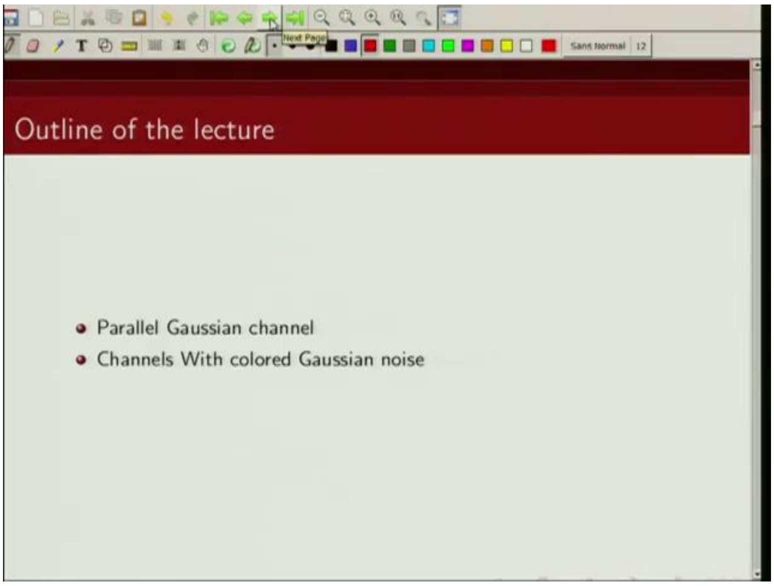
An Introduction to Information Theory
Prof. Adrish Banerjee
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kanpur

Lecture – 12B
Parallel Gaussian Channel

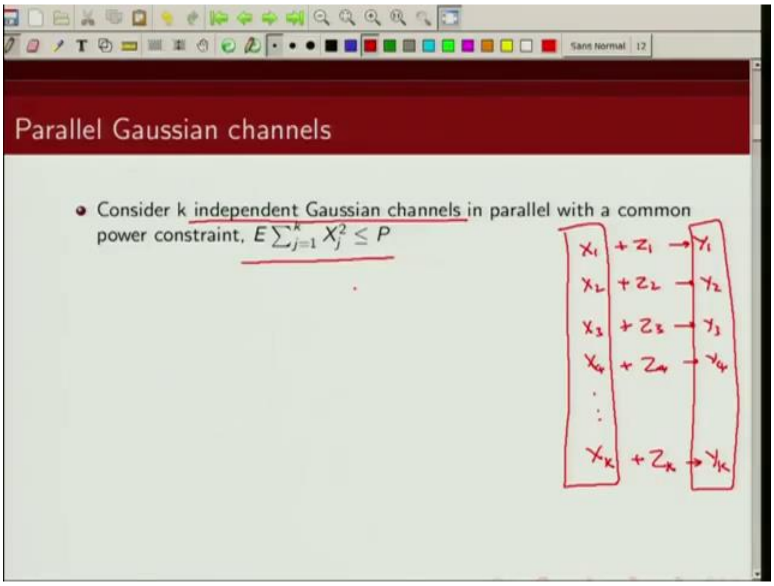
Welcome to the course an Introduction to Information Theory. In today's lecture, we are going to talk about parallel Gaussian channel.

(Refer Slide Time: 00:30)



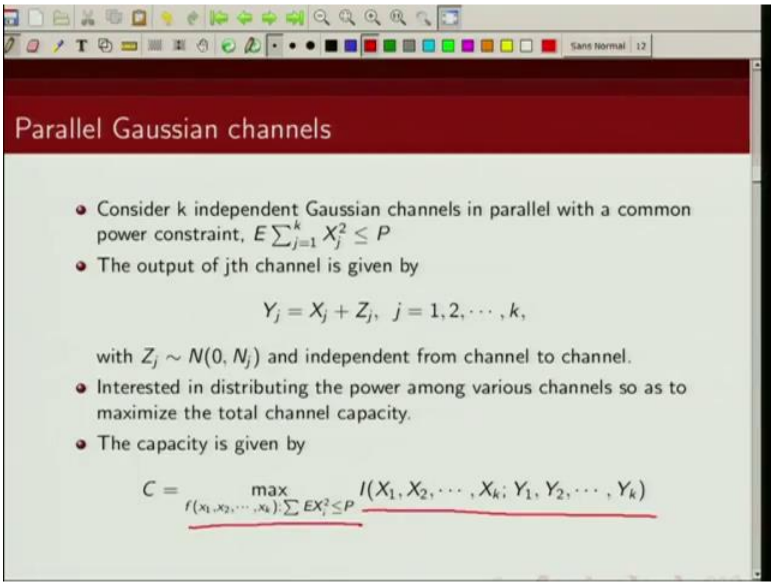
And we are going to talk about what is a capacity of parallel Gaussian channel. And in case of parallel Gaussian channel given an average power constraint, how should we allocate power between various channels, this is what we are going to talk about. Initially, we will assume that the noise is independent across various channels; and subsequently, we will consider colored noise. And we will do the power allocation for parallel Gaussian channel with colored noise.

