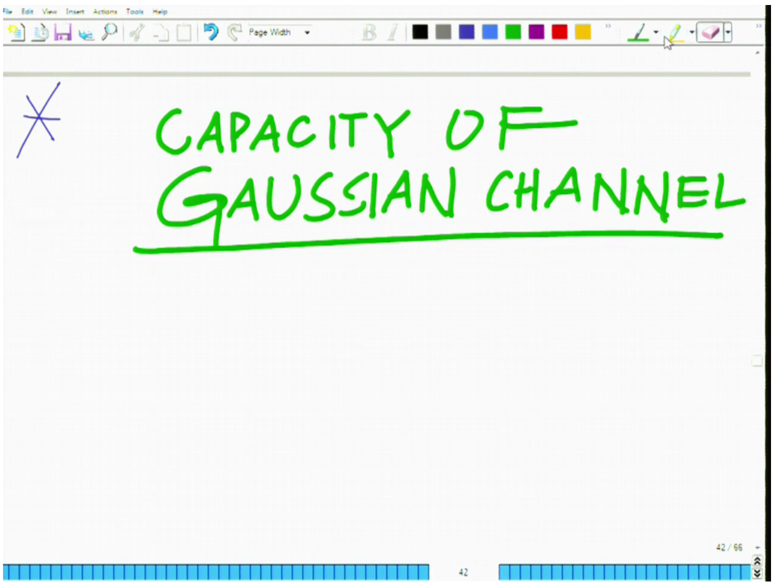
Principles of Communication Systems - Part II
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 40
Capacity of the Gaussian Channel – Part II, Practical Implications, Maximum Rate in Bits/Sec

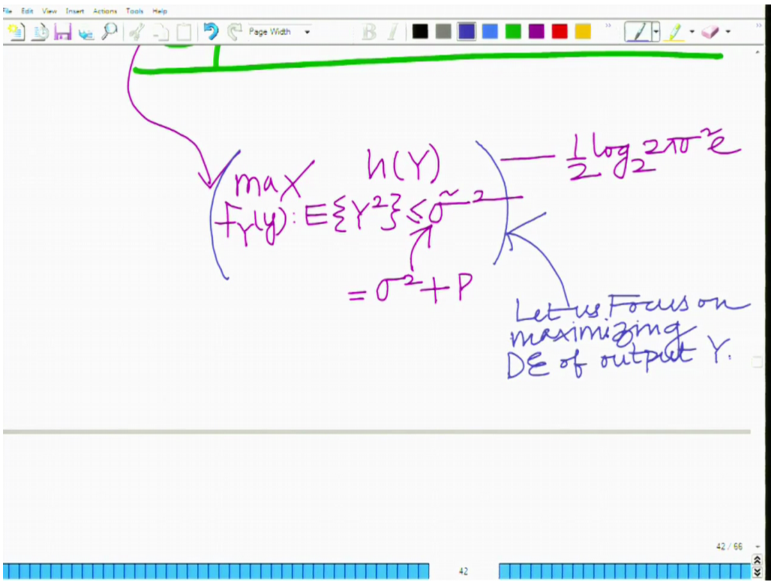
Hello, welcome to another module in this massive open online course. So, we are looking at the capacity of the Gaussian channel correct, we are midway through the derivation of the capacity of the Gaussian channel of communication channel which is basically characterized by the addition of white Gaussian noise to the input, alright.

(Refer Slide Time: 00:35)



And we have shown that the maximum. So, we are looking at the capacity of the Gaussian channel correct.

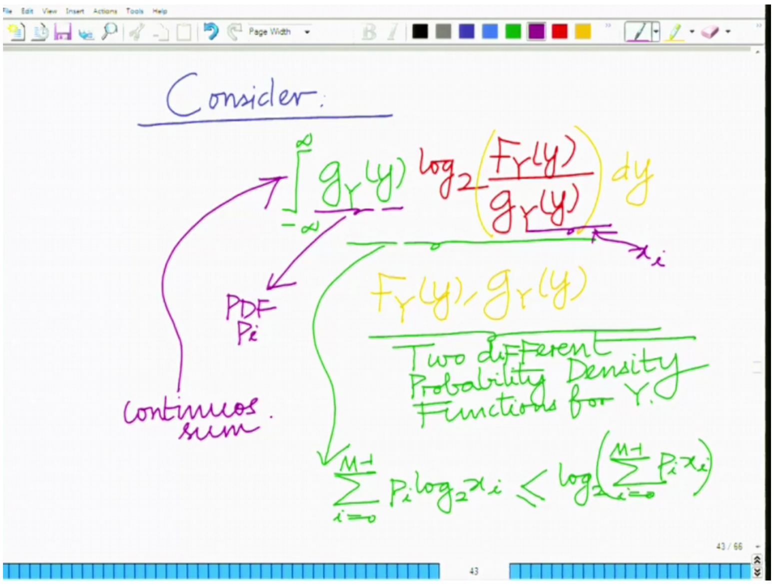
(Refer Slide Time: 01:10)



And well what we have seen is that the capacity of the Gaussian channel when you maximize the mutual information it reduces to the following thing that is we have to reduces in turn to maximize over the set of all probability density functions. Well if you look at what we had written over the set of all probability distribution functions of the output twice such that the power is limited this is less than or equal to sigma tilde square by the way this sigma tilde square this is equal to noise power sigma square plus P I have to maximize h Y minus of course, we have a constant which is of half log the base 2; 2 pi sigma square.

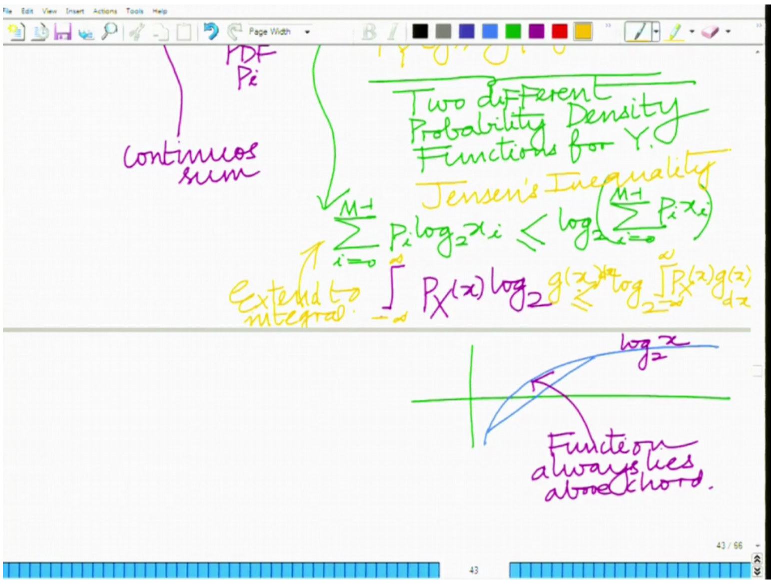
So, first let us focus on maximizing this part focus on maximize that is focus on maximizing the differential entropy of the output wise. So, let us focus, let us focus on maximizing the differential entropy of output Y, now this is slightly laborious correct the procedure for this is slightly lengthy. So, let us go through a procedure to demonstrate. So, we are eventually going to show that corresponding to the variance sigma tilde square the Gaussian probability density function has the maximum differential entropy. However, the procedure corresponding to this is slightly involved and we are going to go through this in a step by step fashion the first thing you have to do is I want to introduce a concept known as the Kullback Leibler divergence and that can be defined as the following. So, we will go through that first let us start by considering let us start by considering.

