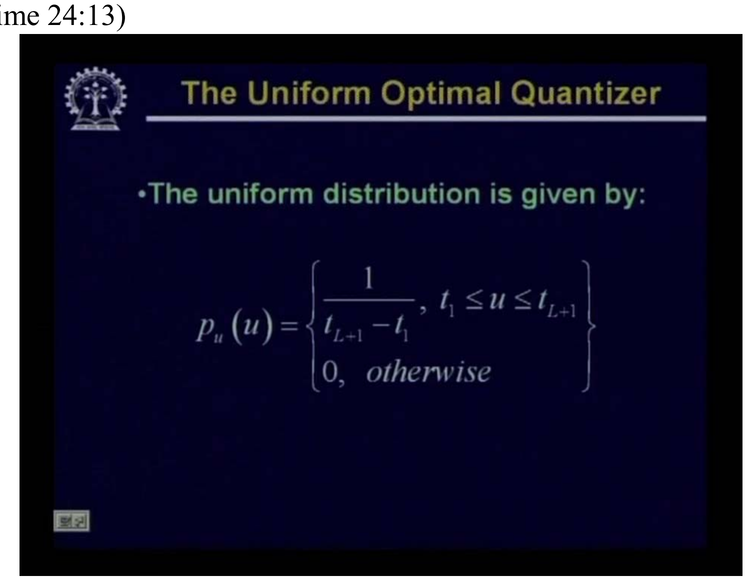
Now once we get the transition levels then we can find out the reconstruction levels by averaging the subsequent transition levels. So once I have the reconstruction levels and the transition levels, then the quantization mean square error can be computed as this, that is the mean square error of this designed quantizer will be $\xi \simeq \frac{1}{12L^2}\left\{\int_{t_1}^{t_{L+1}}\left[pU(u)\right]^{1/3} du\right\}^3$. And this expression gives an estimate of the quantizer error in terms of probability density and the number of quantization levels.

(Refer Slide Time 23:18)



Normally two types of probability density functions are used. One is Gaussian where the Gaussian probability density function is given by there is an well-known expression $p_U(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(u-\mu)^2}{2\sigma^2}\right)$ and the Laplacian probability density function which is given by $p_U(u) = \frac{\alpha}{2}\exp\left(-\alpha|u-\mu|\right)$, where μ and σ^2 denote the mean and variance of the input signal u, the variance in case of Laplacian density function is given by $\sigma^2 = \frac{2}{\alpha}$.

(Refer Slide Time 24:13)

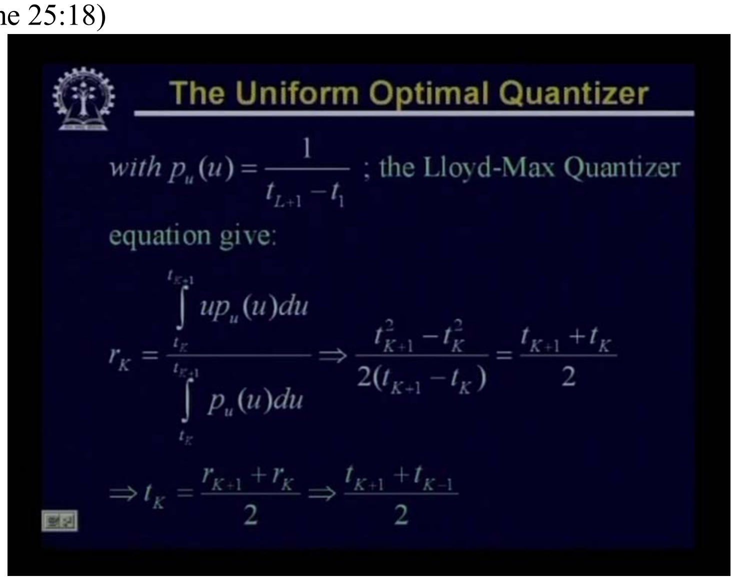


Now find that though the earlier quantizer was designed for any kind of probability density functions, but it is not always possible to find out the probability distribution function of a signal apriori. So what is in practice is you assume a uniform distribution, uniform probability distribution which is given $p_u(u) = \frac{1}{t_{L+1} - t_1}$, where u lies between t_1 and t_{L+1}. And $p_u(u) = 0$, when u is outside this region t_1 and t_{L+1}. So this is the uniform probability distribution of the input signal u. And by using this uniform probability distribution

(Refer Slide Time 25:09)