(Refer Slide Time: 01:09)



So, what is a parallel Gaussian channel? So, let us consider K independent Gaussian channels in parallel. So, you can think of it. So, basically you have x 1, x 2, x 3, x four, x K and redirected by (Refer Time: 01:31) noise let say Z 1, Z 2, Z 3, Z 4, Z K and Z i's are independent of input X i. And this samples are also independent initially we will considered them as independent across all these channels then what we get is output is Y 1, Y 2, Y 3, Y 4, Y of k. So, our input is this and our output is this output of channel is this. So, we are considering K independent Gaussian channel in parallel and each of them have a common parallel constraint which is basically expected value of X 1 square plus X 2 square X 3 square X K square that should be less than equal to E.

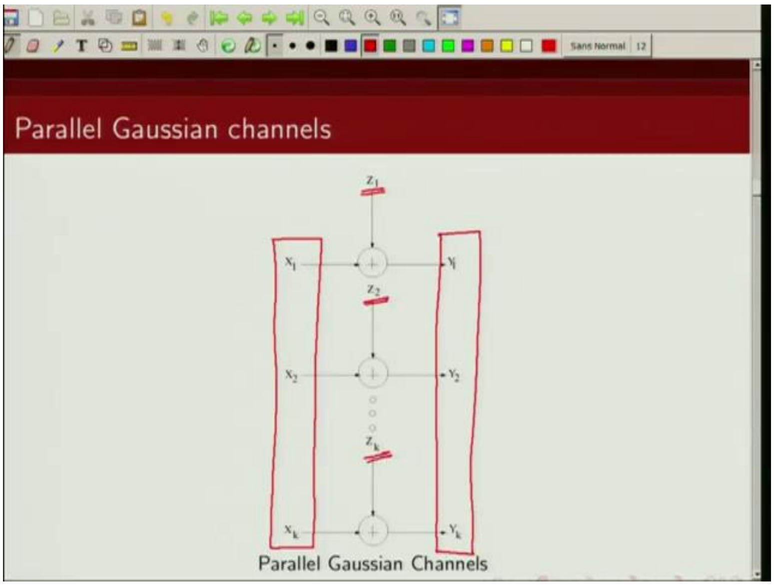
(Refer Slide Time: 02:32)



So, as we said the output is jth channel is given by X j plus Z j at j goes from 1, 2, K. Noise is 0 mean which were is N j and this is considered to be independent from channel to per channel. Subsequently, when we talk about colored noise, we will assume that channel is not independent, and then will compute a capacity expression that well. So, for the first talk, we are assuming that the noise is independent from one channel to another. So, noise N 1 is independent of N 2 is independent of N 3 is independent of N K. Now, we are interested in finding of the power distribution among various channels such that our channel capacity is maximize.

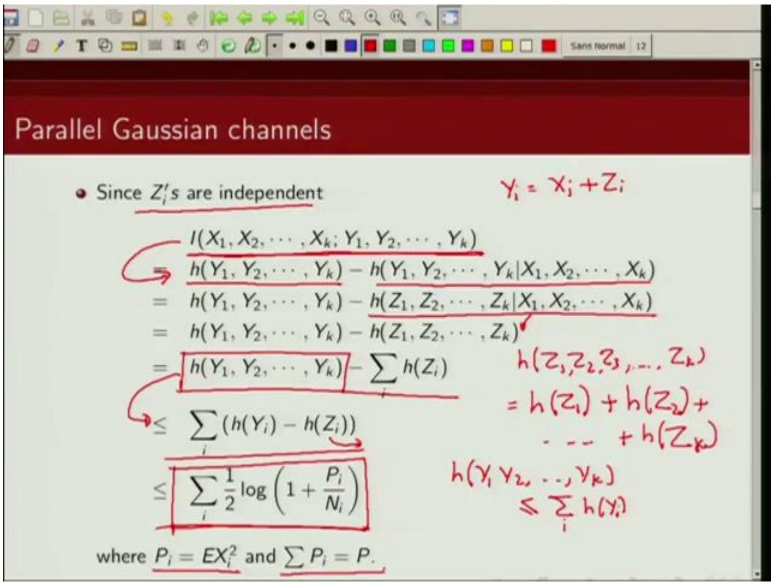
So, we have a common power constraint which is given by this. So, how should we allocate power to X 1, X 2, X 3, X K such that we maximize our channel capacity? This is a problem we are will look at. So, from the definition of channel capacity, we can write channel capacity as maximizing the mutual formation between X 1, X 2, X 3, X K and Y 1, Y 2, Y 3 and Y K. And this maximize over input distribution with this common power constrain which says expected value of j square is less than equal to some constraint p.