(Refer Slide Time: 03:29)



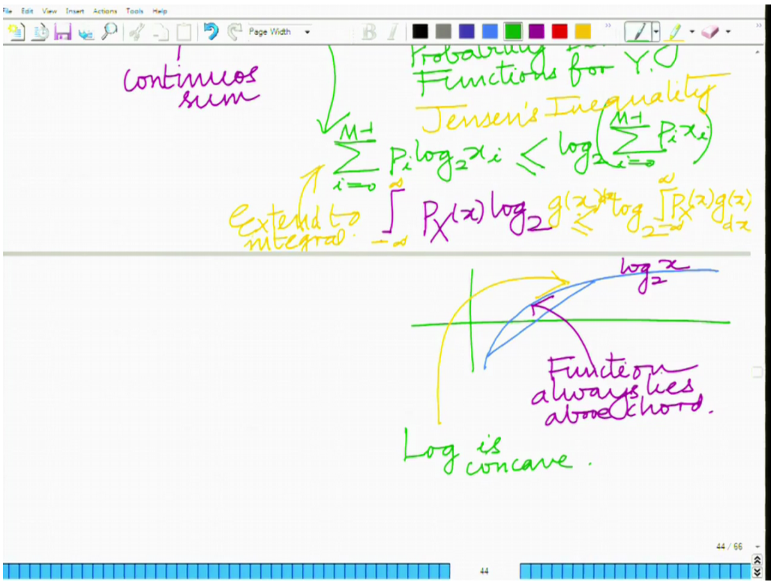
So, consider the following quantity which might sound slightly arbitrary, but its well motivated its defined as g Y of y log to the base 2 F of Y of y divided by g of Y t y where these 2 quantities F Y of y and g Y of y these are 2 different pdfs; 2 different pdfs or probability 2 different pdfs or probability density functions for the random variable Y and now what you are looking at is basically.

(Refer Slide Time: 04:56)



Now, if you look at this we want to use the convexity of the log remember the convexity of the log says that summation from convexity of the log of we have summation P i; I equal to 0 to m minus 1 P i log to the base 2 X i this is less than or equal to log to the base 2 summation i equals to 0 to m minus 1 P i x.

(Refer Slide Time: 05:28)

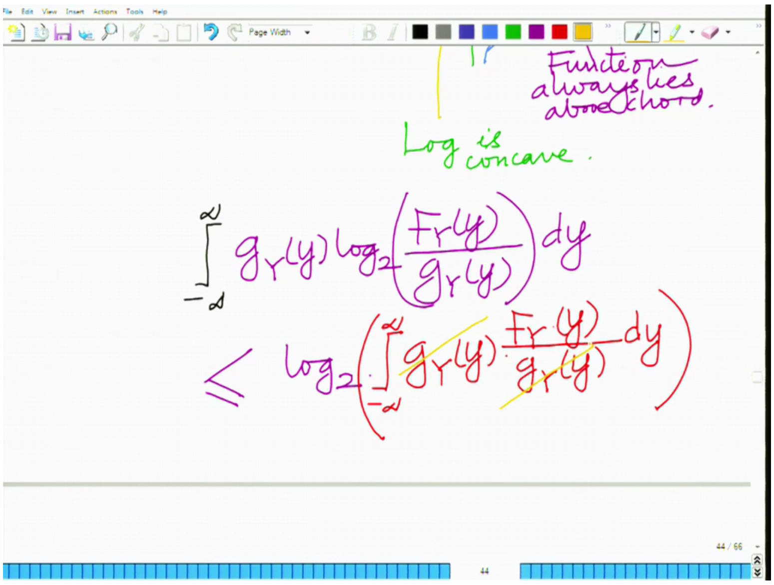


That is if you look at this we said that the function always lies about the cord or the log function that is function always lies about the; this is log 2 to the base X correct. So, you can see function always lies above the function always lies above the cord that is the first aspect.

Now, the second point here if you look at it this is simply. So, look at this; this is a simply probability density function this is a Pdf you can think of this as P i you can think of this quantity as X i and this quantity integral is nothing, but a continuous sum. So, one can also write in this case or instance one can also write instead of a continuous sum i can also consider probability density function P X of x log 2 to the base correct log 2 to the base log to the base 2 X integral correct or log to the base log to the base 2 you can also look at not necessarily the X you can also look at any function of x. So, you can look at for instance any function g of X and that would be less than or equal to log to the base 2 integral minus infinity to infinity P X of x; g X d x of course, has to be d x here also.

So, basically what you have is you can extend this to a continuous integral which is a continuous sum this is your Jensen’s inequality remember inequality for Jensen’s inequality for the concave functions which basically says that the function always lies above the cord, alright. So, what you are saying is that summation the discrete sum can natural be extended to integral alright by replacing the sum by an integral and probabilities is the discrete probability or the probability mass function corresponding to the X is to a probability density function over X. So, that the whole point and this is the Jensen's inequality for a concave function and log of course, is a concave function correct that makes inequality applicable for the log function.

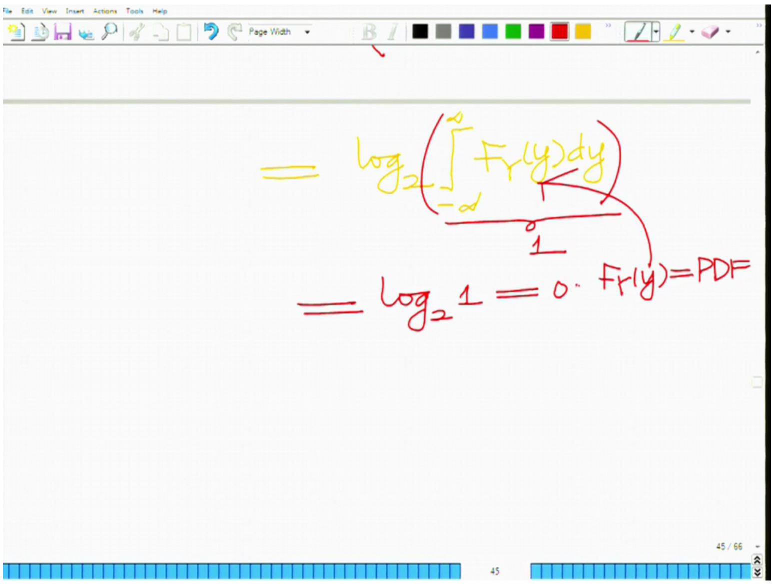
(Refer Slide Time: 08:59)



And therefore, now if you look at this quantity that we have integral minus infinity to infinity we have g Y of y log to the base 2 F of Y divided by g Y d y.

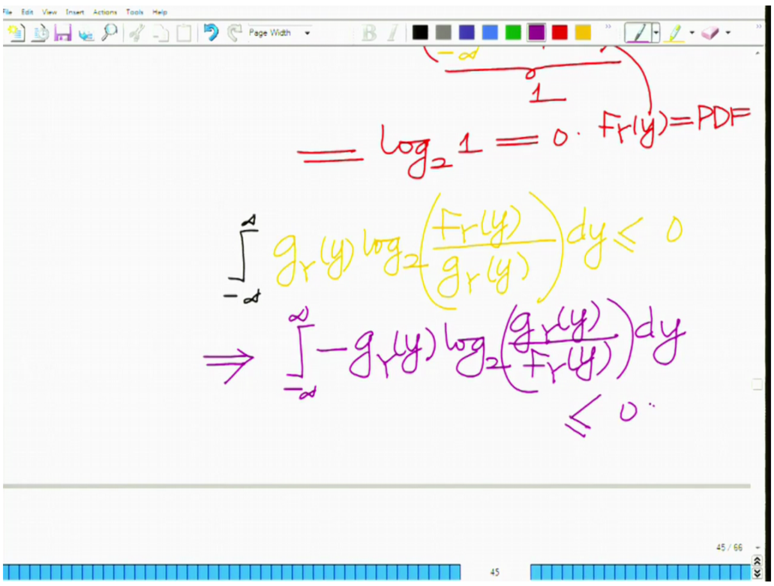
Now, using Jensen's inequality this is less than or equal to log to the base 2 you have to bring integral the summation inside the log g Y of y into the argument if a Y y divided by g Y of y times d Y and now you observe that the g Y cancel.

