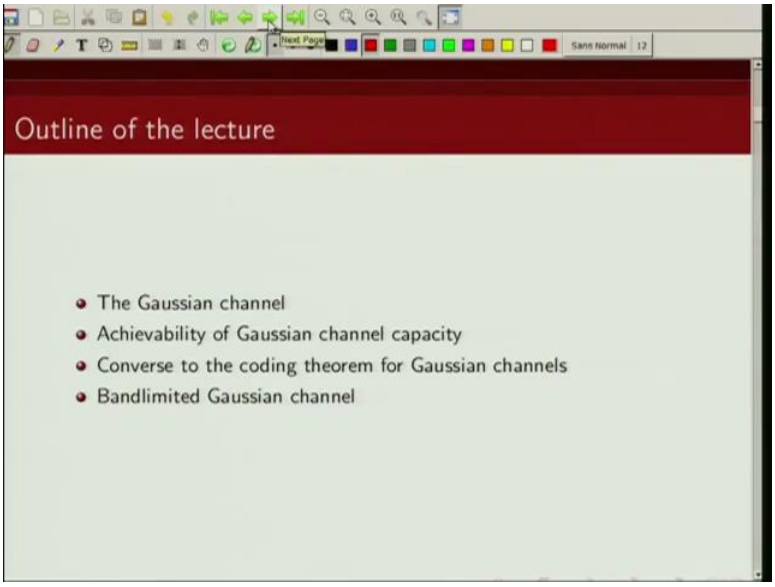
**An Introduction to Information Theory**
**Prof. Adrish Banerjee**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology, Kanpur**

**Lecture – 12A**
**Gaussian Channel**

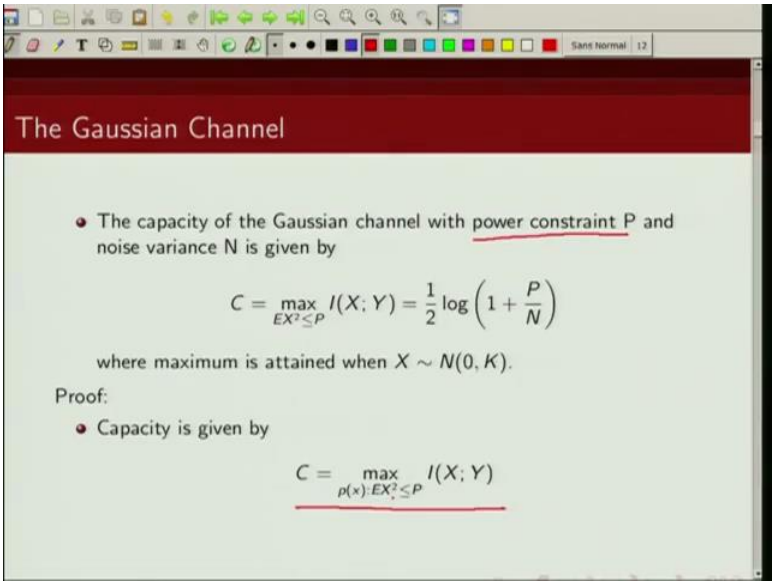
Welcome to the course on an introduction to information theory. So, today we are going to talk about Gaussian channel in the last class we talked about entropy for continuous random variable one of the important channels continuous values channel is this Gaussian channel.

(Refer Slide Time: 00:47)




And today we are going to talk about capacity of a Gaussian channel, we will first talk about capacity of a Gaussian channel and we are going to prove the achievability of Gaussian channel capacity subsequently. We are going to show the converse to the coding theorem for Gaussian channel and finally, in this lecture we are going to talk about band limited Gaussian channel and we will compute the capacity of band limited Gaussian channel.

(Refer Slide Time: 01:12)

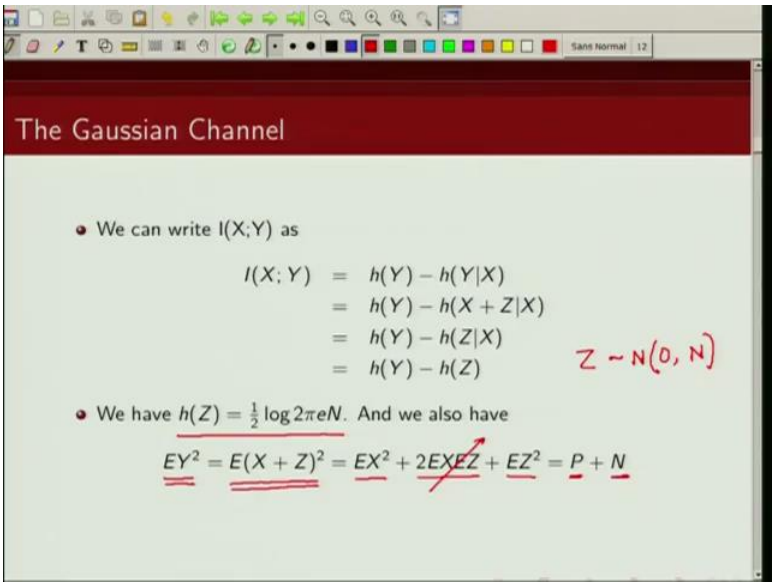



So, as I mentioned Gaussian channel are very important continuous value channel. If we write is corresponding discrete motion, we can write it. Let us see the input is x i and noise is additive which is Gaussian distribution denoted by z i what we get as output is y i this is a discrete motion of a continuous value Gaussian channel. So, first result that we are going to show you is the capacity of the Gaussian channel, with power constraint p and noise variance n is given by this expression. Where the maximum is attained when, x is Gaussian distributed with 0 mean and variance k. Now let us first look at what is the capacity. If we do not have these constraints power constraints or if you do not have noise it is very easy to see, if the noise is 0 then, whatever we send will be reliably received. So, the channel capacity is infinite in that particular case similarly is there is no power constraint.

We can make other input arbitrarily large such that it is received correctly at the receiver. So, if we do not have this power constraint then also the capacity of this channel is infinite. Now let us look at the capacity of this Gaussian channel, when we do have a power constraint and we do have a noise variance given by n. So, from the definition of channel capacity its maximum information between the input x and output y and this maximization is taken over all input distribution p of x. Remember, we have a power constraint here it is as an average power constraint. So, expected value basically of x

square is less than p. So, we need to compute maximum mutual information under all input distribution given this power constraint, from the definition of the mutual information.