(Refer Slide Time: 04:28)



So, as I said this is how a my parallel Gaussian channel looks like I have these inputs X 1, X 2 and X K and these are my output Y 1, Y 2, Y K the noise samples are independent denoted by Z 1, Z 2, Z K.

(Refer Slide Time: 04:53)

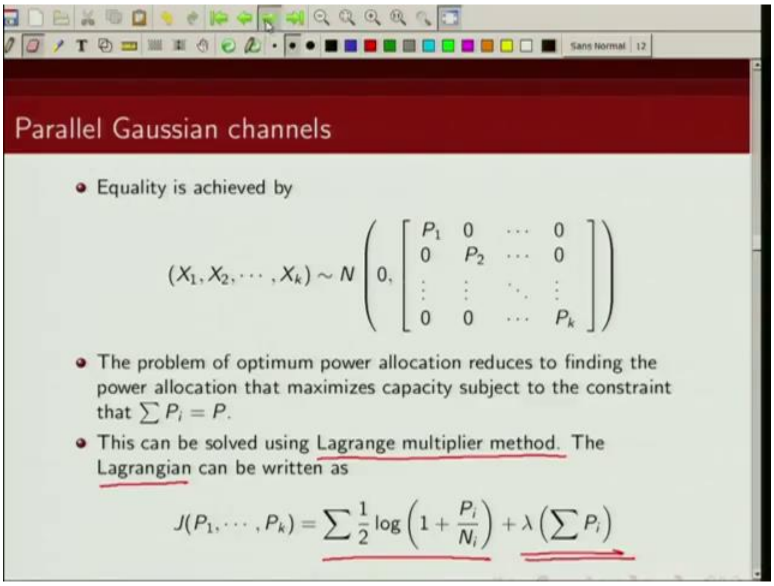


Now, since my noise is independent, so first this is the mutual information between the input of the channel and output of the channel for the parallel Gaussian channel. Now, from the definition of mutual information, I can write this as differential entropy of Y 1, Y 2, Y 3, Y K minus differential entropy of Y 1, Y 2, Y K given X 1, X 2, X 3, X K. Now, what are these Y i's now Y i's are X i plus Z i and X i's are independent of Z i. So, what would be, so given X 1, X 2, X 3, X K, the uncertainty in Y 1, Y 2, Y 3, Y K can be written as differential entropy of Z 1, Z 2, Z 3, Z K given X 1, X 2, X 3, X K. And what is a differential entropy of Z 1, Z 2, Z 3, Z K given X 1, X 2, X, 3 sine Z i and X i are independent, this can be written as uncertainty in Z 1 Z 2 Z 3 Z K.

And since these noise over these channels parallel channels this noise are independent, we can write this joint entropy of Z 1, Z 2, Z 3, Z K, we can write this as an differential entropy of Z 1 plus differential entropy of Z 2 plus differential entropy of Z K this is because z Is are independent. So, then what we have is mutual information between X 1, X 2, X 3, X K and Y 1, Y 2, Y 3, Y K is given by this expression. Now, we can upper bound this point differential entropy, this joint differential entropy Y 1, Y 2, Y K using apply chain rule and using the fact that conditioning cannot increase entropy using these two fact we can upper bound this by this. Joint entropy, you can write it in terms using chain rule in terms of some of conditioning entropy and we know that conditioning cannot increase entropy. So, those can be upper bounded by individual Y i's and that is how we are getting this inequality.

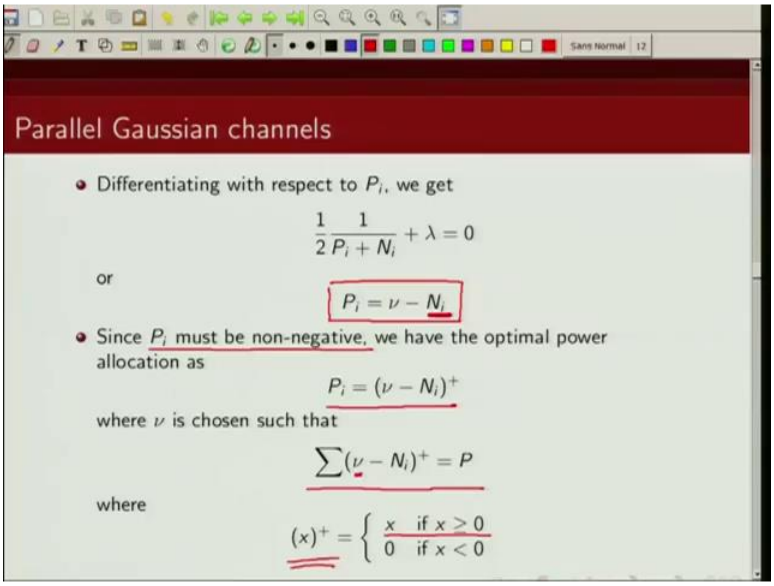
So, if I plug that result here what I can do is I can write this mutual information between X 1, X 2, X K and Y 1, Y 2, Y 3 and Y K as upper bounded by this quantity. Now, this can be upper bounded by, so differential entropy of Z i, this is basically half log 2 pi E times N which is noise variation. And differential entropy of Y i can be upper bounded by differential entropy of Gaussian distribute random variable with various given by P. So, if you plug that value in here, we can write that mutual information can be upper bounded by this term. Please refer to the previous lecture where we have you know derived this. So, mutual information can be upper bounded by this term, where this pi is given by expected value of X i square and there is a common power constraint.