(Refer Slide Time: 09:57)



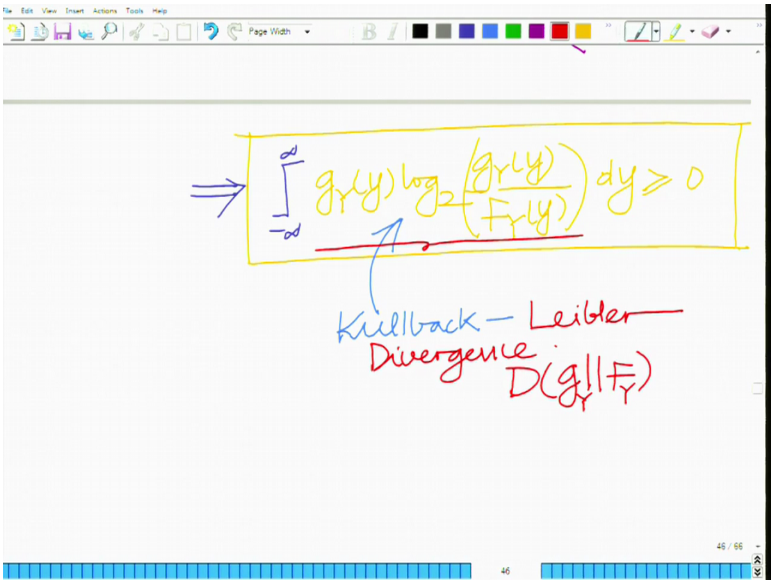
And therefore, what we have is this is equal to nothing, but log to the base 2 what remains is integral F Y of y, d y, but this quantity is area under the probability density function because F Y; F Y is any probability density function. This is area under the probability density function this quantity integral minus infinity to infinity F Y d y is the area under the probability density function therefore, this is equal to what the total area under the probability density function integral minus infinity to infinity F Y; d y and therefore, this is equal to log 2 to the base 1 which is equal to 0.

(Refer Slide Time: 10:52)



And naturally therefore, we have the very important and interesting inequality that integral g Y of y log to the base 2 F Y of y u Y of y d y less than or equal to 0 implies log F Y of Y divided by g Y of Y is nothing, but minus of log g Y of y divided by F Y y. So, this is equal to integral minus infinity to infinity minus g Y F y log to the base 2 d Y and this is less than or is equal to 0.

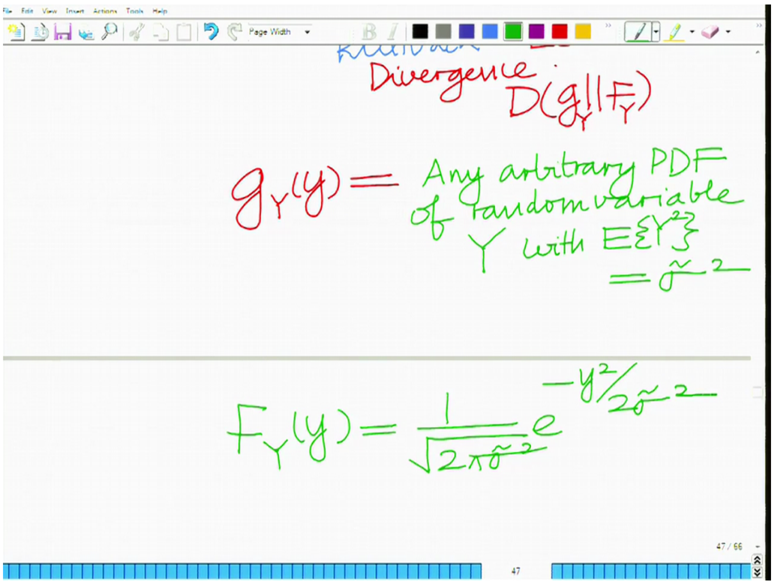
(Refer Slide Time: 11:56)



And finally, you bring the negative side to that side that implies minus infinity to infinity g of Y log to the base 2 g Y of y correct divided by F Y divided by F Y of y divided is greater than or equal to 0 and this quantity integral minus in g Y of Y log to the base 2 g Y y divided this is known as the Kullback this is known as the Kullback Leibler the Kullback Leibler divergence denoted by d.

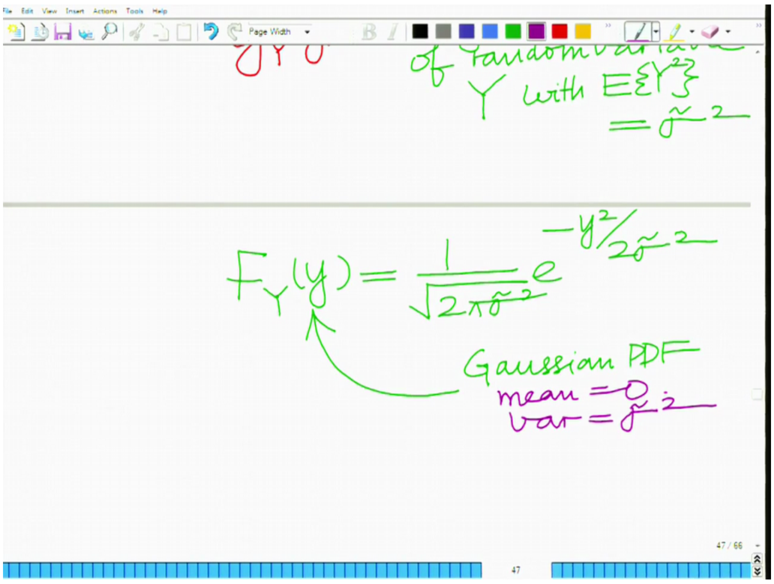
So, this is a Kullback Leibler divergence and this is always greater than or equal to 0. So, we have shown. So, we have defined this quantity this is known as the Kullback Leibler divergence and we have shown that this quantity is always greater than equal to 0 and we are going to use this property further in deriving the probability density function that maximizes the differential entropy remember. So, that is our original aim alright we have not yet to do that which we are going to shortly which is to basically derive the differential entropy first find the probability density function Y which maximizes the differential entropy corresponding to the constraint that the power is less than equal to sigma tilde square right which we have defined as P plus sigma square.

(Refer Slide Time: 14:15)



Now, let us now set this Kullback Leibler divergence in this Kullback Leibler divergence let us set g Y of Y as any arbitrary Pdf; Pdf of random variable Y with expected Y square equal to sigma tilde square. So, this is any arbitrary Pdf and F Y of y we will choose this as the Gaussian Pdf with 0 mean that is 1 over square root of 2 pi sigma tilde square e raise to minus e raise to minus Y square divided by 2 sigma tilde square.