(Refer Slide Time: 03:46)

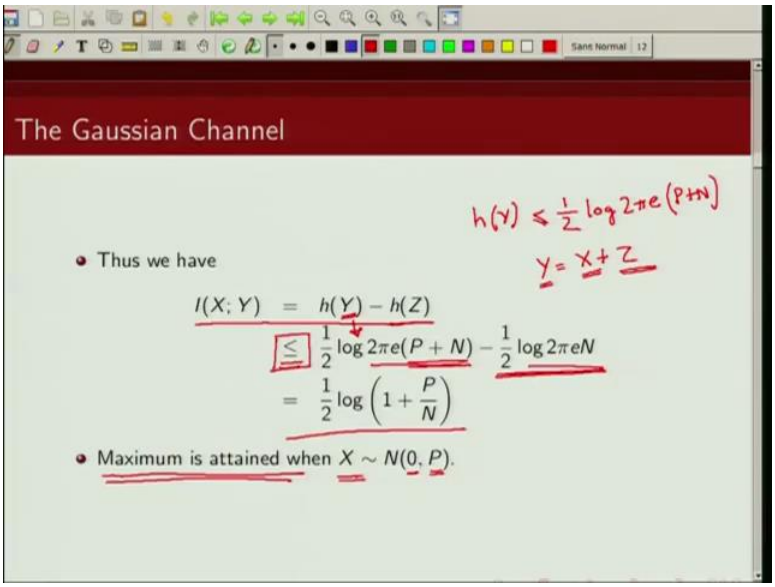



We can write mutual information between x and y as differential entropy of y minus differential entropy of y given x. Now what as I said y is the output of the channel which is nothing, but input x plus additive noise additive y additive Gaussian noise, which is z. So, i can write this as differential entropy of y minus differential entropy of x plus z given x. Now what is the uncertainty also what is the entropy of x plus z given x that only depends on z given x. So, we can write this expression as differential entropy of y minus differential entropy of z given x, now z is which is noise is independent of the signal x. So, the differential entropy of z given x is equal to differential entropy of z because z the noise is independent of the signal x. .

So, we can write this mutual information between x and y as differential entropy of y minus differential entropy of z. Now we know that z is Gaussian distributed and if we have a Gaussian distributed random variable. We know what is its differential entropy is. So, if we assume that z is Gaussian distributed its mean. Let us say 0 and variance n then, the differential entropy of z is given by half log of 2 pi e times n 2, we can compute the

expected value of y square y is nothing, but x plus z. So, this can be written as expected value of x plus z square which is nothing, but expected value of x square plus expected value of z square plus 2 times expected value of x and expected value of z. Now since the expected value of z is 0 this term will be 0. So, what we would get is expected value of y nothing, but expected value of x square plus expected value of z square now expected value of x square is p and expected value of z square is n. So, expected value of y square is p plus n.

(Refer Slide Time: 06:52)

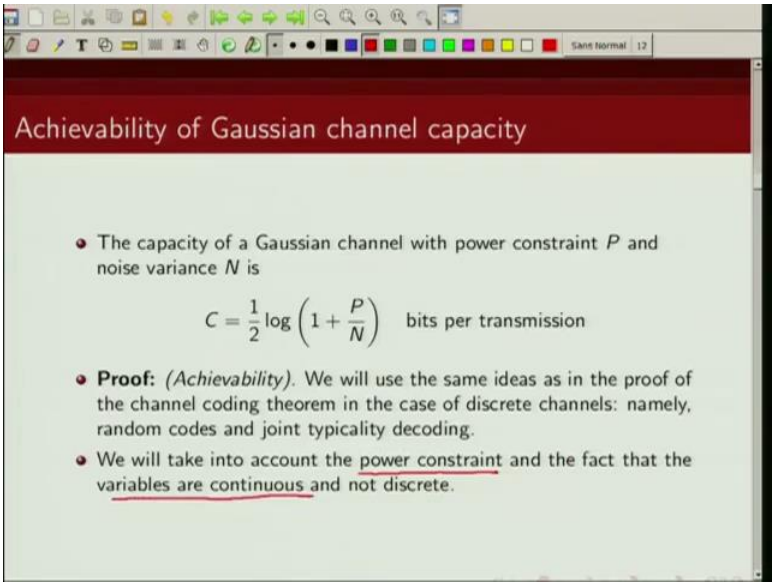



Now we have just shown that the mutual information between x and y can be written as differential entropy of y minus differential entropy of z. Now we know, what is the differential entropy of z that is because, z is a normal distributed with 0 mean and variance n, the differential entropy of z is given by this. We also know if y is a random variable which has variance p plus n then the differential entropy of y is upper bounded by the differential entropy of Gaussian random variable. So, we can upper bound the differential entropy of y by the differential entropy of a Gaussian random variable with same variance and in this case expected value of y square is p plus n. So, we can upper bound this differential entropy of y by differential entropy of a Gaussian random variable with variance given by p plus n, now variance given by p plus n.

So, now, basically what we have said this results follows from a theorem, that we have proved for differential entropy that is y is the random variable with variance p plus n then its differential entropy is upper bounded by differential entropy of a Gaussian random variable with same variance. So, clearly this is equal when y is also Gaussian and y is Gaussian, when x is also Gaussian because y is x plus z we know z is Gaussian distributed, If x is also Gaussian distributed then y will be sum of 2 Gaussian distributed random variable will also be a Gaussian distributed, in that case y will also be a Gaussian distributed random variable and in that case this would have been equality.

So, combining the terms, this is log e minus log v kind of term which can be combined. So, that becomes this. So, what we have shown is mutual information is upper bounded by half log of one plus p times n now remember the maximum is attained. When x is also Gaussian distributed with mean 0 and variance p in that case as i said y is also going to be Gaussian and this inequality would be equality. So, what we have shown is capacity of a Gaussian channel with power constraint p and noise variance n is given by this expression. Where the maximum is attained when input x is also Gaussian distributed with mean 0 and variance k.

(Refer Slide Time: 10:20)

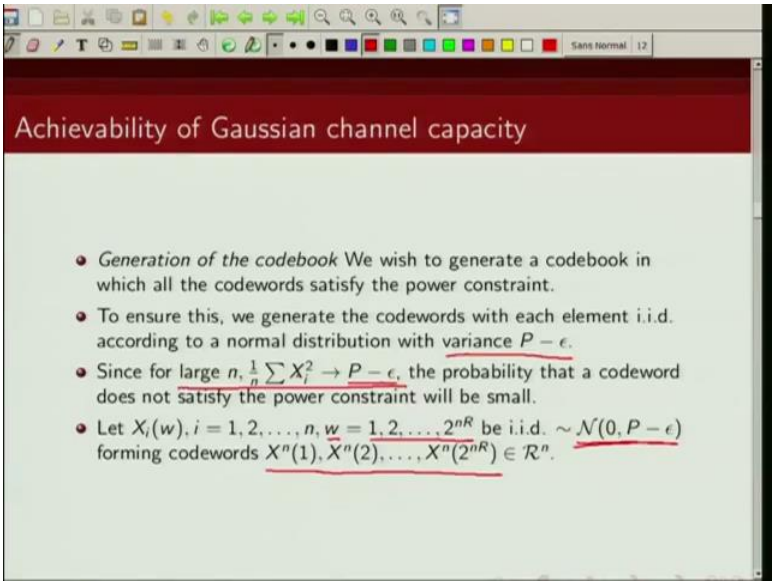



Now, we are going to show the achievability of this. So, all as long as we choose rate less

than this capacity we are going to get reliable communication. So, let us prove the achievability part of this capacity theorem. So, we are going to use the same ideas that we have used earlier to prove channels, noisy channel, coding theorem the difference here is, now in that case we had discrete random variables.