(Refer Slide Time: 09:34)



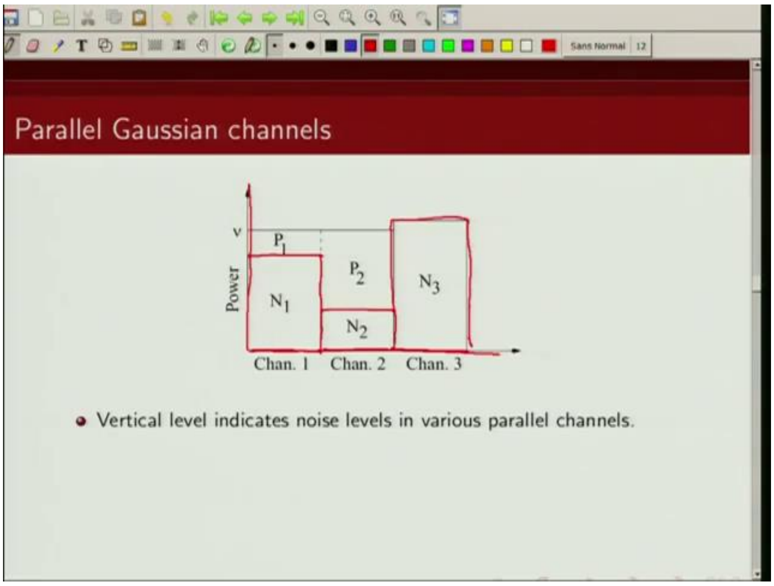
The equality is achieved when the input X is Gaussian distributed with mean 0 and varies basically covariance matrix given by this. This we already know from a result that we have proved in the lecture on differential entropy. So, then the problem of optimum power allocation reduces to finding the power allocation that maximizes the power subject to the constraint that summation of this P i is basically is equal to P. Now, we can solve this using Lagrange multiplier method. So, we form the Lagrangian which is this capacity expression plus lambda time these constraints.

(Refer Slide Time: 10:35)



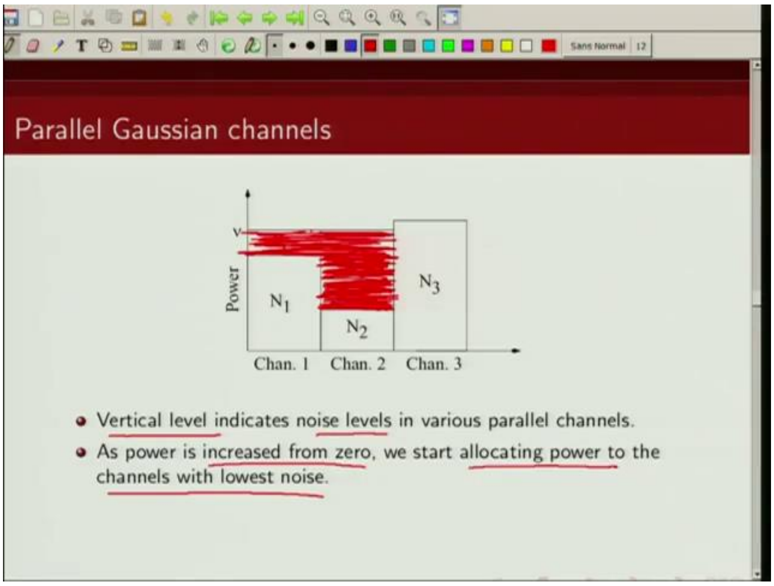
Now, if we differentiate this Lagrangian with respect to P i, what we get is this that P i should be equal to nu minus N i, I just define what my nu is. Now, power must be non-negative right. So, I am writing this P i as nu minus N i plus where this operator means that this is equal to x, if x is greater than equal to 0; otherwise this is 0. And how do I get this value of this nu the summation of this should add up to my power constraint that I have. So, what does this P little attention to the power allocation policy that we are getting? So, let us say we have some nu, now what does it says. So, if the noise is less where in a particular i th channel, you are going to allocate more power, whereas for the channels which have higher noise you are going to allocate less power, so that is what this optimal policy said.