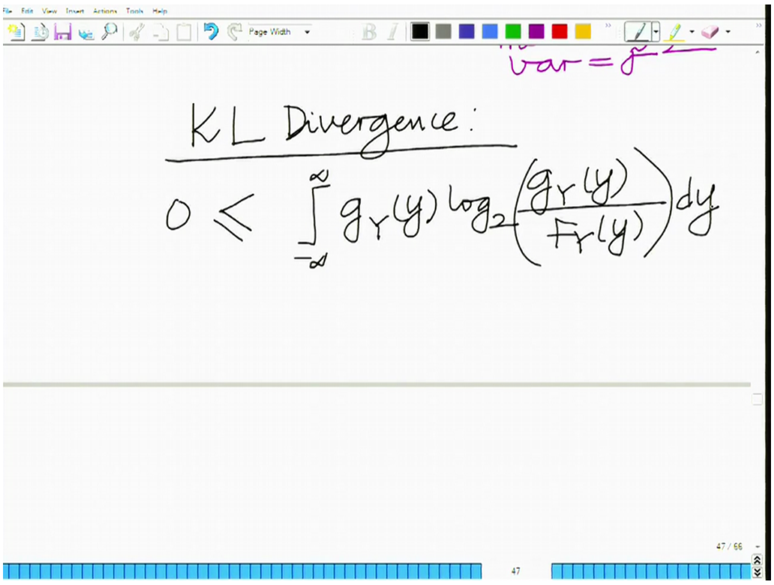
(Refer Slide Time: 15:18)



So, we set F Y of y as the Gaussian Pdf mean 0 this is your Gaussian Pdf mean equals 0 variance is equal to sigma tilde square. So, we are choosing 2 Pdf. So, g Y y let be any arbitrary Pdf like except that we remember it has to satisfy the constraint that the power that is expected Y square is equal to sigma tilde square and F Y we are setting it in particular as the Gaussian Pdf with mean 0 variable sigma tilde square.

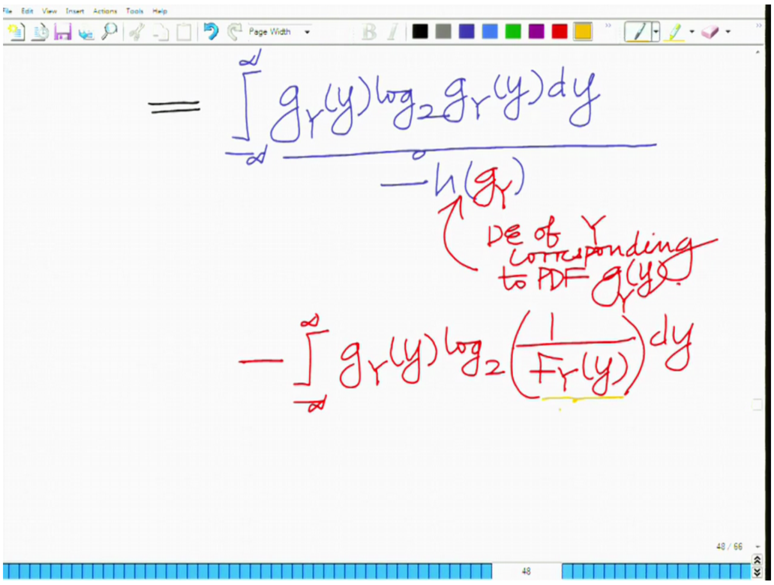
Now, let us look at the Kullback Leibler divergence between these 2 now if you look at the Kullback Leibler divergence between these 2.

(Refer Slide Time: 16:03)



Or the which is also known as KL which is also known as the simply abbreviated as KL divergence also abbreviated as simply the KL divergence what you see is we have 0 remember the Kullback we have put that the KL divergence is always less than or equal to KL divergence is always greater than or equal to 0. So, I have minus infinity to infinity g Y of y log to the base 2 g Y of y divided by F Y of Y d y.

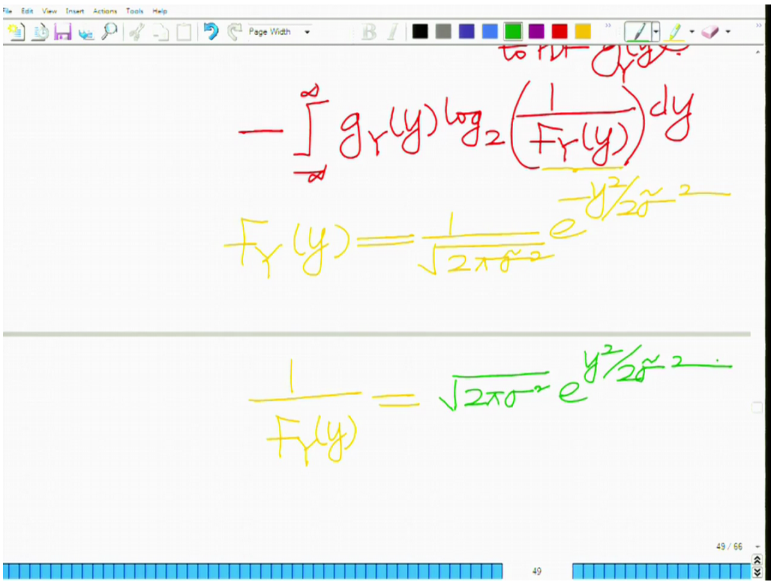
(Refer Slide Time: 16:55)



Now, if you simplify this; this is equal to nothing, but integral minus infinity to infinity g Y of Y log to the base 2 g Y of Y d y that is log to the base 2 g Y of y over F Y can be written as log to the base to g Y minus log to the base to F Y, correct.

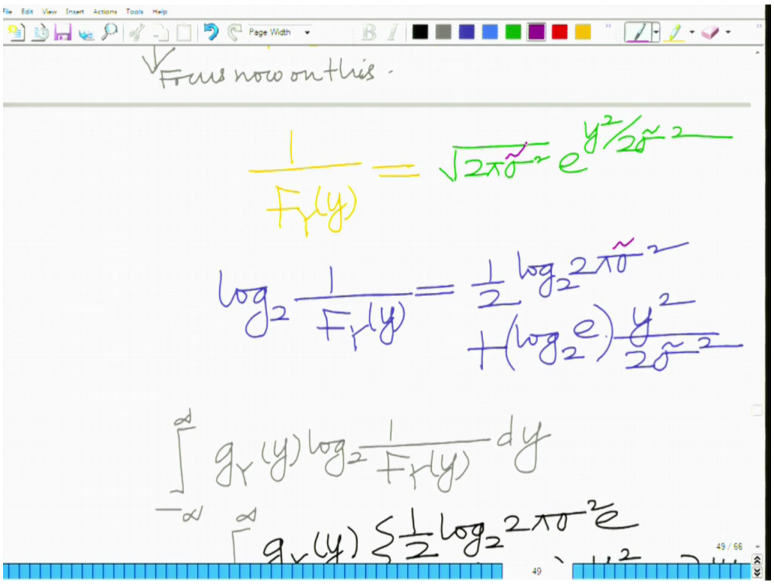
Now, if you look at the first term the first term is nothing, but minus the differential entropy of Y corresponding to the probability density function g. So, I am going to simply write it as minus h of g y that is basically differential entropy of Y corresponding to the Pdf g Y g y minus integral minus infinity to infinity well g Y of y log to the base 2 1 over F of Y y d y. Now remember F of Y y this is the Gaussian Pdf. So, we have let us simplify this quantity although we have already done it before when we are calculating the differential entropy of the Gaussian random variable, but of Gaussian sources.

(Refer Slide Time: 18:39)



But let us do it once again 1 over square root of 2 pi sigma tilde square e raise to minus Y square by 2 Sigma tilde square 1 over F Y of Y this is equal to well naturally this will be square root 2 pi sigma tilde square P power Y square by 2 sigma tilde square.