Here we have continuous random variables this difference here we also have this average power constraint which we did not have let me prove a channel noisy channel coding theorem. So, we are going to use a same ideas as the as we have used including channel coding theorem for discrete channels and what was those ideas we are going to use random randomly. We are going to generate this code word and we are going the receiver, we are going to use typical set coding we are going to use this ideas of random codes and joint typical decoding as i said, we are going to take into account that now there is a constraint on powers is that is violated basically that is the error and the variable are now continuous not discreet. So, that is the difference from the proof that we had done earlier.

(Refer Slide Time: 12:03)

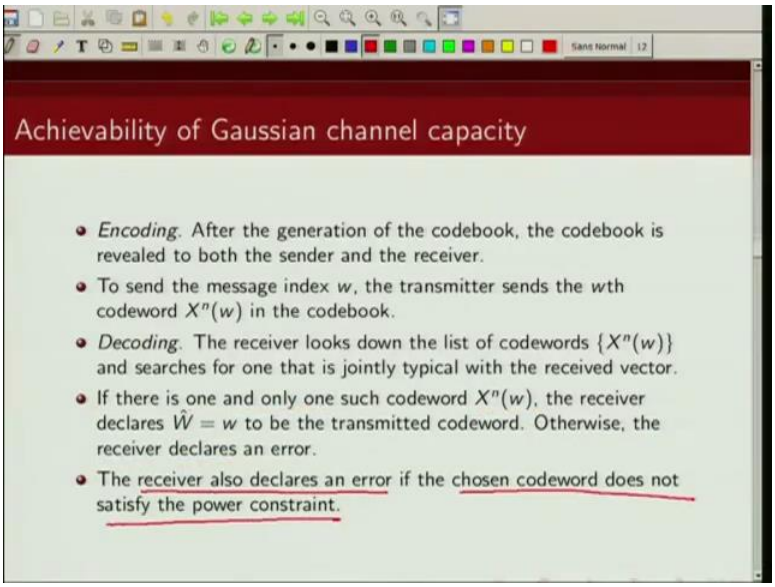



So, first step is generation of the codebook. So, we are going to generate a codebook in which all code words are going to satisfy the power constraint. So, to ensure this we are going to generate code words with each element identically or independently generated

according to Gaussian distribution, with variance given by p minus epsilon. Where p epsilon whether is small positive number now by choosing noise variance p minus epsilon we are going to ensure that the average power constraint is not violated. So, we can see from last large numbers.

If we generate each code word normal distributed variance gained by p minus epsilon by law of large numbers the average value will basically tend to p minus epsilon and hence the probability that a code word does not satisfy, this average power constraint will be small. Let x i w where w are these 2 raise power n minus n r code words be i i d distributed and as we said we are generating this code words with variance given by p minus epsilon. So, let these code words be x n 1 x n 2 x n 2 raise power n r. So, these are n length code words which are randomly generated with Gaussian distributed with variance p minus epsilon.

(Refer Slide Time: 14:05)

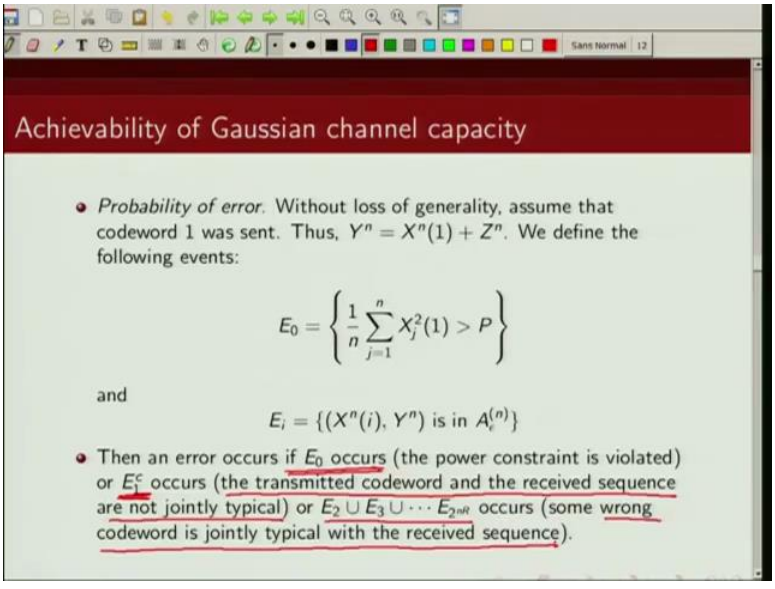



So, after the generation of the codebook the codebook is revealed to both the sender and receiver. Now if we want to send message w, what the sender does it sends the wth code word x n w in the codebook. Now at the decoder we are going to do typical set decoding. So, how does typical set decoding works. So, we are going to see which code word is jointly typical with a received codebook and, if there is only 1 code, code word which is

jointly typical with this and there is know the code word which is jointly typical with a received codebook then, the decoder will correctly decode otherwise there is a decoding error also if the power constraint is valid then also there is an error.

So, the receiver is going to look down the least of code words which are just denoted by this x and w and it searches for this one code word which is jointly typical with this received vector y n w, it tries to find that one code word which is jointly typical with this received vector. Now if there is one and only one note this if there is one and only one such code word then, the receiver and let see that code word is x and w then, the receiver is going to declare that the transmitted code word was w otherwise the receiver will make an error since. There either more than, 1 code word which is jointly typical with the receive sequence or there is no code word which is jointly typical with the receive sequence then, the decoder will make an error also when power constraint is valid it then also receiver will make error. So, as I said the receiver also declares an error if the chosen code word does not satisfy this average power constraint.