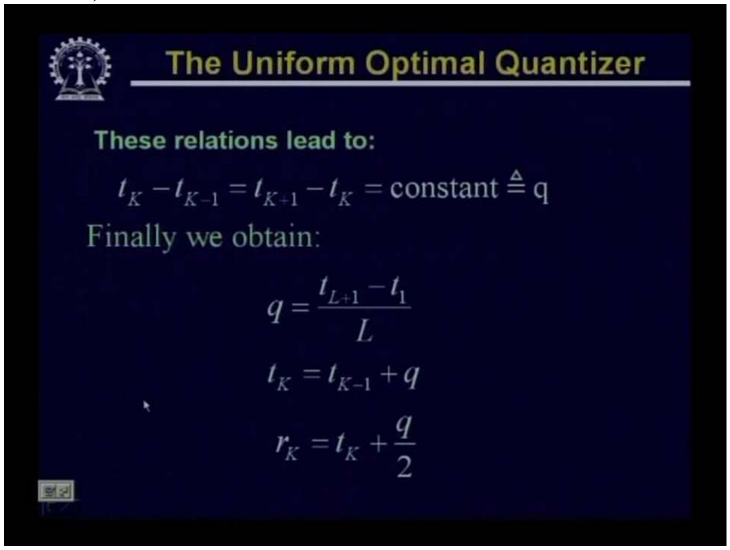
the same Lloyd-Max quantizer equations give r_k as,

(Refer Slide Time 25:18)



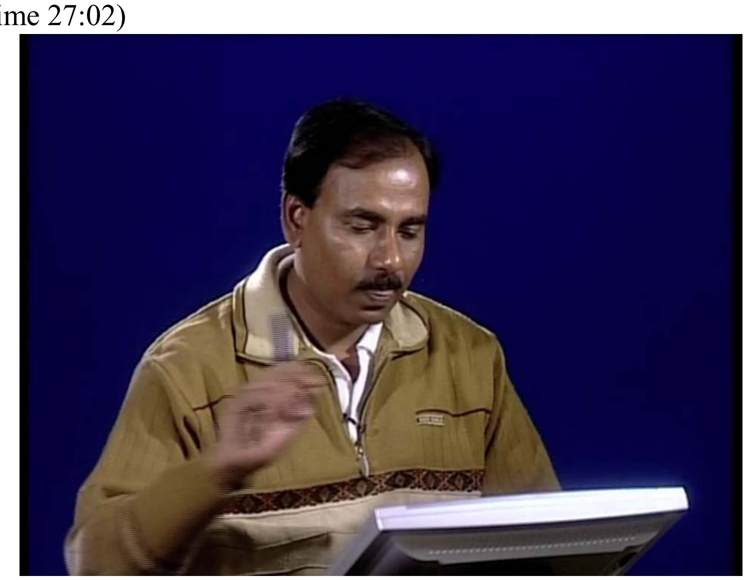
if I compute this, then you will find that the reconstruction level r_k will be nothing but $r_k = \frac{t_{k+1} - t_k}{2}$, where t_k will be $t_k = \frac{r_{k-1} + r_k}{2}$, which is same as $\frac{t_{k+1} - t_{k-1}}{2}$. So I get the reconstruction levels and the decision levels for a uniform quantizer.

(Refer Slide Time 25:50)



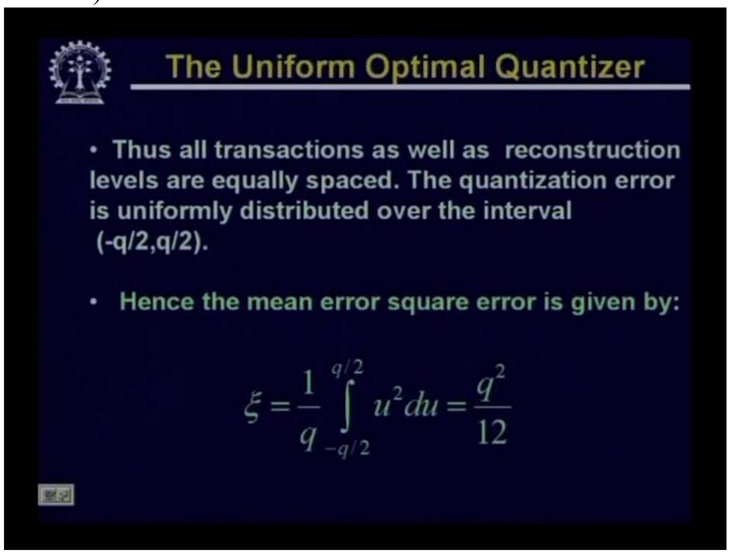
Now these relations lead $t_k - t_{k-1}$ is same as $t_{k+1} - t_k$ and that is constant equal to q, which is known as the quantization step. So finally, what we get is the quantization step, is given by $q = \frac{t_{L+1} - t_1}{L}$, where $t_{L+1}$is the maximum transition level and t_1 is the minimum transition level and L is the number of quantization steps. We also get the transition level t_k in terms of transition level t_{k-1} and q as $t_k = t_{k-1} + q$ and the reconstruction level r_k in terms of the transition level t_k as $r_k = t_k + \frac{q}{2}$. So we obtain all the related terms of a uniform quantizer using this mean square error quantizer design which is the Lloyd Max quantizer for a uniform distribution.