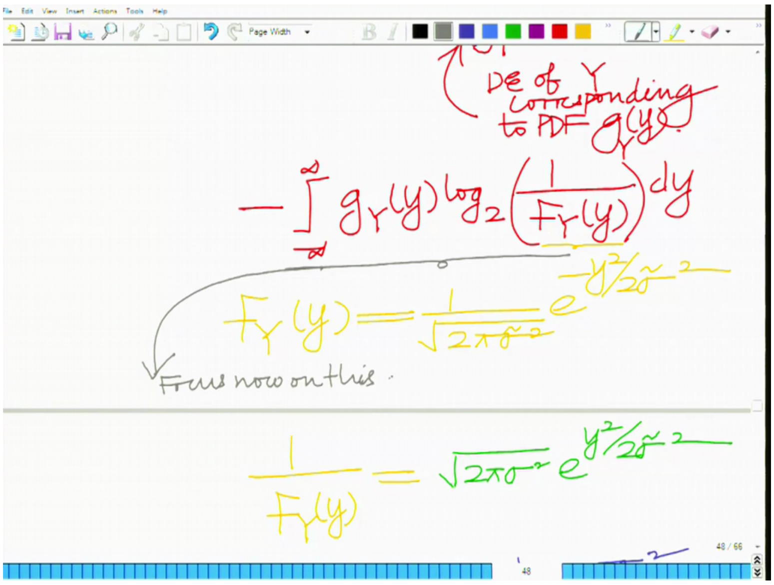
(Refer Slide Time: 19:18)



And log to the base 2 1 over F Y of y is basically half log to the base 2; 2 pi sigma square plus log e to the base 2 into log of e raise to Y square by 2 sigma tilde square to the base e. So, that will simply be Y square divided by 2 sigma tilde square.

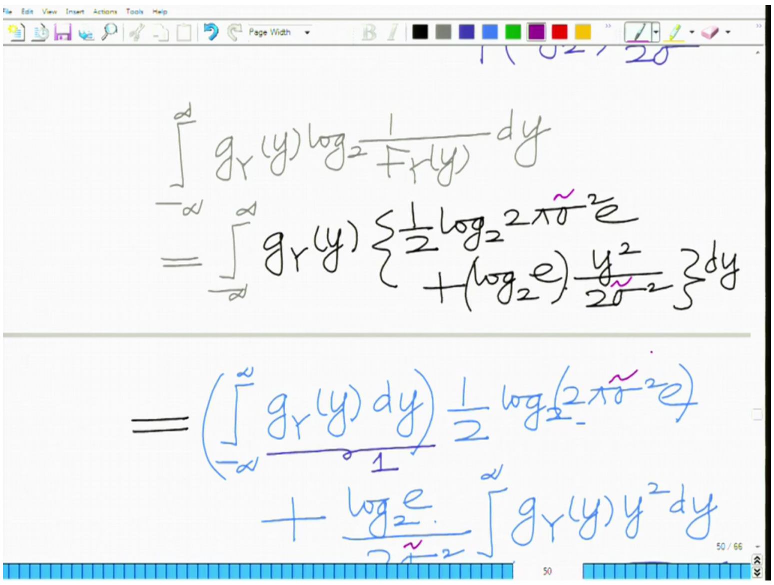
So, we have simplified this quantity log to the base 2 1 over F Y of y. So, I have simplified this quantity. Now I am going to substitute it in that integral and evaluate that now if you look at the previous integral that will be when I am saying previous integral I am saying this.

(Refer Slide Time: 20:09)



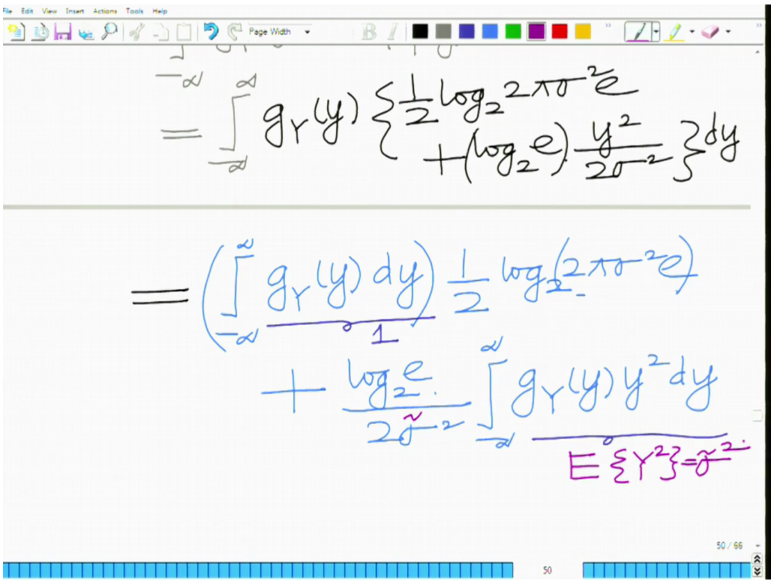
So, we are focusing on this now the first integral is there focus now on this integral.

(Refer Slide Time: 20:21)



And this integral is integral minus infinity to infinity g Y of y log to the base 2 1 over F Y of y d y is integral minus infinity to infinity well g Y of y substitute log to the base 2 1 over F Y of Y that is half apologize for this. This is half log to the base 2 2 pi sigma square e plus log e to the base 2 into Y square divided by 2 sigma square d Y which is equal to if you look at this; this will be integral minus infinity to infinity g Y of y.

(Refer Slide Time: 21:24)

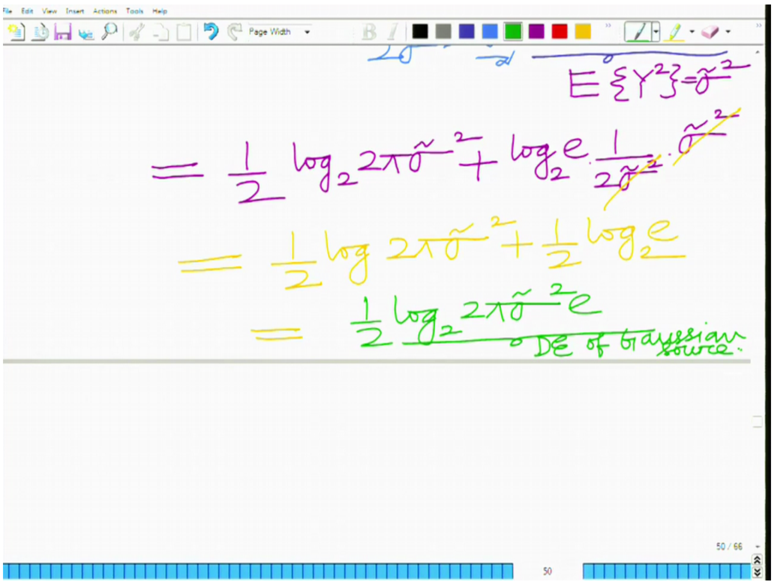


Well this half log 2 to the base of log to the base 2 2 pi sigma square is a constant. So, this will come outside half log to the base 2; 2 pi sigma square e plus now again log e to the base 2 is a constant divided by 2 sigma square is a constant this will be minus infinity to infinity g Y of y y square d y.

Now, if you look at this integral infinity to minus infinity g Y of y d y this is nothing, but one this is the total area under probability density function by the property of the total probability correct that has to be one the integral that is the area under the probability density function is one. Now if you look at this quantity this is interesting integral minus infinity to infinity g Y into y square d y this is nothing, but expected I am sorry this has to be by sigma tilde square expected Y square this is equal to sigma tilde square its everywhere I am writing sigma square this has to be sigma tilde square.