(Refer Slide Time: 16:36)



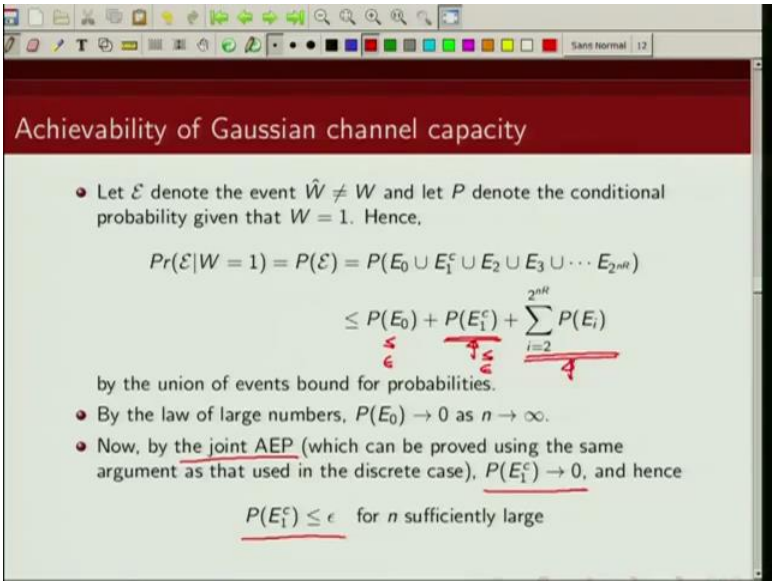

Now, let us compute what is the probability of error. So, when the decoder is doing joint typical set decoding. Let us compute what is the probability of error when, we are sending this wth code, word over this Gaussian channel. So, without loss of generality

we will assume let say the code word one was sent. So, we are going to assume code word 1, was sent without loss of generality then the received code word y n is going to be x n 1 plus z n this is a noise this is the code word that was transmitted and this is what we have received.

Now, let us define the error events the first error event which i am defining by e0 is the error event corresponding to the violation of average power constraint. So, e naught is if the average power constraint is violated then, the event e 0 happens. So, this corresponds to violation of power constraint the average power constraint that we had let e i be the event where, x n i and y n belongs to this jointly typical set. So, e i corresponds to the event when x n I and y n are jointly typical. So, clearly an error happens if event e 0 occurs or if e 1 compliment occurs why because, we have sent code word 1. So, there will be no error if y n is jointly typical with x n 1; however, event e 1 com compliment will occur. If x 1 x n 1 and y n are not jointly typical and that is this event e 1 compliment.

So, if the transmitted code word and the receive sequence are not jointly tip typical that corresponds to the event e 1, compliment then also error happens or if any of these events occur e2, e3, e4, e2 raise power n r; that means, if any other codeword other than the code word number one is jointly typical with the receive sequence y n then there is an error and that that is denoted by this union of events e2, e3, e4, e2 raise power n r. So, this corresponds to a wrong code word which was not sent to be jointly typical with the receive sequence.

(Refer Slide Time: 20:02)



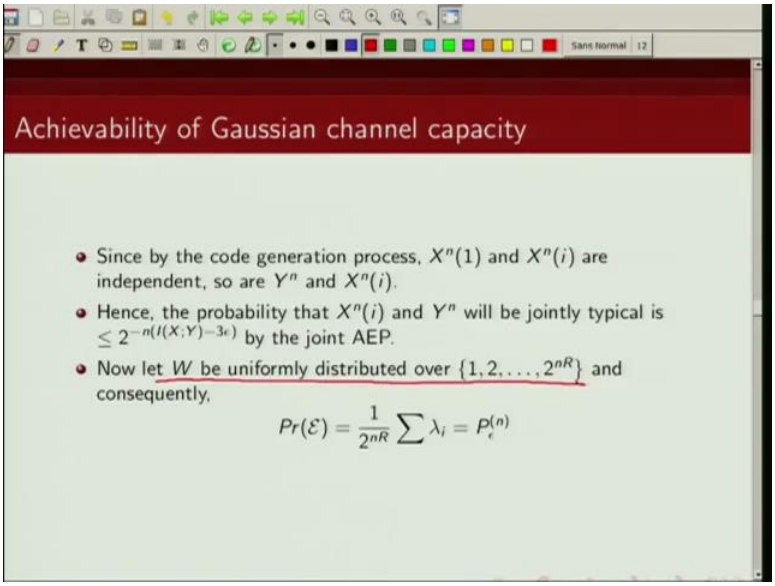

So, now that we have a numerated error events let us compute the probability of error given in the code word 1 was transmitted. So, let e denotes the event that what, which is an estimate of the code word that we sent is not same as w and let p be the conditional probability of error given w is 1. So, this is given by probability that event e0 happens which corresponds to violation of power constraint union with event e1 compliment which is the event that x and 1 and y n are not jointly typical or union with e2, e3, e4, e2 raise power n r which corresponds to the event that are wrong code word is jointly typical with the receive sequence y n.

Now this can be upper bound bit using union bound using union bound, we can upper bound this probability of union of these events by. So, we can upper bound this using union bound this is probability of e 0 upper bounded by probability of e 0 plus probability of e one compliment plus probability of e 2 plus probability of e 3 up to plus probability of e 2 raise power n r.

Fine next now, we are going to evaluate these probabilities now first let us, look at probability of e naught now remember we were generating these code books with variance p minus epsilon where epsilon is a positive number small number close to 0. So, if we are generating our code books like this by law of large number probability of

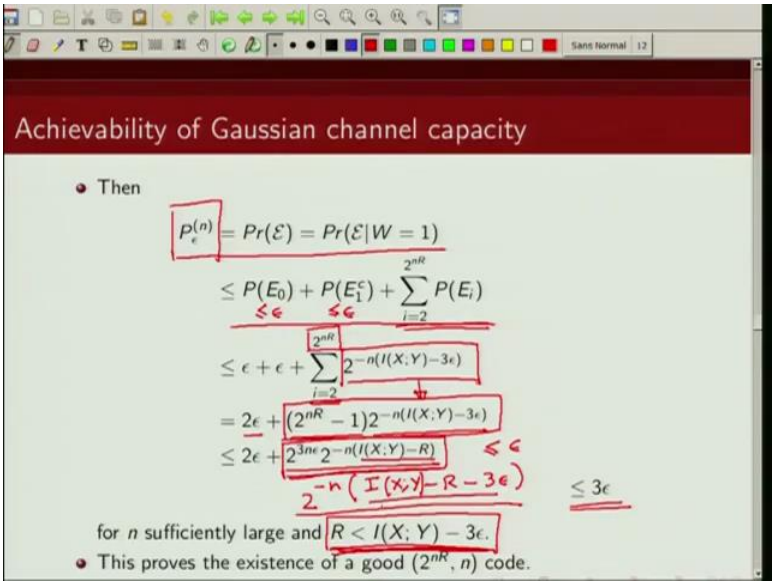
occurrence of this event e0 is 0 as n is very large. So, this probability can be upper bounded by small quantity epsilon similarly, let us look at this probability of occurrence of the event that e one that x n one is not jointly typical with y n. Now we know from the property of joint ep which we proved in the last class that probability of this event is also 0 as the n goes to infinity. So, this can also be upper bounded by epsilon. So, we have upper bounded this by epsilon we are upper bounding this by epsilon.