(Refer Slide Time: 12:01)



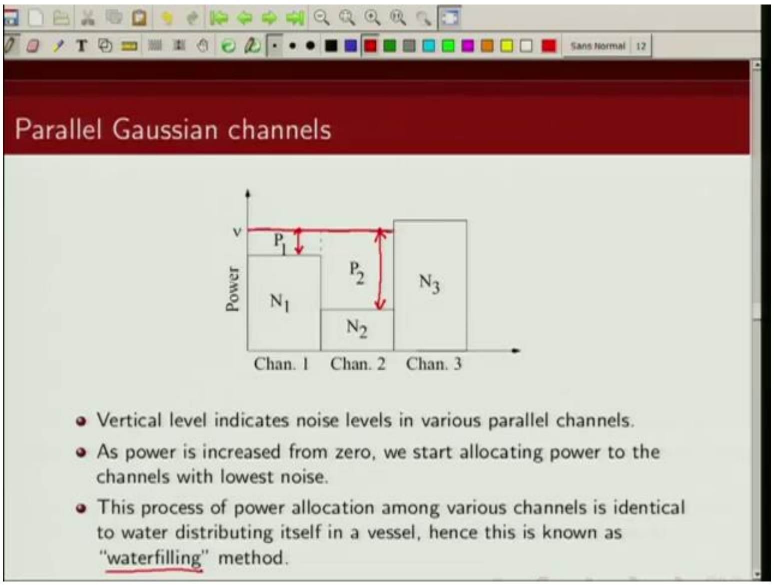
Now, this can be visualized in this fashion. So, let say on the x-axis, I am plotting these key channels. So, I have this channel 1, channel 2, channel 3 like that I am plotting. And here I am plotting power. So, initially I plot the noise power. So, if you see this level high, it means the noise level is higher. So, noise level of N 1 is more than N 2, but it is less than the noise in the channel 3, so that is what this height of this N 1 into N 3 means.

(Refer Slide Time: 12:48)



Now, the power allocation strategy that I derived that is we that I should first add power, I should first add power to this channel first add to power this channel, then I should start filling power to this giving power to this channel and this channel, and this is my new then this channel would not get any power. So, as I said the vertical level indicate the noise levels and various parallel channels. And hence, the power increased from 0, we start allocating power to the channels with lowest noise. As I just showed you I first going to pore power into this channel number 2, until I reach this level of and then I start pouring channel for power in both of these channels.

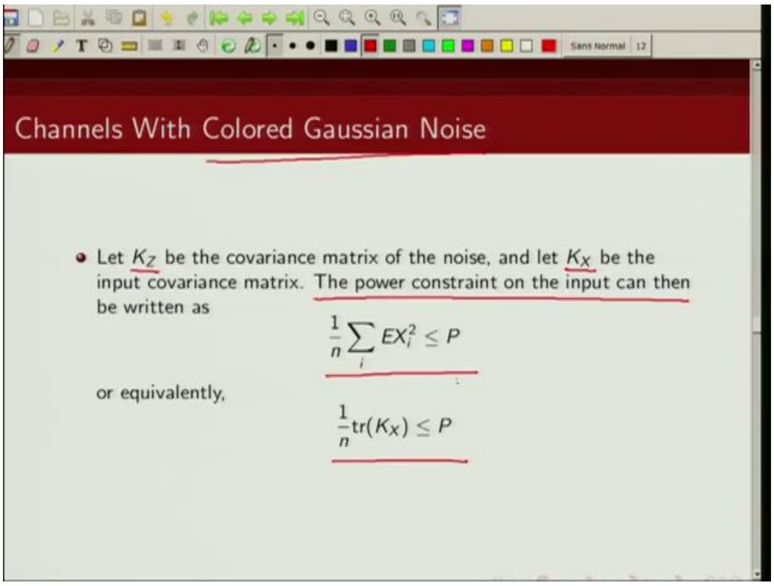
(Refer Slide Time: 13:46)



Now, the process of allocating power to various channels is similar to the way you will pour water in a vessel. Let say you have a bucket and I put some things in a bucket. So, some items which tall and some items which are smaller and I start pouring water into this bucket. So, how will water get filled? So, first it was start filling over the items which are small right start filling them, subsequently start filling water on the items which are tall. So, this is precisely the way this power allocating strategy is happening; you can think of those items that put in a bucket as noise level. So, I will start once I start allocating power, I give power to those channels which have less noise until power plus noise become equal to noise of a mother channel and then that channel will also start getting some power.