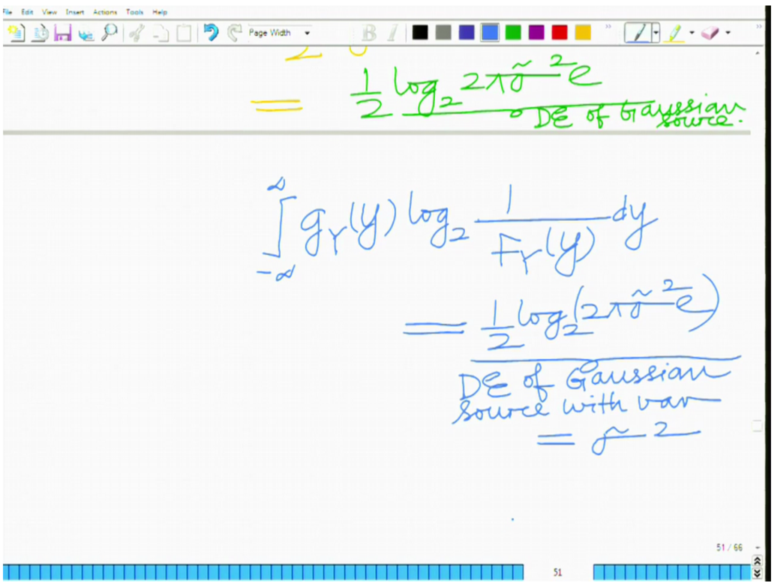
So, this has to be this is also sigma tilde square yeah. So, this is also sigma tilde square sigma tilde square sigma tilde square.

(Refer Slide Time: 23:16)



And this has to be and therefore, what we have is half log to the base 2; 2 pi sigma tilde square now e 2 pi sigma tilde square 2 pi sigma tilde square half log to the base to sigma tilde square plus log e to the base 2 into 1 by 2 sigma tilde square into sigma tilde square. And now the sigma tilde square cancel and therefore, what you have is half log to the base 2; 2 pi sigma tilde square plus half log e to the base 2 which is equal to nothing, but this is basically your half log to the base 2; 2 pi sigma tilde square this is nothing, but the differential entropy of the Gaussian source.

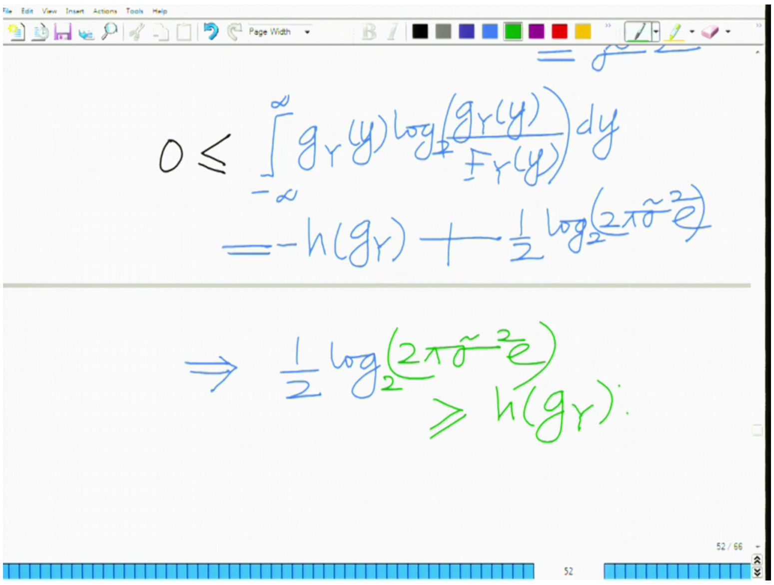
(Refer Slide Time: 24:39)



So, what we have shown is basically if you look at this what we have shown is basically a minus infinity to infinity g Y of y log to the base 2 1 over F Y of y d y this is equal to half log to the base 2 2 pi sigma tilde square e which is nothing, but d e of differential entropy of Gaussian source with variance equal to sigma tilde square.

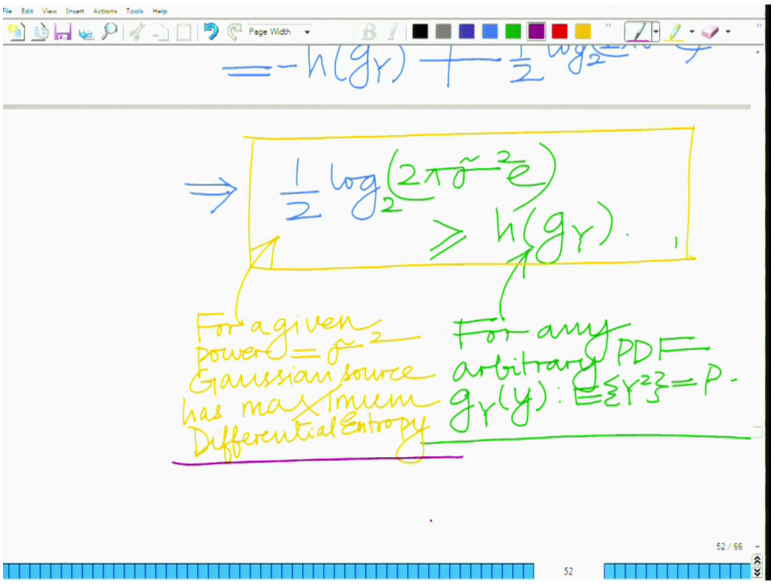
So, we have shown that this quantity is indeed the differential entropy of a Gaussian source with the variance equal to sigma tilde square and now finally, taking this all back substituting it this integral.

(Refer Slide Time: 25:46)



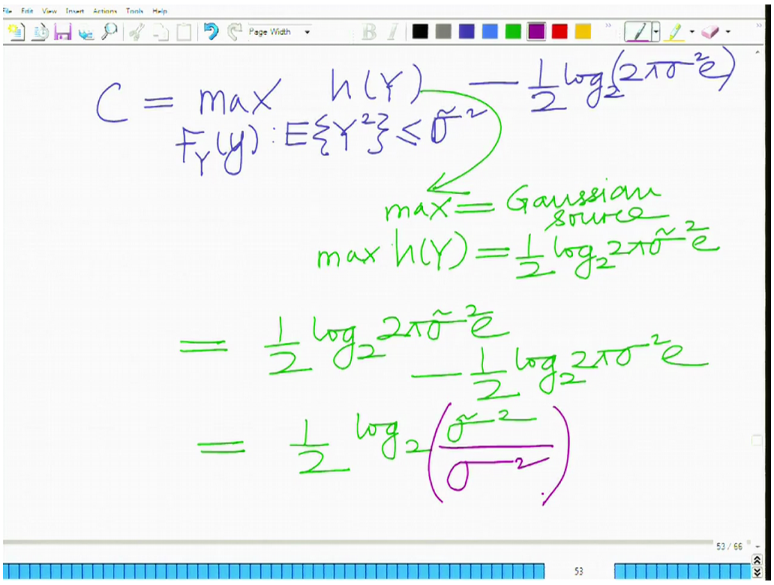
We have evaluated what we have is if you do a recap we have 0 less than or equal to if you just are willing to spend some time just reflect or go back 0 is less than or equal to that Kullback Leibler divergence log 2 to the base well F Y d y, this is equal to well h of g Y that is the differential entropy of Y corresponding to the probability density function g Y minus half log to the base 2 2 pi sigma tilde square e. Which means if I am sorry this is minus and this is plus which means if you look at this what we get is half log to the base 2 2 pi sigma tilde square e is greater than or equal to h of g of Y for any arbitrary probability density function for any arbitrary probability density function g Y of Y such that expected Y square is equal to P that is interesting.