(Refer Slide Time: 23:27)




Now, let us look at this error event now since the way we are generating our codebooks x n one and x n i where, I is the mother code, codeword they are independent. So, y n and x n i are also going to be independent and we have shown that we have shown a result in the last class that what is the probability that if xn, xn i and y n which are chosen independently. What is the probability that they will be jointly typical? So, what is the probability that x n i and y n are jointly typical this we have proved in the last class by the property of joint a e p this is upper bounded by 2 raise to power minus n times mutual information between x and y minus 3 epsilon. So, if we let w to be uniformly distributed over all these 2 raise power n r code words then subsequently the probability of error is given by this average probability of error.

(Refer Slide Time: 24:43)

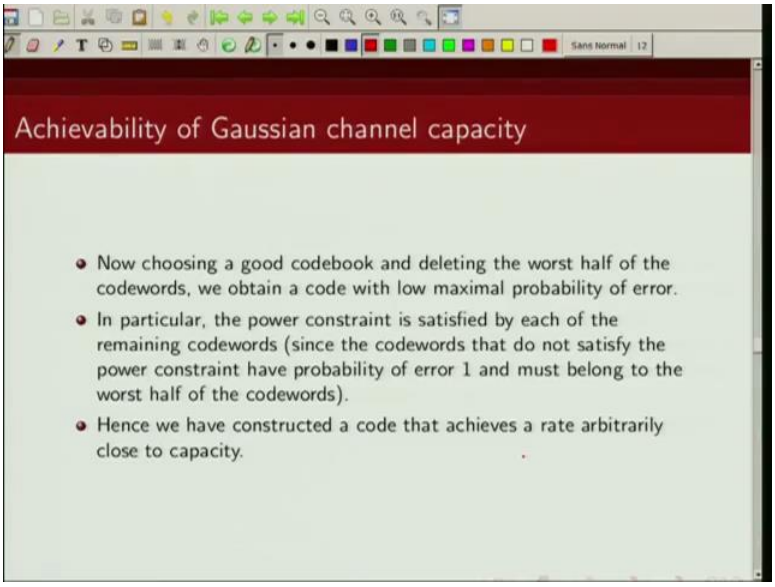



Now as we said average probability of error without loss of generality can be written as probability of error given code word 1, was send using a union bound this was upper bounded by sum of these error events we have already shown, this is upper bounded by epsilon this is upper bounded by epsilon and this is upper bounded by this quantity this follows from the properties of joint a e p which we have proved in the last class. So, then this term does not depend on i. So, I can sum over all i going from p to 2 raise power n r. So, i get this epsilon will be 2 epsilon and this summation over i from 2 to 2 raise power n r this will be this term and then, we have this one which is this now combining the terms containing n i get this this can be written as 2 raise power 3 n epsilon into 2 raise to power minus n mutually is a minus r and this quantity.

If we choose our r to be less than mutual information being x and y minus 3 epsilon see this term, we can write as 2 raise to the power minus n mutual information between x and y minus r minus 3 epsilon can write in this way. So, if we choose our r to be less than mutual information between x and y minus 3 epsilon then, this term will be positive and. So, for large n this will also go to 0. So, this whole term can be upper bounded by epsilon. So, in other words this probability of error can then be upper bounded by 3 epsilon provided my rate is less than, mutual information between x and y minus 3 epsilon. So, this shows that there exists good codes with rate given by this and in that

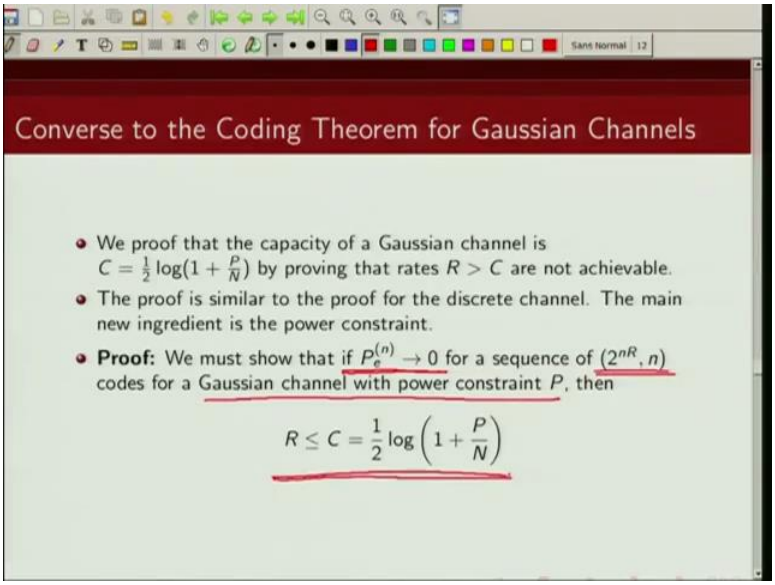
case the probability of error will go towards 0 as n goes to infinity.

(Refer Slide Time: 27:52)



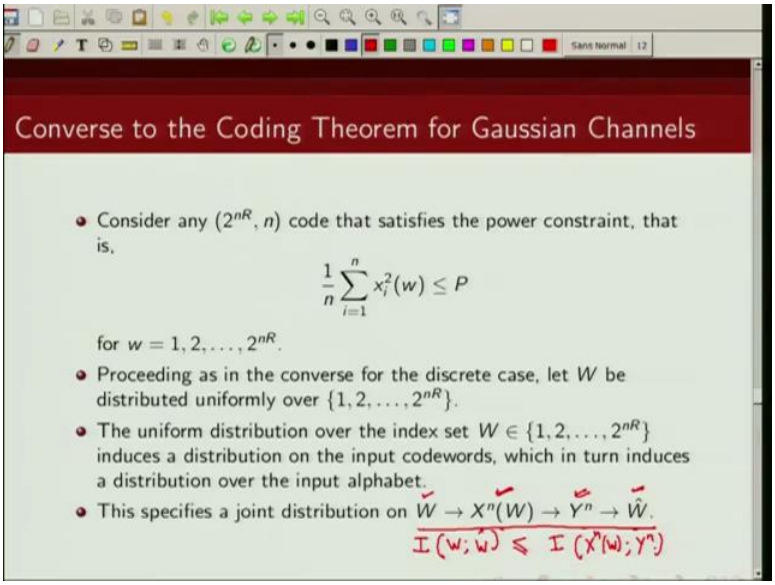

Now, if we delete half of this words code words, we obtain a code with low maximum probability of error in particular the power constraint will be satisfied by this remaining code words because we are already throwing of these. So, called gauss code words and hence we are constructed a code that achieves a rate arbitrarily goes to capacity and gauss probability of error which is which goes probability of error which goes to 0 as n goes to infinity. So, this proves the achievability of Gaussian channel capacity.