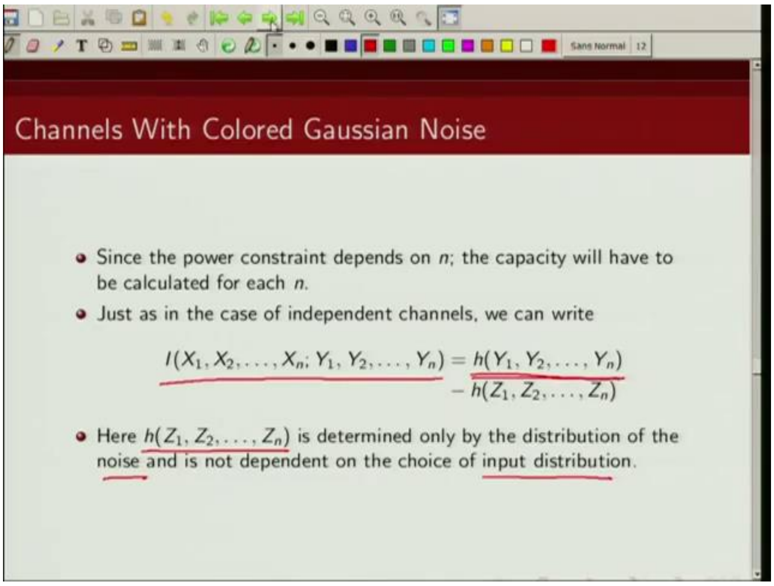
So, you can see here, from this example is my new was this then my power allocation strategy would have been I would have located power P 1 here which is less than the power I located with channel 2 which is P 2 and I am not locating any power to channel three. So, this is known as water filling this way of power allocations known as water filling because this is precisely the way water would fill if you have things kept in the bucket of various types.

(Refer Slide Time: 15:34)



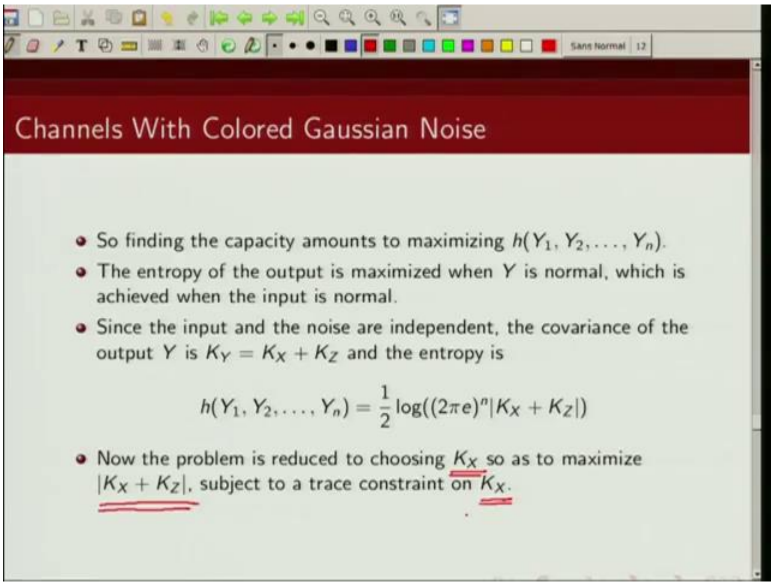
So, we just talked about parallel Gaussian channel where noise was independent. So, noise at channel 1, noise at channel 2, those are all independent. Now, what happened if the noises are not independent that is a case corresponding to colored Gaussian noise? So, now we are considering that noise is not independent across the channels. So, let K of z be the covariance matrix of the noise, and let K of x be the input covariance matrix. And we can write the power constraint on the input like this which can be equivalently written in terms of trace of this covariance matrix on the input covariance matrix K of x. So, this can be written as the trace of 1 by n time of space of input covariance matrix that is less than equal to P.

(Refer Slide Time: 16:42)



Now, here the power allocation depends on n, so we have to calculate capacity for each n. So, this is similar to the case of independent channel; if you recall, the mutual information between X 1, X 2, X 3 and X n, and Y 1, Y 2 and Y n can be written as differential entropy of Y 1, Y 2, Y 3, Y n minus differential entropy of Y 1, Y 2, Y n given X 1, X 2, X 3, X n. And what are these Y is Y is are X i plus Z i. And since input X i are independent of Z i, we can write that preferential entropy of Y 1, Y 2, Y 3, Y n given X 1, X 2, X 3, X n as differential entropy of Z 1, Z 2, Z 3, Z n. However, there is one difference from what we had earlier, since earlier the noise was independent we could just write this as summation of differential entropy of Z i which we cannot do because now we have colored noise. Now, this differential entropy of Z 1, Z 2, Z 3 and Z n is determined by the distribution of noise, and it does not depend on input distribution. So, to maximize this capacity, we need to concentrate on this term joint entropy of Y 1, Y 2, Y 3, Y n.