(Refer Slide Time: 27:08)



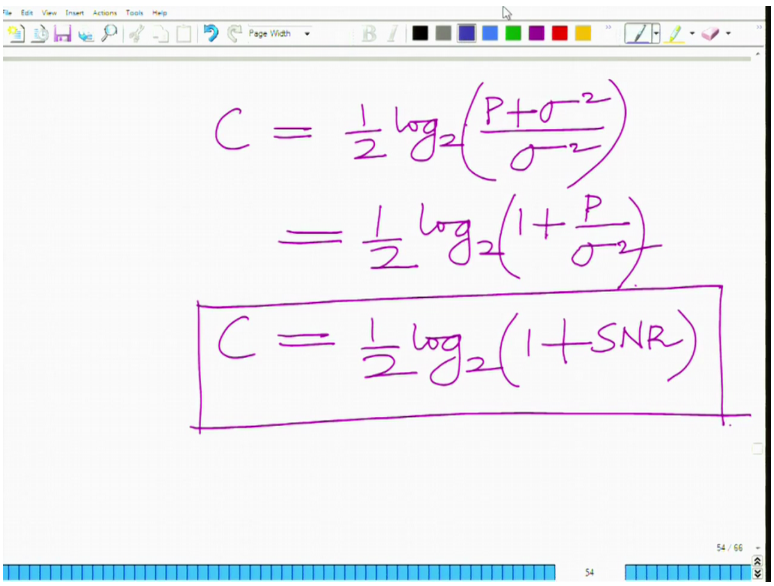
So, what we have shown is that for a given variance for a what if you reflect on this and we have shown a very important result if you reflect on this what you will realize is what we have shown is that for a given variance that is consider a random variable with a given variance and mean 0 the Gaussian probability density function maximizes the different differential entropy that is for a given power amongst all the sources Gaussian source has the maximum Gaussian source has a maximum differential entropy. And now, therefore, if you go back to original problem that is maximizing which is remember this is all aside to derive that probability density function which maximizes the differential entropy now if you go back.

(Refer Slide Time: 28:58)



All the way remember we have this problem of maximizing the capacity is maximizing the mutual information subject to e Y square less than or equal to well sigma tilde square h Y minus half log to the base 2 2 pi sigma square e we now know what maximizes. This we now know max equal to max occurs for the Gaussian source and max h of Y is basically half log two to the base 2 2 pi sigma tilde square e and therefore, the capacity will be a half log to the base 2 2 pi sigma tilde square e minus half log to the base 2 2 pi sigma square e which is half log to the base 2 sigma tilde square divided by sigma square and we know sigma tilde square is P plus sigma square.

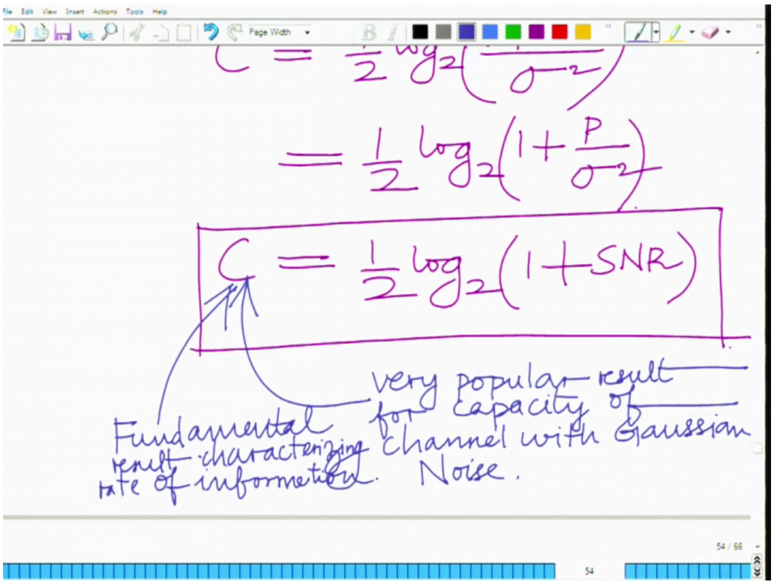
(Refer Slide Time: 30:30)



So, we have finally, C equal to half log to the base 2 P plus sigma square over sigma square equals log to the base 2 1 plus P over sigma square, but remember P over sigma square remember P is the power of the transmitted symbols X sigma square is the noise power. So, P over s sigma square is nothing, but the SNR.

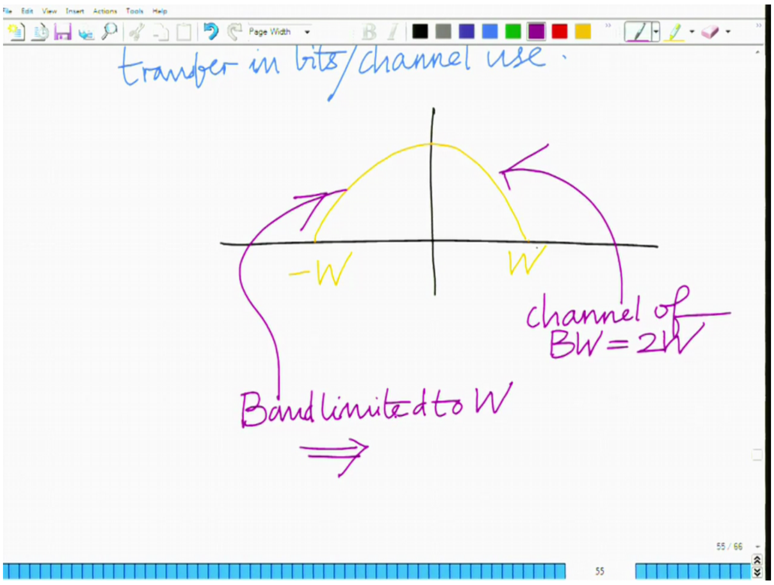
Therefore we have the celebrated result for the capacity of the Gaussian channel that is log to the base 2 1 plus SNR.

(Refer Slide Time: 31:20)



And this is the capacity this is well this is the very popular result popular and fundamental popular result for capacity of channel with Gaussian noise and also very practically relevant because as we have said many times before in this course the Gaussian channel model is one of the most frequently occurring channels. And one of the most popular models to represent or to model the channel in a typical communication system this is one of the most fundamental result which characterize what is the maximum rate at which information can be transmitted over bits per second across a Gaussian channel.

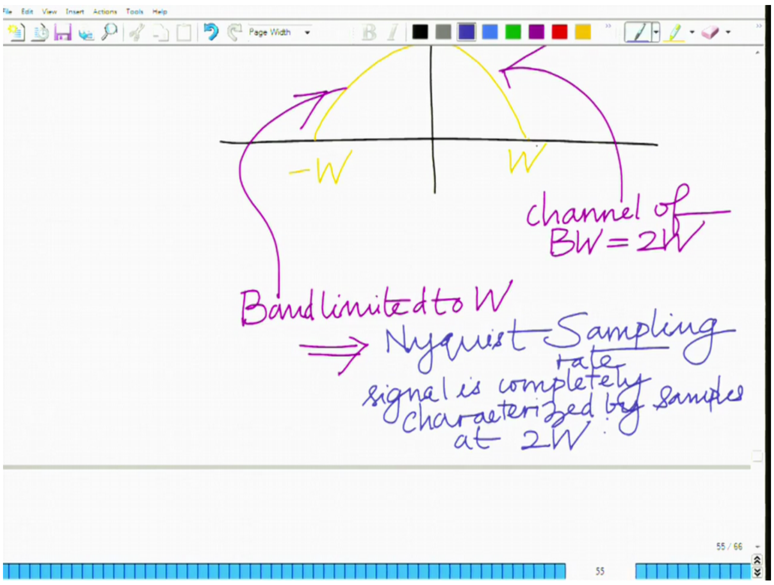
(Refer Slide Time: 32:42)



So, this is fundamental result characterizing rate of information transfer in terms of bits per channel in terms of bits per channel use.