(Refer Slide Time: 28:40)



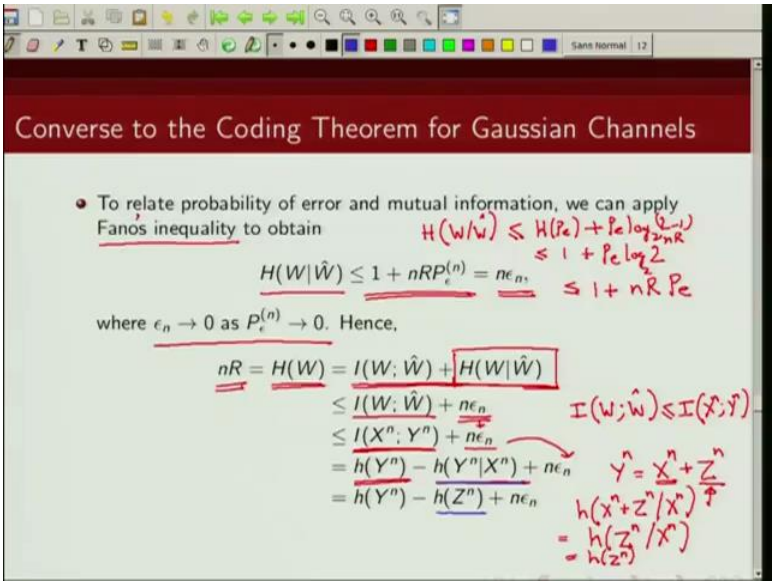

Next we are going to show the converse of the channel coding theorem for Gaussian channel. So, we are going to show that we prove that a channel capacity of Gaussian channel is given by this and we prove that rate greater than capacity is not achievable or we are going to show that; if probability of error goes to 0 then the rate is less than, channel capacity. Now the converse proof is very similar to the proof of discreet channel there is one additional constraint that we have here now, which is the average power constraint. So, we are going to show you now that this probability of error goes to 0 for a sequence of code given by these parameters for a Gaussian channel, with power constraint p as long as rate is less than capacity which is given by this expression.

(Refer Slide Time: 30:03)




So, we are considering a 2 raise power n r comma n code which satisfies average power constraint which is given by this expression and this is satisfied by all code words w going from one to 2 raise power n r. Let w be uniformly distributed now this uniform distribution over this index set induces a distribution of input code words which in intern induces distribution on the input alphabet and we can show, that w x n w y n and w hat they form a mark of chain. So, w is my code word index x n is my encoder encoded codebook y n is the received codebook and w hat is an estimate of w based on what I receive which is y n.

(Refer Slide Time: 31:17)



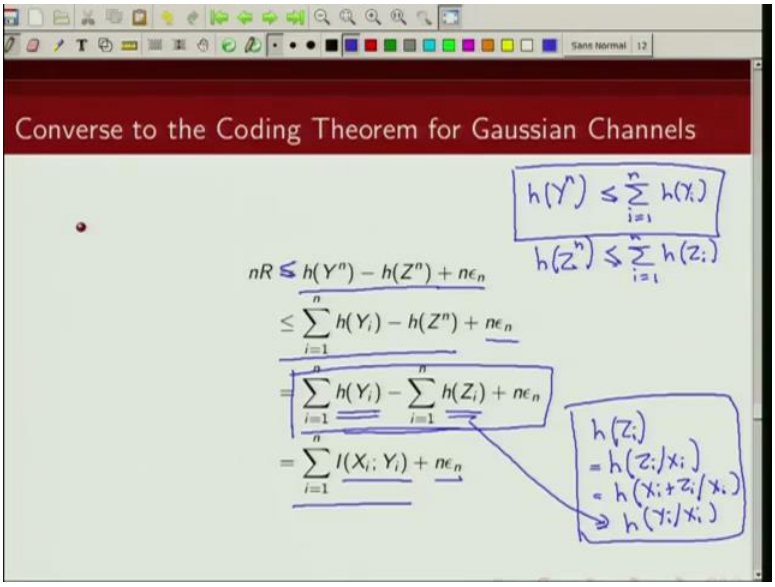

Now, we are going to take help of Fanos lemma to relate probability of error and to mutual information. Now we in the class has used this version of Fanos lemma which was given by now we did each of p e plus p e log of l minus 1 we use this portion and we said this can be weak end because binary entropy function is less than equal to 1. So, this can be weak end also 1 and this can be we further we can p e log of all the possibilities l. So, number of code words here is 2 raise to power n r. So, if we take log of that this will become less than equal to one plus n times r n probability of error.

So, this is the version that we are using uncertainty in w given w hat is upper bounded by 1 plus n r times probability of error and as we said here we are going to show that, if probability of error goes to 0 for large n then the rate has to be less than c. So, this is you know. So, uncertainty in w given w hat is less than equal to n times. So, small value of epsilon where epsilon goes to 0 and probability of error goes to 0. Now w is uniformly distributed over this index said 1 2 2 raise power n r. So, uncertainty in w is given by log of 2 raise power n r which is nothing, but n r now from the definition of mutual information we can write uncertainty in w in terms of mutual information between w and w hat plus conditional entropy of w given w hat. Now if probability of error goes to 0 as, n goes to infinity we have shown that this term is upper bounded by n times.

So, I can write then uncertainty in w is upper bounded by mutual information between w and w hat plus n times epsilon. Now we just saw in the last slide that w x n w y n and w hat they form a mark of chain. So, then from data processing lemma we know that mutual information between w and w hat is going to be less than mutual information between x n w and y n. So, this follows from the data processing lemma. So, then we can write this that the mutual information between w and w hat is less than equal to mutual information between x n and y n. So, that is what we are writing here and this comes as n times epsilon n now following the definition of mutual information, we can write the mutual information between x n and y n as differential entropy of y n minus differential entropy of y n given x n and this is n times epsilon.

Now what is y n y n is the output of the Gaussian channel which is x n plus z n where z n is the Gaussian noise this 0 mean n variance n. Now uncertainty in y n differential entropy of y n given x n can be written as differential entropy of x n plus z n given x n and this is nothing, but uncertainty in z n given x n now note that the noise z n and the signal x n are independent and hence this can be written as uncertainty in differential entropy of z of n.