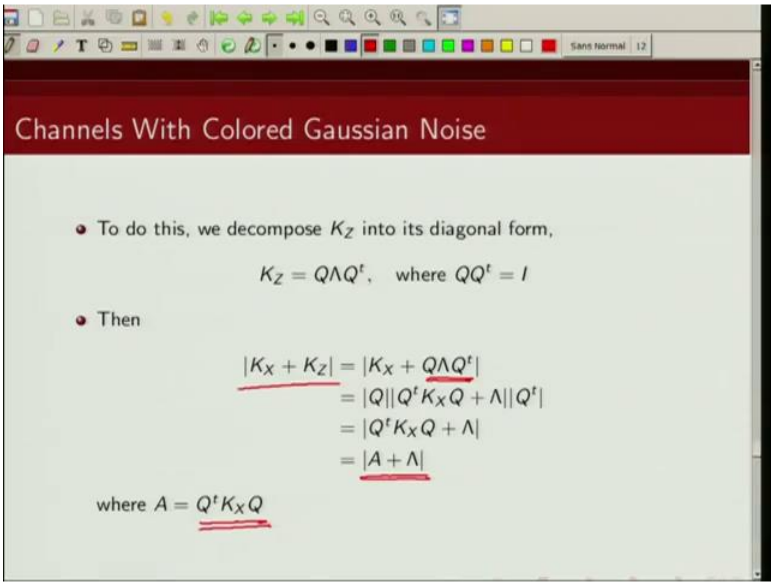
(Refer Slide Time: 18:31)



So, as I said maximizing capacity is basically we have to find out what is a maximum value of this differential entropy joint and differential entropy h of Y 1, Y 2, Y 3, Y n. Now, we know that entropy is maximize Y is normal distributed. So, for Y to be normal distributed because Z is normal distributed X also has to be normal distributed some of two Gaussian random distributed variable is also a normal distribute random variable. So, the entropy is maximize when Y is normal distributed, which is achieved when input is also normal distributed.

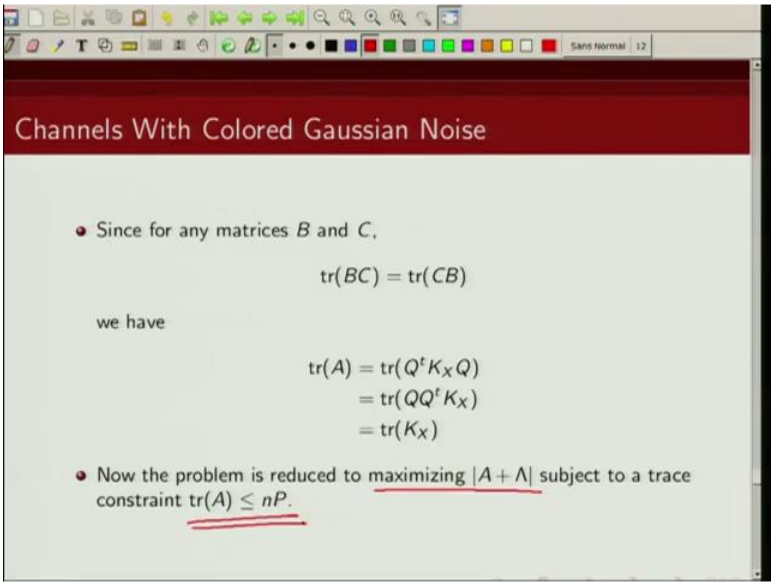
Now, since input and noise are independent. So, the covariance of the output which we write as denote by K of Y is equal to K of x because of the co variance of the input plus co variance of the noise which is K of Z. So, differential entropy of Y 1, Y 2, Y 3 and Y n can be written as half log of 2 pi e raise power n and determinant of this covariance matrix of the input plus covariance matrix of the noise. So, the problem of maximize is this then reduces to choosing our input covariance matrix K z such that this determinant is maximize, subject to the input power constrain which we wrote in terms of condition on tries of this input covariance matrix.