Now, there is one small modification for this not a small modification, but if you want to find the bits per second if I consider a channel of bandwidth channel of bandwidth let us say that is we have now from the channel of bandwidth W remember we can have a signal. So, channel is band limited to W implies from the Nyquist sampling rate the signal.

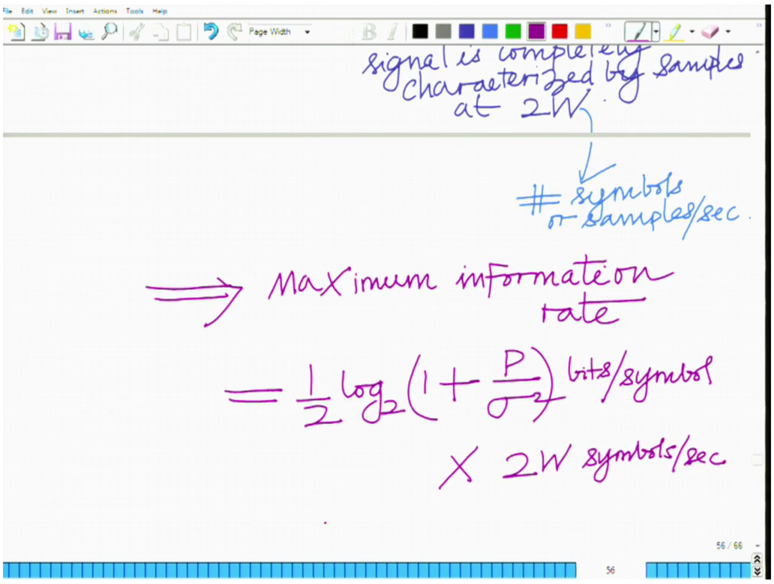
(Refer Slide Time: 33:58)



Now, if you go back to our Nyquist we have that signal is completely characterized by samples at rate 2 W right at samples at rate 2 W that is sampling rate that is the sampling rate if we have 2 W samples per second that is the maximum number of distinct samples that is signal from the one of the ways to interpret the Nyquist sampling theorem is that if your signal that is band limited to W correct, I can sample it at the rate 2 W and that completely characterize and I can reconstruct the signal from these sample, so, the entire signal. So, the information in the entire signal is basically embedded in this 2 W samples right 2 W samples per second.

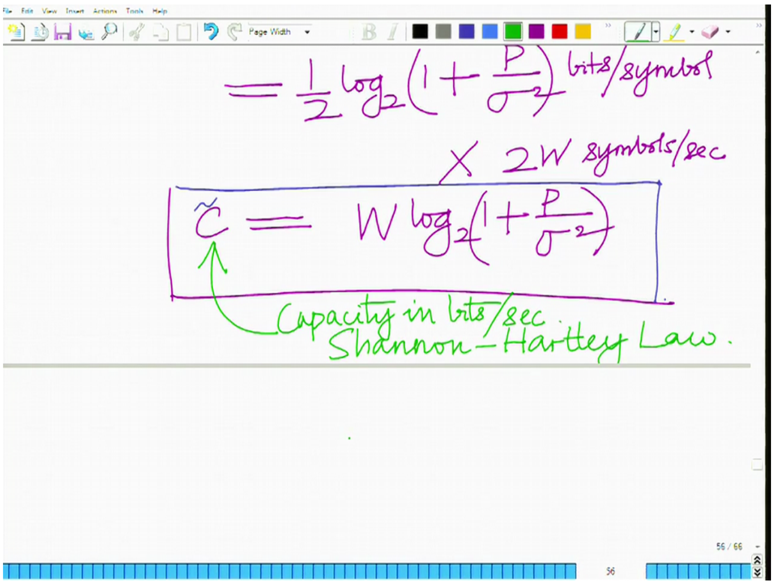
So, we can sample the signal at 2 W samples per second I can basically represent the entire signal I can capture the entire signal. So, these are the maximum number of novel sample or these are basically this is the rate at which I can transmit novel symbols or novel samples across the channel and now we have seen that per channel use we can transmit half log to the base 2 1 plus SNR bits per channel use.

(Refer Slide Time: 35:44)



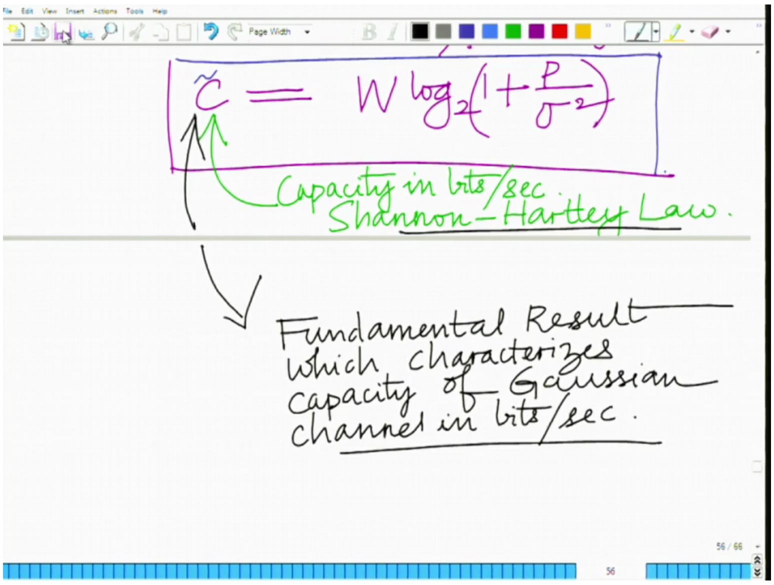
Therefore maximum information rates, so, this is the number of symbols per you can think of it number of symbols or samples per second implies maximum information rate the maximum information rate is equal to half log to the base 2 1 plus your sigma square that is bits per symbol times 2 W symbols per second.

(Refer Slide Time: 36:45)



And this gives your basically W log to the base 2 1 plus your sigma square that is a capacity of Gaussian channel in or you can use it as C tilde that is this is in bits per second capacity in previously in the bits per channel use this is and needless to say this also known as the Shannon Hartley law this is known as Shannon Hartley law. Shannon of course, again no need to remind you already described him the father of information theory from whose fundamental works all these result is regarding capacity of various channel follow in particular of course, Shannon's work is much more general characters provides a framework to characterize the capacity of any general channel in particular this result is for the Gaussian channel and its very fundamental its relevant its fundamental its own right. Because as I have told you the Gaussian channel is one of the most frequently occurring channels and its one of the most popular channels models and also widely applicable in practice.

(Refer Slide Time: 38:22)



So, therefore, this result the sub result of the general capacity of framework is a fundamental relevance and also very popular. So, this is a fundamental result and very elegant and simple result needless to say. So, this is a fundamental result which characterizes which characterizes the capacity of Gaussian channel in bits per seconds we already said this is the Shannon Hartley law.

So, in this module we have completed the derivation of one of the fundamental result that is the capacity of the Gaussian channel which we have shown basically in terms of bits per channel uses half of log 1 plus SNR in terms of bits per second over a channel of bandwidth W is W times log to the base 2 to 1 plus SNR alright. So, we will stop here and look at another aspect in the subsequent modules.

Thank you very much.