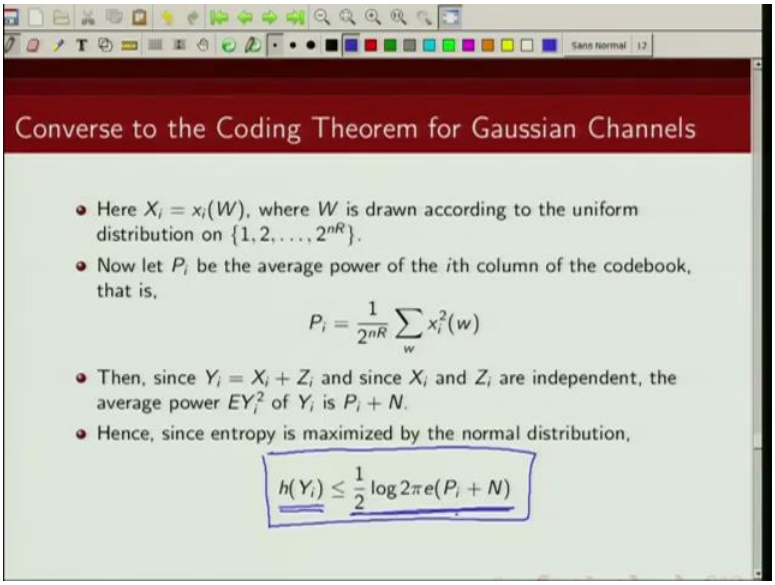
(Refer Slide Time: 36:55)




So, I can write differential entropy of y n given x n as differential entropy of z n next. So,

what we have shown. So, far is n r is actually less than equal to h of y n minus differential entropy of z n plus n times epsilon. Now this can be upper bounded by this again this is straight forward to prove we have prove this using chain rule we can write h of y n in terms of y1 y2 y3 yn conditioned on these y i and we know that conditioning cannot increase entropy. So, from that we get this result similarly we can write h of z n as less than equal to summation over i from 1 to n h of z n. So, this is what we are writing here plus n times epsilon n this is this term here now, we have just shown h of z i is same as h of z i given x i and this is same as h of x i plus z i given x i which is same as h of y i given x i. So, this is differential entropy of y i this is differential entropy of y i given x i. So, then this can be written as mutual information between x i and y i.

(Refer Slide Time: 39:17)

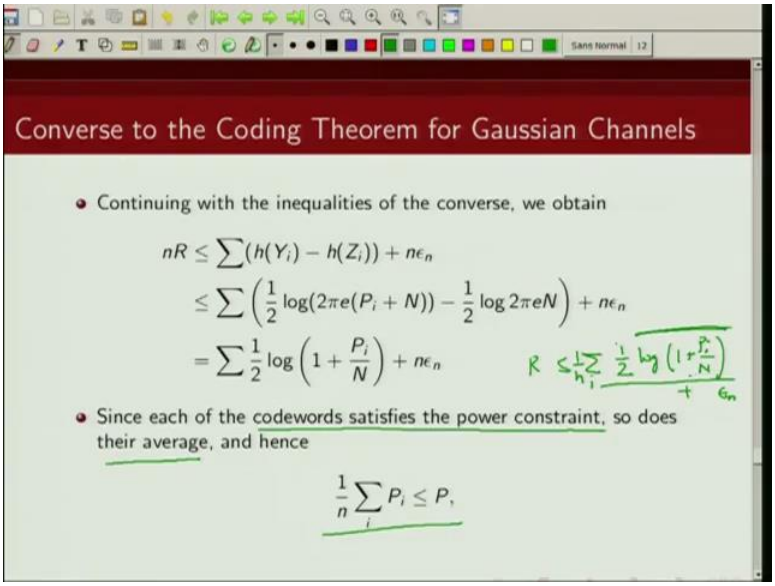



So, from this result and this result we can write n r is upper bounded by summation of mutual information between x i and y i where, summation is over I going from one to n plus n times epsilon now let x i be w i w where x i w where w is drawn uniformly over this index set and let p i be the average power of the I th column of this codebook we know the x i signal and z i the noise they are independent.

So, the average power of the received sequence y i can be written as p i plus n where p i is the power of x i n is the noise variance now again this result will have proved in the

previous class that if, y has y i have the same second order movement then Gaussian random variable will have the maximum differential entropy. So, h of y i is upper bounded by the differential entropy of an equivalent of differential entropy of Gaussian random variable with same variance. So, basically h of y i is an upper bounded by half log of 2 pie e times p i plus n. So, the differential entropy of y i is upper bounded by the differential entropy of Gaussian random variable which has 0 mean and variance given by p i plus n actually mean can be anything because translation does not change the differential entropy.

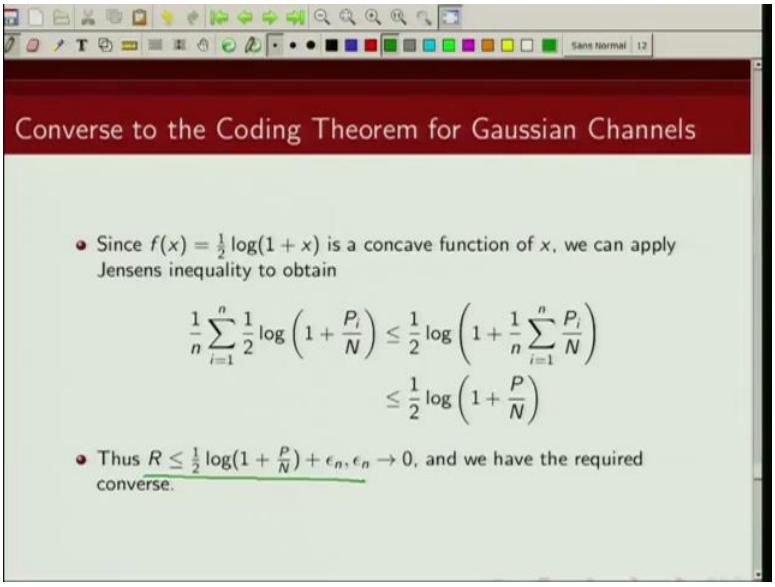
(Refer Slide Time: 41:13)




Now continuing with the inequalities that we had we just had the expression that n of r is upper bounded by summation over all I. So, I 1 to n h of y i minus h of z I, now we just have now showed you that differential entropy of y i is upper bounded by this quantity and since z i is Gaussian distributed with variance n is differential entropy is given by this expression. Now, combining these 2 log times this term and this term we get this expression the n r is upper bounded by summation from I goes to one to n half log of one plus p i divided by n plus n times epsilon. Now the way we generated these codebooks code words they all satisfied average power constraint. So, they all satisfy power constraint because we were we are generating these code word with variance given by p minus epsilon. So, this code words satisfying power constraint. So, that average is also

going to satisfy the power constraint.