(Refer Slide Time: 20:38)



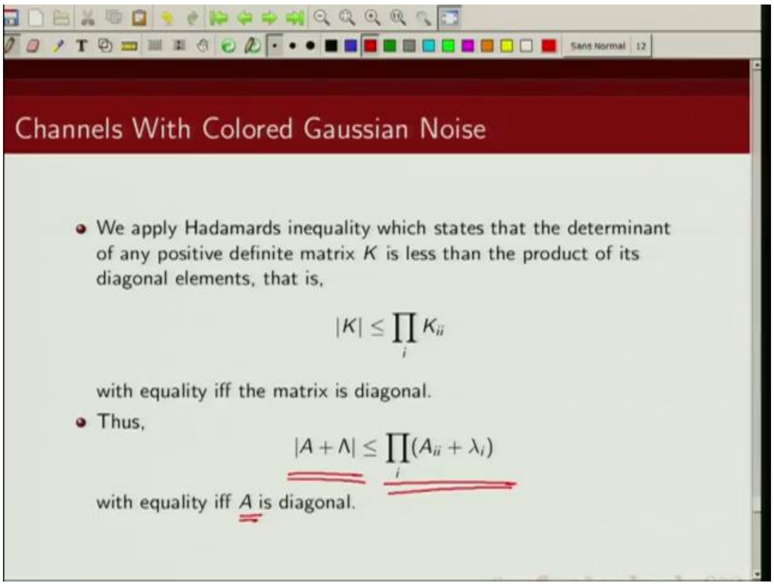
Now, we can decompose this covariance by K Z into Q diagonal matrix and Q transpose where Q Q transpose is a identity. So, if we do that, we can write this determinant of K Z plus K Z as K X plus this is the decomposition of this matrix K of Z. And after we do simple matrix manipulation, we can show that this can be written as determinant of matrix A plus some diagonal matrix lambda, where this matrix A is given by Q transpose covariance matrix of the inputs times Q.

(Refer Slide Time: 21:28)



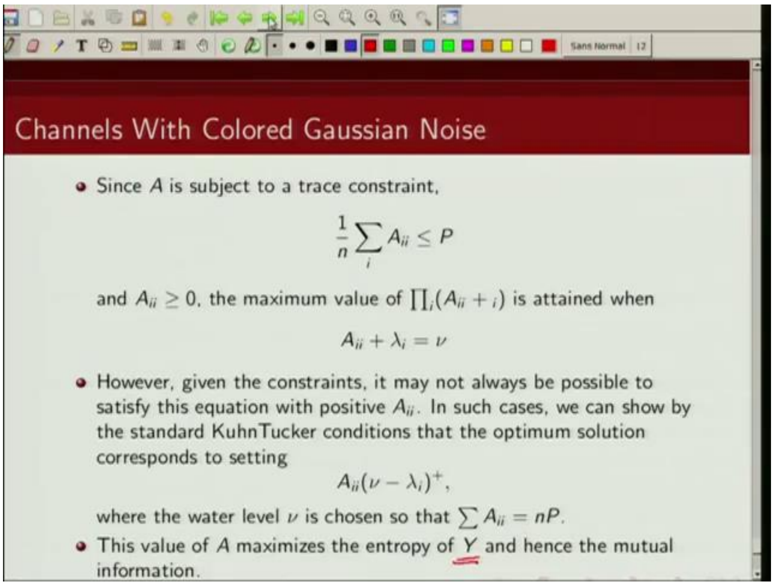
Now, we are going to make use of property of matrix, which says if we will have any matrix B and C trace of this matrix B and C is same as trace of the matrix CB. So, we can write the trace of matrix A as trace of matrix Q transports K X Q which we can write as trace of Q Q transport K X. Now this is identity matrix. So, then trace of A is nothing but trace of this input co variance matrix K of x. So, then the problem reduces to maximizing this determinant of A plus lambda subject to this trace constraint where the trace of A is less than equal to n times P.

(Refer Slide Time: 22:22)



Now to solve this, we are going to take help of Hadamards inequality. Now, recall in the lecture for differential entropy, we have proved this Hadamards inequality. So, Hadamards inequality stays that determinant of any positive definite matrix K is less than the product of its diagonal element. So, this determinant of K is less than equal to summation is product of these diagonal elements with equality happening only iff the matrix is diagonal. So, then determinant of this matrix A plus lambda is going to be less than product of the diagonal elements of this matrix and this will happen with equality only iff this matrix A is also a diagonal matrix.

(Refer Slide Time: 23:30)



Now, since A is subject to this trace constraint then by this A i has to be greater than equal to 0 then the maximum value of this product of this diagonal elements is obtained when this is equal to some quantity mu. However, A i's cannot be negative. So, it may not be always possible to satisfy this constraint on A i's. So, in such case, we show the standard Kuhn Tucker conditions that the optimum solutions corresponds to basically this where we choose this water level nu such that summation of A i is equal to n times P. So, this is similar to basically water filling in this spectral domain. So, the value of A that maximizes the entropy is basically given by this; and this will in turn because this value of A will maximize deferential entropy of Y 1, Y 2, Y 3, Y n, this in turn will also maximize the mutual information. So, this is how we are going to do in power allocation in case of Gaussian channel with colored noise. So, with this, we are going to conclude our discussion on parallel Gaussian channel.

Thank you.