(Refer Slide Time 27:02)



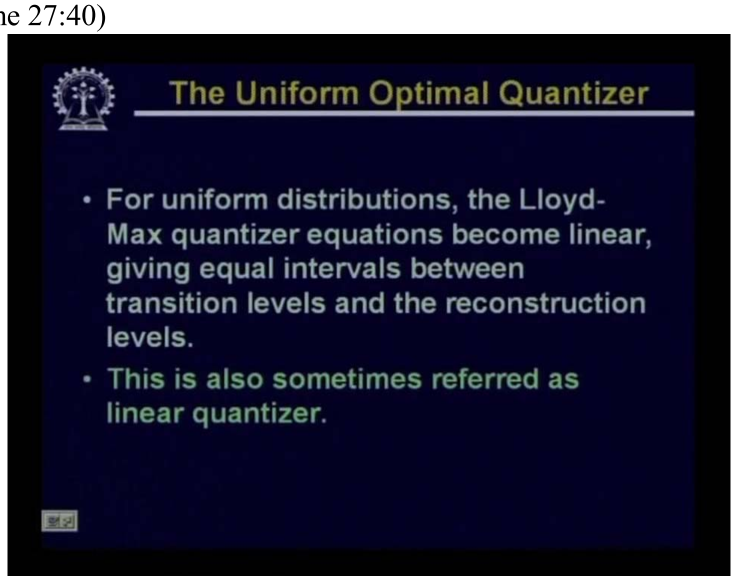
So here you find that all the transition levels as well as the reconstruction levels are equally spaced and the

(Refer Slide Time 27:11)



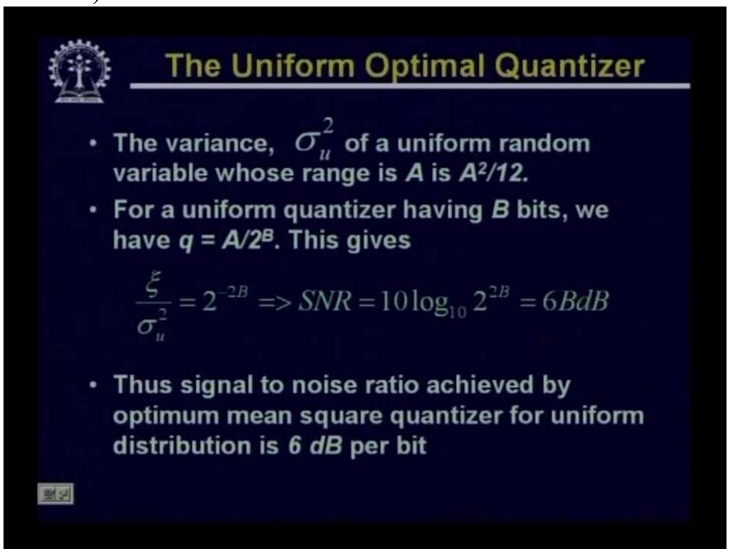
quantization error in this case is uniformly distributed over the interval -q /2 to q/2. And the mean square error in this particular case if you compute will be given by $\xi = \frac{1}{q}\int_{-q/2}^{q/2} u^2 du$, which will be nothing but $\frac{q^2}{12}$.

(Refer Slide Time 27:40)



So for uniform distribution the Lloyd Max quantizer equation becomes linear because all the equations that we had derived earlier, they are all linear equations giving equal intervals between transition levels and the reconstruction levels and so this is also sometimes referred as a linear quantizer.

(Refer Slide Time 28:04)



Ok. So there are some more observations from this linear quantizer. The variance, σ_u^2 of a uniform random variable. whose range is A, is given by $A^2/12$. So for this, you find that for a uniform quantizer with B bits. So if we have a uniform quantizer, where every level has to be represented by B bits we will have $q = \frac{A}{2^B}$ because the number of steps will be, 2^B number of steps and thus the quantization step will be $q = \frac{A}{2^B}$ and from this you find that the $\frac{\xi}{\sigma_u^2} = 2^{-2B}$ and from this we can compute the signal to noise ratio. In case of a uniform quantizer the signal to noise ratio is given by $SNR = 10\log_{10} 2^{2B}$ 10, this is nothing but 6B dB.

So this says that signal to noise ratio that can be achieved by an optimum mean square quantizer for uniform distribution is 6 dB per bit that means if I increase the number of bits by 1. So if you increase the number of bits by 1, that means the number of quantization levels will be increased by 2, by a factor of 2. In that case you gain a 6 dB in the signal to noise ratio in the reconstructed signal.

(Refer Slide Time 29:59)

So with this we come to an end on our discussion on the image digitization process. So here we have seen that how to sample an image or how to sample a signal in one-dimension, how to sample an image in two-dimension. We have also seen that after you get the sample values, where each of the sample values are analog in nature, how to quantize those sample values so that you can get the exact digital signal as well as exact digital image. Thank you