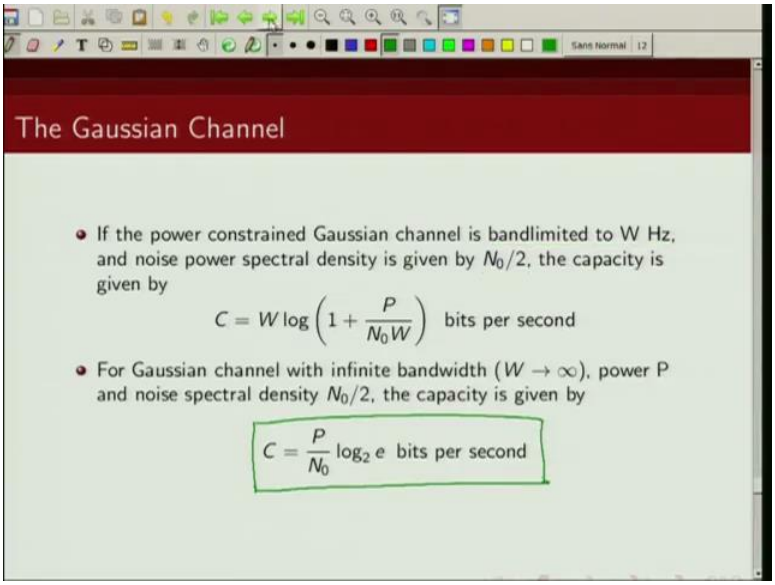
(Refer Slide Time: 42:53)




Now we know that log is a concave function and if a function is concave, we know from jensens inequality that expected value of the function is less than equal to function of expected value. So, this is what we know from jensens inequality. If f of x is a concave function in this case log is a concave function of x. So, then here we have this expression that n r is less than equal to summation over all I one half log of 1 plus p i n plus n times epsilon n if I divide this by n what I would get is this will go away this will go away and I will have here one by n.

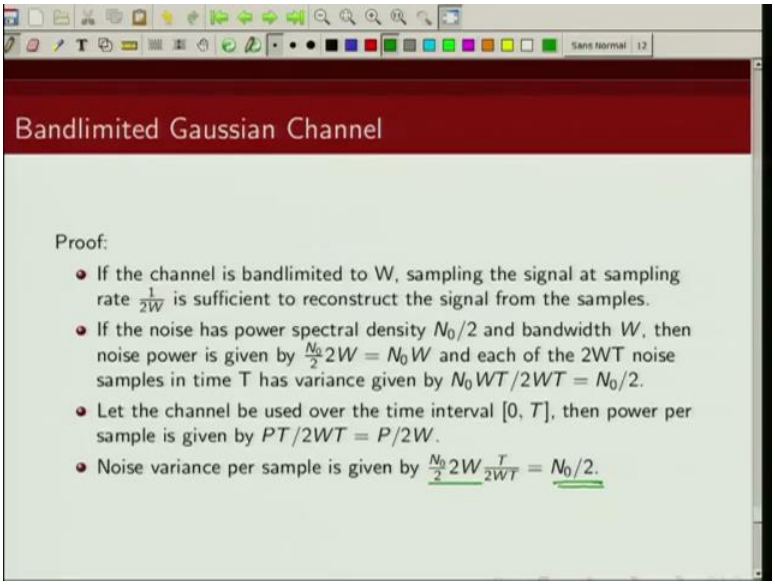
So, have I have this expression r is less than equal to 1 by n summation this term plus epsilon now I know that log one plus x is a concave function of x. Then expected value of log function will be upper bounded by log of expected value of this x. So, that is what we are doing here. So, this is expected value of the log function this is from jensens inequality upper bounded by function is log of the expected value alright now this one by n summation p i is nothing, but average power constraint. So, what we have shown here now is that r is less than equal to one half log 1 plus p by n plus epsilon. So, what we have shown is if probability of error goes to 0 then, r is also less than channel capacity. So, r is less than channel capacity if probability of error goes to 0 as n goes to infinity.

(Refer Slide Time: 45:48)




So, this proves the converse of the channel coding theorem for Gaussian channel next, we are going to talk about what is the capacity of a Gaussian channel. If the Gaussian channel is band limited to w hertz and the noise power spectral density is given by n naught by two. So, we are going to show you that if, the Gaussian channel is band limited to w hertz and its noise power spectral density is given by n naught by 2 then, the capacity of this band limited Gaussian channel is given, by w log of one plus p divided by n naught w this is the capacity of the band limited Gaussian channel and if we let w go to infinity then the capacity is given by this expression.

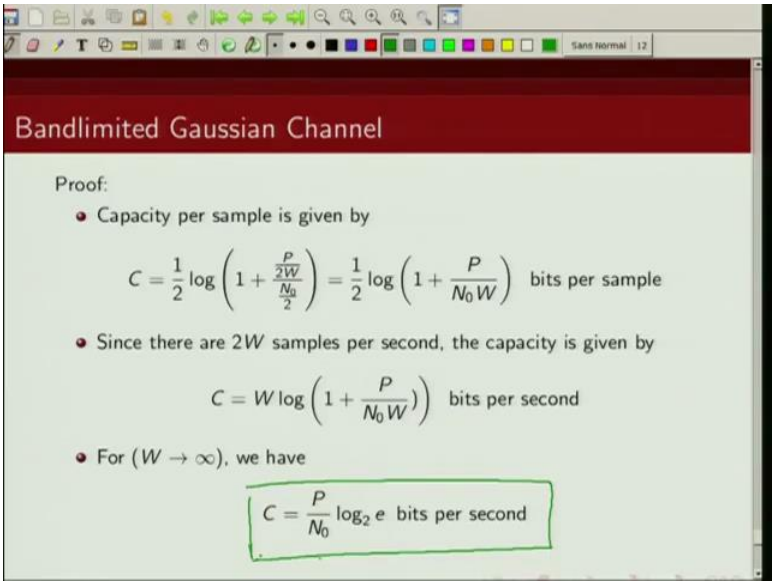
(Refer Slide Time: 46:50)




So, let us prove this. So, if the channel is band limited to w then we can sample this signal at rate 1 by 2 w and that would be sufficient to reconstruct the signal from the samples this follows from the criteria. Now if the noise has power spectral density given by n naught by 2 and the band width is w then the noise power is given by n naught by 2 times w. So, you have basically noise power spectral density given by n naught by 2 and you have band width of w to w. So, this noise power is given by n naught by 2 into 2 times w which is n naught w and over of period from 0 to t have, many noise samples because we are sampling at rate one by 2 w.

So, each of these 2 w t noise samples in time t will have variance given by n naught w times t divided by 2 w t which is nothing, but n naught by 2. Now let us use the channels over time 0 to t, power per sample will be p times t divide by total samples which is 2 w t and that comes out to be p divided by 2 times w. So, similarly we can compute noise variance per sample is n naught into 2 w t divided by 2 w t which is n naught by 2.

(Refer Slide Time: 48:45)




So, capacity per sample is given by half log one plus power per sample this is noise per sample if you go back here, noise variance per sample is n naught by 2 and power per sample is given by p by 2 w. So, then the capacity is given by half log of one plus p by 2 w divided by n naught by 2 and this is nothing, but half log of one plus p divided by n naught w now this is capacity per samples and how many sample we have per second we have 2 w samples per second.

So, then the capacity is given by w log of 1 plus p divided by n naught w bits per second this is signal power this is noise power and we can similarly compute, if you let w go to infinity and take the limit the capacity comes out to be this expression p times p divided by n naught into log of e bits per second. So, this is the expressions for capacity for band limited Gaussian channel. So, with this we will conclude our discussion on Gaussian channel. In the next lecture we will talk about parallel Gaussian channel.

Thank you.